



Asian Research Association



Endo-CNN: A Novel Deep Learning Model for Gastrointestinal Diseases

Esha Saxena ^{a,*}, Suraiya Parveen ^a, Mohd. Abdul Ahad ^a, Meenakshi Yadav ^b

^a School of Engineering Sciences and Technology, Jamia Hamdard, 110062, New Delhi, India

^b Galgotias College of Engineering and Technology, Gr. Noida, 201310, Uttar Pradesh, India

* Corresponding Author Email: eshasaxena.1987@gmail.com

DOI: <https://doi.org/10.54392/irjmt2632>

Received: 27-03-2025; Revised: 01-04-2026; Accepted: 08-04-2026; Published: 15-04-2026



Abstract: Gastrointestinal (GI) diseases often represent the most frequent and common high-risk diseases. Wireless capsule endoscopy (WCE) has changed the landscape of diagnosing and treating patients. Endoscopists commonly utilize wireless capsule endoscopy to assess the majority of intestinal conditions, particularly with respect to polyps and ulcers. The use of WCE has shown a ten percent increase in Indian hospitals. Medical assessments are typically time-consuming and expensive, especially given the necessity to investigate directly from endoscopic videos. These confines are alleviated with the assistance of artificial intelligence and deep learning, which provide an efficient platform for instantaneous defect detection. The objective served by this examination is to assist endoscopic image classification work for clinical investigators. The paper proposed a deep-learning model named Endo-CNN based on convolutional neural network to classify endoscopic images according to the identified disease. The classes of images include polyps, ulcerative colitis, esophagitis and a healthy colon. Data augmentation occurs to reduce the imbalance of datasets and to evaluate the model performance that exceeds 48,000 images. The model achieves a positive accuracy rate with all the image classes. There are various aspects of an identified disease because of the variety of sizes, shapes and textures as well as colors. The paper also performs a comparative study of the designed model and against other pre-trained models (VGG16, ResNet50 & DenseNet). This paper can act as a baseline for many future solutions in the field of gastroenterology.

Keywords: CNN, Data augmentation, Deep learning, Endoscopic images, Gastroenterology, WCE.

1. Introduction

AI is exploring gastroenterology in every possible way. Deep learning techniques provide strength to large and complex problems. Many medical fields like cardiology, nephrology, radiology, dermatology and so on are taking advantage of DL methods [1]. Globally, the demand for more advanced neural architectures is increasing with respect of problem domain. In the medical field, number of images are obtained by different sources, which help to diagnose the affected part of the body. The intestinal investigation is one of the major research areas in gastroenterology. The diseases related to large and small intestine is not easy to diagnose by other investigation methods. Mostly, endoscopy is performed as an investigation tool to obtain images. Wireless capsule endoscopy (WCE) empowers the doctors to examine the entire intestinal tract in detail with just an endoscopic capsule [2], which is swallowed by the patient (carrying a battery, light-enabled wireless minute camera, light-emitting diodes, and a radio frequency emitter within it). This is a painless procedure that records a long 8-9 hours of video. The

challenging task for doctors is to study such a long video in short time span. With the help of the computerized system, RGB images are recorded [3], and it's very difficult to identify the disease from thousands of images. Generally, to examine gastrointestinal tract disorders, an exhaustive series of steps are done. Most of the hospitals and gastroenterology centers have the facility to conduct WCE. Every electronic device processes RGB color model over images and shows distinct color patterns, and texture properties on imaging devices [4]. This creates demands for different DL models for various medical imaging systems.

The digestive tract [5] encompasses the organs that facilitate digestion, beginning from the mouth and ending at the anus. The Tract incorporates all supporting organs of digestion [6]. There are many gastrointestinal tract-related diseases worldwide that need to be diagnosed early before permanent organ damage occurs. Common disorders are polyps, recognizable colon cancer and bleeding problems. All the above disease forms are considered for treatment. For better accuracy and speed to classify images, you need to

Implement complex neural architectures. Doctors and researchers are stressing the intense application of DL techniques in higher-accuracy implementations of image problems. The actual purpose of our research is to recognize the smallest portion of undesired GI tract abnormality achieved by designing more efficient CNN-based novel architectures than previously trained DL models [7]. AI Techniques have shown great success in analyzing medical data. Many methods have already been developed and applied to classified intestinal diseases, but this paper seeks to position a model that may detect a multitude of diseases in a dataset. A Convolutional Neural Network (CNN) is the most advanced neural network for medical images and all images comprise a grid (2D) of pixels. Many deep learning [8, 9] architectures, such as AlexNet, VGG, ResNet, GoogleNet.

Deep convolutional neural network (DCNN) [10], which is a dense neural network with several processing layers and learns features from input data (text/image). Deep learning techniques are continuously working to enhance the accuracy of endoscopic images, reduce overfitting and also decrease study time. Many researchers follow CNN architectures [11, 12] for feature extraction, image classification and segmentation. The architecture comprises of input layers, output layers and multiple hidden layers. The convolution layer applies a number of filters to an image, which can vary according to the architecture design selected. A feature map is obtained for processing [13]. To reduce spatial dimensions pooling layer is used and the most popular layer is Max pooling layer. Mostly, neural networks suffer from overfitting and vanishing gradient problems. The work presents a novel deep learning model (Endo-CNN) for the classification of endoscopic images into four types: polyps, ulcerative colitis, esophagitis, and healthy colon. The original dataset was composed of 8,000 images and was expanded with data augmentation to over 48,000 images (x6 times) to address class imbalance. In contrast to general pre-trained architectures, such as VGG16 (138M parameters) or ResNet50 (25M parameters), Endo-CNN was designed specifically for medical imaging with a more computationally efficient and domain-optimized usage of compute resources.

The comparative experiments show Endo-CNN was able to deliver a positive accuracy across all classes, while significantly outperforming the pre-trained architecture in sensitivity and F1 score for the less represented category, esophagitis, and showing the promise of AI-based approaches in lowering diagnostic error and improving clinical decision making in diagnostic gastrointestinal endoscopy.

2. Related Work

Research articles are being gathered from both conference and publication methods to explore the use

of deep learning techniques within healthcare using multiple sources. The table would be the list of objectives and findings from existing research publications that allow the researcher an avenue for analyzing how they will approach this study. For example, [5, 2] refers to gaining knowledge regarding the digestive system and the role of endoscopy. Articles such as [8, 9, 17, 20] provide brief knowledge regarding deep learning techniques and convolutional neural networks. Table 1 provide an overview of the findings from the articles, which facilitate identifying and issues within the use of AI technologies for analysing endoscopic images

3. Methodology

Deep learning methods have been used in this section to study intestinal endoscopic images obtained by WCE-recorded videos. The proposed methodology (figure 6) will help quickly diagnose the diseases.

3.1 Data Description

Wireless capsule endoscopy is one of the most advanced and widely in practices for detecting intestinal diseases. The images were obtained by 8-9 hours of video recorded. In our research, the dataset is chosen from a freely available online platform, Kaggle [14]. The site is enriched with a collection of datasets, offering a huge platform for researchers. We have used the Kvasir-V2 dataset (www.kaggle.com/datasets/yasserhessein/the-kvasir-dataset), contains 8,000 endoscopic images from 8 different classes. They are Dyed-lifted polyps, Esophagitis, Dyed-resection margins, Normal pylorus, Normal cecum, Polyps, Normal z-line, Ulcerative colitis.

Figure 1 depicts the endoscopic images from the Kvasir dataset from each class. The dataset contains endoscopic images showing different GI tract diseases [15]. The images are of size 720 * 576 pixels. The dataset is found to be a good choice for this investigation as it contains the most common diseases of GI tract.

3.2 Data Preparations

The data pre-processing [16, 17] is done beforehand, feature extraction using various strategies such as noise reduction, data augmentation, contrast enhancement, dimensionality reduction and image sharpening. Python is used as the programming language. Many libraries exist for image pre-processing in Python and some of the more popular ones would be OpenCV and Tensorflow. Python is a great platform for data understanding and data visualization [18].

Image preprocessing involves transforming raw image files into something usable and meaningful. By preprocessing image data, you can remove unwanted distortions and improve specific attributes.

Table 1. LR table

Paper	Objective	Key Findings
[11]	This research is about a method of medical image processing with the use of multi-feature fusion to better extract image features from chest, lung, brain, and liver images. A result of using this method has been an increased retrieval accuracy of more than 5% compared to traditional techniques, which addresses some difficulties associated with extracting features from a sample medical image.	<ul style="list-style-type: none"> - The proposed image processing technique utilizes multi-feature fusion and shows remarkable results for feature extraction with respect to images of the chest, lungs, brains and livers, therefore providing a very accurate means of image retrieval. - The data gathered indicates that a large number of medical images have not been studied and compared and therefore represent substantial areas for additional future research.
[14]	This article describes how wireless capsule endoscopy (WCE) is used to identify stomach problems and uses many different forms of artificial intelligence to analyse and classify the images obtained through WCE. Pre-trained models like VGG16 were used to achieve very high levels of accuracy and precision when accurately identifying diseases based on the WCE photographs taken.	<ul style="list-style-type: none"> -VGG16 had an overall accuracy score of 96.33% along with recall, or sensitivity, of 96.37%, and precision, or positive predictive value, of 96.5%. Finally, the model achieved Cohen's kappa score of 0.96, and a Matthews Correlation Coefficient (MCC) measure of 0.95 as well. -ResNet-18, on the other hand, had low classification metric scores because most misclassifications were in specific disease categories. Overall, the proposed system for detecting irregular WCE images had 92% precision.
[21]	A new model named "BIR" is introduced in this work, which allows for the classification of wireless capsule endoscopy images based on gastrointestinal tract diseases. BIR was developed using MobileNet and CNN technology. Evaluation has shown that BIR performs better than current technologies in both accuracy and precision at classifying these images while overcoming the issues of having an imbalanced dataset and providing a more powerful and efficient diagnosis in medical imaging.	<ul style="list-style-type: none"> -The classification of WCE images by using the proposed BIR model attained maximum accuracy (0.993), precision (1.000), recall (0.994), F1 score (0.997) and Cohen's kappa (0.995). According to the results obtained when testing the BIR model using Google-collected WCE images, its classification performed with an accuracy level of 0.978. Therefore, this study indicates positive and encouraging results regarding the performance of the BIR in terms of its accuracy. -The experiments were performed using both an unbalanced data set with a weighted cross-entropy and a balanced data set using data augmentation methods.
[25]	The study analyses how well different DL techniques, specifically RCNN series, are at detecting early gastric cancer (EGC) by using gastroscopic images. AI methods have been shown to outperform human endoscopists for detecting EGC via endoscopic exam, and it is imperative for clinicians to diagnose EGC early to improve patient outcomes.	<ul style="list-style-type: none"> -The Mask RCNN method had an accuracy of 93.5% and specificity of 90.08% when detecting EGC, whereas the accuracy of the Faster RCNN was less than that of the Cascade or Mask RCNN model. Thus, the lower accuracy of the Faster RCNN as compared to Cascade and Mask; means less performance in doing so. In terms of Specificity, Cascade had 94.6%; which is higher than 90.8% for both Faster and Mask. -While these three models produced similar results when accurately diagnosing the same patients with gastric cancer via GST, none of the models were able to meet the demands of the clinical environment and additional research is needed to improve on the accuracy of diagnosis using Artificial Intelligence (AI).
[30]	In this article, we will look at how recent advances have been made in the fields of deep learning (DL) and specifically, DL image data augmentation techniques have improved model performance on tasks such as image classification and object detection. We will address the prospects of	<ul style="list-style-type: none"> -Using augmentation techniques such as random cropping and PCA color augmentations has been effective at reducing overfitting on Deep Learning networks, resulting in improved overall accuracy (>1% reduction in error rate).

	<p>leveraging both traditional and neural augmentation methods for a better working experience going forward within this domain.</p>	<p>-The paper notes that artificially increasing the dataset size through augmentation methods can simulate big data benefit from small amounts of data.</p> <p>-GAN Generated augmentation methods showed better classification boundaries and can lead to better results in analyzing biomedical images.</p> <p>-Combining traditional augmentation methods with neural network-generated methods may yield the best results according to this research.</p>
--	--------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

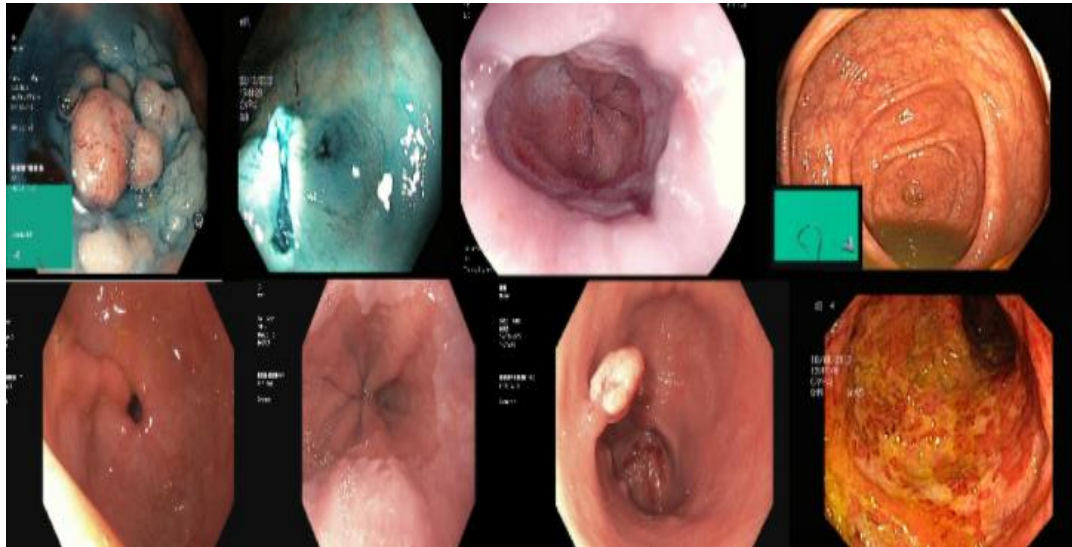


Figure 1. Endoscopic views of colonic and gastric pathologies

Preprocessing is an important initial step for preparing an image before inputting it into AI models. The image preprocessing techniques include:

- Resizing:** Images are resized to a standard size is ideal for deep learning/machine learning approaches to run correctly. The `resize()` method of OpenCV can resize images.
- Grayscale:** Changing colored images to grayscale possibly simplifies the image data and decreases analytical space for models.
- Noise reduction:** You can use two techniques called smoothing and blurring to reduce noise. Smoothing and blurring are usually completed using the `GaussianBlur()` and `medianBlur()` methods.
- Normalization:** It is the technique of changing the intensity values of pixels to a set range, in many cases, it's 0 to 1. Normalization may lead to better ML/DL models. The `normalize()` from scikit-image can be implemented.
- Contrast enhancement:** You can change the contrast of an image using histogram equalization. The `equalizeHist()` method is utilized to enhance the contrast of images.

Data augmentation techniques will help with the need to overcome the imbalance of datasets. It is a

method that helps to raise the number of input images without collecting new ones by including transformations. Figure 2 shows the following transformations.

- Rotate images by up to 40 degrees
- Translate horizontally by up to 20%
- Translate vertically by up to 20%
- Shear transformations
- Zoom in/out by up to 20%
- Randomly flip images horizontally

The easiest way to get rid of over-fitting is to have a very large dataset and train the model. In this paper, the image's pixels are normalized. Transformations were done such as rotations, zoom in/out, and flipping [19].

The classes have 8,000 images in total and after data augmentation, all the parameters are checked over 48,000 endoscopic images. We have used 20% of the data for validation (`validation_split=0.2`) and 80% for the training set. The dataset is then transferred for cross-validation, where the trained model is evaluated for a performance check over the validation set. The image size is reduced to reduce memory usage.

3.3 Proposed Methodology

- a. Input: Upload the endoscopic dataset of 8000 images articulated into 8 classes. Each class is labelled with a unique disease name and number.
- b. Pre-processing: The image dataset is pre-processed before the learning set is processed. Different methods are applied to enhance image features. Data augmentation was applied (rotate, translate horizontally/ vertically, zoom, shear transformations and flip). The total number of images invaded by data augmentations was 48,000.
- c. Learning set: Label hyperparameters, epochs and batch size. The endoscopic dataset is split into training, test and validation sets. This step

is referred to as the learning stage of your model.

- d. Evaluation: The model learns from the classifier to predict the labels. Feature extraction was completed using a deep neural network. Evaluate metrics were evaluated and compared to the pre-trained models.

Figure 3 explains the designed methodology for the proposed model Endo-CNN. The model is inspired by CNN's internal architecture [20] and follows all techniques to make this model work efficiently. An automated identification system processes and analyzes a large dataset of images using advanced deep-learning methods. In our proposed model, the dataset is pre-processed and trained with data augmentation [21]. The model consists of 32, 64 and 128 filters and is free from the problem of overfitting.

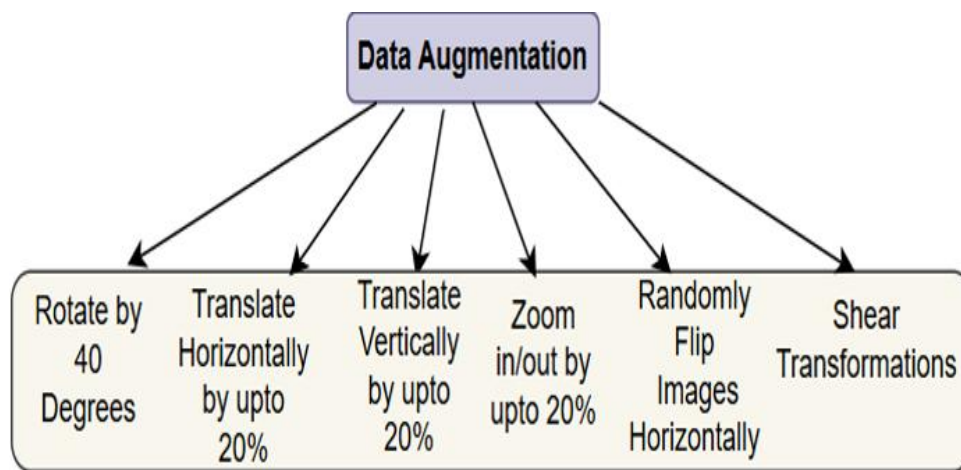


Figure 2. Data Augmentation transformations

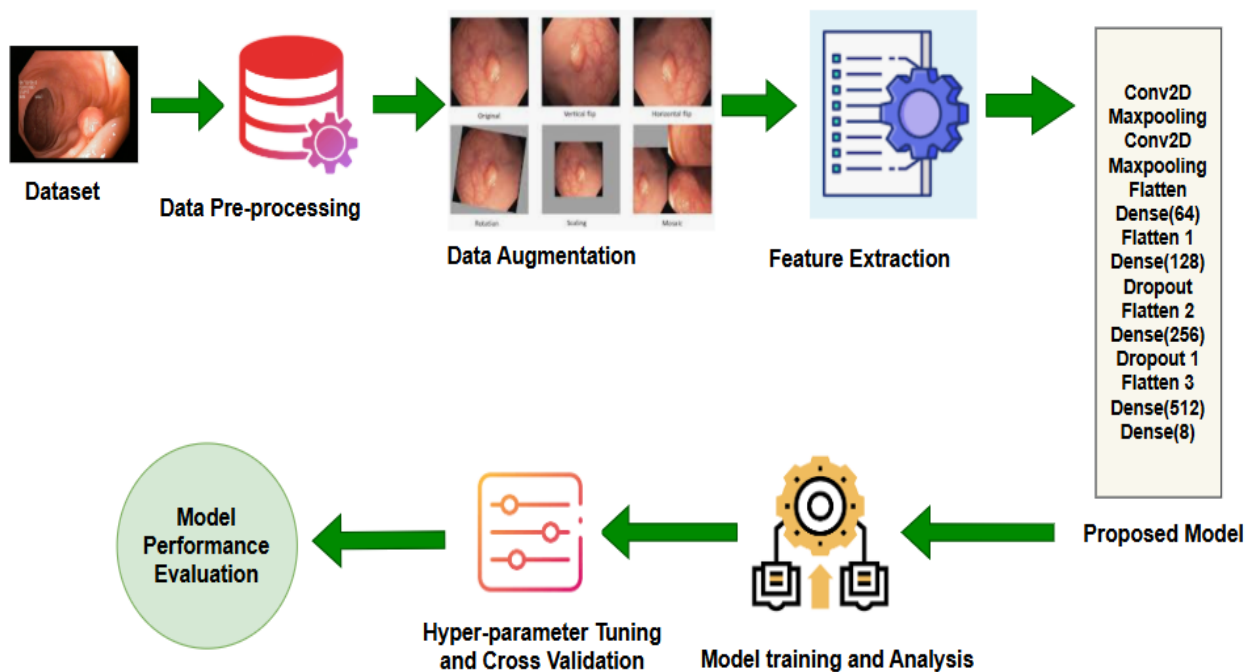


Figure 3. Proposed methodology

It shows a good accuracy rate with different endoscopic image classes. Hyperparameter tuning [22] is performed to help in finding the best model parameters. Image size (224, 224) is taken to reduce the memory usage. Activation function=relu, total parameters are 6,159,208, input_shape = (224, 224, 3).

4. Architectural Overview of the Proposed Model: Endo CNN

The Endo-CNN model, an endoscopic model, is primarily based on the architectural design of CNN, which contains convolutional layers for feature extraction, a MaxPooling layer, a Flatten layer, and dense fully connected layers for classification [23]. Dense (64, 128, 256) is being used to understand the composed features of the endoscopic images. The total parameters are 6,159,208. The dropout is a regularization technique [24] that keeps the model from overfitting and reduces interdependent neural learning.

Table 2 describes the model, which shows a convolutional layer with 32 filters. The shape of the feature map after the 2D convolutional layer: (None, 222, 222, 32), kernel size= (3, 3), input_shape= (224, 224, 3) RGB images, pool_size= (2, 2) and activation function ReLU. The dropout of 20% is applied to balance overfitting. The image is resized to reduce memory usage, img_size = (128, 128). Using the Keras and TensorFlow software packages, training was performed on the previously described 6 search space configurations with a learning rate of 0.001. The model size is 23.50MB and it was trained for 15 epochs at a reduced batch size of 8. Training cost calculated as 41sec * 15 epochs=~10 min, resulting in a total training time of approximately 10 minutes when using an NVIDIA RTX 3090 GPU. During inference, the final model produced an average processing time of 12ms/image, making it suitable for real-time clinical implementation, particularly in high-pressure settings such as endoscopy.

4.1 Steps to Design: Endo-CNN Algorithm for Disease Identification

Step 1. Data collection

Collect annotated endoscopic image datasets covering all target disease classes

Step 2. Data Preparation

Dataset Partitioning: Split the dataset into training, validation, and testing subsets.

Normalize and resize images, Resize image I to size S

$I \in \mathbb{R}^{H \times W \times C}$ where $H=224$, $W=224$ and $C=3$.

Apply data augmentation (rotation, flips, noise) to reduce overfitting.

$$D = D_{\text{train}} \cup D_{\text{val}} \cup D_{\text{test}}$$

Perform class balancing.

Step 3. Model Definition

Convolutional Layers: Generate a series of hierarchical feature maps.

Batch Normalisation and Dropout Layers: Stabilizes training and reduces overfitting.

Flatten Layer: Converts the feature maps to a one-dimensional vector.

Fully Connected Layers and output Layer.

Conv Layers: $F = \text{ConvNet}(I'; \theta_c)$

Flatten: $v = \text{Flatten}(F)$

Fully Connected: $h = \phi(W_{fc}v + b_{fc})$

Output: $\hat{y} = \text{Softmax}(W_{out}h + b_{out})$ (multi-class)

Where $\theta = \{\theta_c, W_{fc}, b_{fc}, W_{out}, b_{out}\}$ and ϕ denotes activation (typically ReLU)

Step 4. Model Training

- Loss Function: Weighted Cross-Entropy Weighted categorical cross-entropy (for classes $j = 1, \dots, K$):

$$L(\theta) = -\frac{1}{M} \sum_{i=1}^M \sum_{j=1}^K w_j \cdot \mathbb{1}(y_i = j) \log \hat{y}_{i,j}$$

where w_j is the class weight, M is the mini-batch size.

- Regularization: Dropout, Batch Normalization. Apply dropout p and batch normalization at designated layers:

$$h' = \text{Dropout}(\text{BatchNorm}(h), p)$$

- Hyper parameter tuning: Batch size, LR grid search.
 - Optimizer
- Update θ using Adam: $\theta \leftarrow \theta - \alpha \nabla_{\theta} L(\theta)$

Learning rate $\alpha = 0.001$

Early Stopping: If L_{val} does not decrease for T epochs ($T = 10$), stop training.

Step 5. Evaluation

- Calculate Accuracy, Precision, Recall, Specificity, and F1 Score.
- Monitor ROC-AUC for imbalance robustness.

$$\text{ROC-AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx$$

Step 6. Deployment and Learning

- Implement inference pipeline: For new I_{test} , obtain $\hat{y}_{\text{test}} = f_{\theta^*}(I_{\text{test}})$ Where θ^* are the final trained parameters. Continuous learning: periodically retrain with new clinical data.

Table 2. Description of Endo-CNN layers

Layer	Output shape	Parameters
Conv2d (Conv2D)	None, 222, 222, 32	896
Max_pooling2d (MaxPooling2D)	None, 111, 111, 32	0
conv2d_1 (Conv2D)	None, 109, 109, 32	9,248
max_pooling2d_1 (MaxPooling2D)	None, 54, 54, 32	0
flatten (Flatten)	None, 93312	0
dense (Dense)	None, 64	5,972,032
flatten_1 (Flatten)	None, 64	0
dense_1 (Dense)	None, 128	8,320
dropout (Dropout)	None, 128	0
flatten_2 (Flatten)	None, 128	0
dense_2 (Dense)	None, 256	33,024
dropout_1 (Dropout)	None, 256	0
flatten_3 (Flatten)	None, 256	0
dense_3 (Dense)	None, 512	131,584
dense_4 (Dense)	None, 8	4,104

5. Result and Discussion

5.1 Performance Evaluation Measures

Medical image analysis uses key metrics measured for its evaluation, such as accuracy, precision, sensitivity, specificity and F1-score [25].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (3)$$

$$Specificity = \frac{TN}{(TN + FP)} \quad (4)$$

$$F1 \text{ score} = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (5)$$

The accuracy calculates the overall authenticity of the model. Precision calculates low false positives (FP). Sensitivity calculates low false negatives (FN). Specificity calculates how many actual negatives were accurately predicted. Calculating F1 score is useful for imbalanced datasets.

5.2 Performance Comparison

The Kvasir dataset is hosted on Kaggle and is a freely available multi-class image dataset for gastrointestinal diseases. It is trained on three different CNN architectures (ResNet-50, VGG-16, and

DenseNet) that are mainly for image classification purposes. The coding was run on Google Colab, which utilizes the Python programming language to write code. The models are trained on different classes of endoscopic images; these images cover polyps, esophagitis, ulcers and normal intestinal view.

Deep convolutional neural networks architecture (VGG) by Visual Geometry Group have become popular for image classification tasks due to their many layers. The word 'deep' refers to how many layers make up the network VGG-16 has 16 convolutional layers and VGG-19 contains 19, by increasing depth, accuracy will increase. ResNet is a type of artificial neural network that uses an identity shortcut connection, which allows it to bypass one or more layers in the network during training. This concept helps to eliminate vanishing gradients and improves the ability to train networks with many more layers than previously possible due to the use of the identity shortcut. Consequently, ResNet has improved the performance of neural networks on challenging computer vision problems. DenseNet is a deeper and more sophisticated convolutional model. The architecture consists of dense blocks that connect each layer in a feed-forward connection. This provides benefits including feature reuse, parameter efficiency and implicit deep supervision. DenseNet minimizes the effects of the vanishing gradient problem. The following illustration is an overview of the evaluation process conducted on the endoscopic image dataset.

Table 3 shows the comparative study on ResNet 50, VGG 16 and DenseNet by computing averages of

accuracy, precision, recall, specificity and F1 score. The comparison depicts that the VGG-16 and DenseNet are the most appropriate models, with an average accuracy of 0.7907 and 0.7784, respectively on endoscopic image dataset [26, 27].

The table below presents the readings of the proposed model. The performance is calculated over evaluation parameters: accuracy, precision, sensitivity (Recall), specificity and F1-score [28, 29].

Table 4, during the assessment of the model as proposed. The top accuracy achieved was for the dyed-lifted polyps (0.8294), normal z-line (0.8206) and ulcerative-colitis (0.8263). There are multiple processes available to improve the accuracy of the model by leveraging hyperparameter tuning, data augmentation, and optimization. A technique for validation used by AI models is cross-validation [30], which works to validate the performance by splitting the dataset into subsets. Once the model is trained, it is possible to see how well

it performed on the data it has seen and not seen without overfitting/underfitting. Cross-validation outputs better results because it reduces bias and variance by splitting the dataset. Data augmentation [31] is a methodology used to create synthetic samples to improve generalization based on the sample data. After data augmentation, the evaluation of all 48,000 endoscopic images was performed the hyper-explore the hyperparameters. The images are now at less risk of overfitting. The weighted loss function is used to handle class imbalance. Performance metrics: Mean ± Standard Deviation.

The proposed model achieved an average accuracy of 0.8656 ± 0.0347 over five independent runs with different random seeds. Table 5 shows the variations in the accuracy rate and figure 4 depicts confusion matrix. The confusion matrix demonstrated strong diagonal dominance with limited inter-class misclassification.

Table 3. shows the average performance comparison of different models on endoscopic classes

Model	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. Specificity	Avg. F1 score
VGG 16	0.7907	0.1992	0.1344	0.8766	0.1375
ResNet 50	0.7688	0.1250	0.1429	0.8750	0.2222
DenseNet	0.7784	0.1132	0.1131	0.8734	0.1136

Table 4. Analysis of the proposed model on endoscopic images

Disease classification	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1 score
Dyed-lifted-polyps	0.8294	0.1524	0.0800	0.9364	0.1049
Dyed-resection-margins	0.7375	0.1233	0.1800	0.8171	0.1463
Esophagitis	0.7331	0.0903	0.1250	0.8200	0.1048
Normal-cecum	0.7762	0.1473	0.1650	0.8636	0.1557
Normal-pylorus	0.7750	0.1491	0.1700	0.8614	0.1589
Normal-z-line	0.8206	0.0935	0.0500	0.9307	0.0651
Polyps	0.7406	0.1004	0.1350	0.8271	0.1151
Ulcerative-colitis	0.8263	0.1020	0.0500	0.9371	0.0671

The figures 5 and 6, shows the accuracy and loss plot. The train/validation loss is calculated to predict defects like overfitting and underfitting. Ideally, a model should stop training if validation accuracy stops improving or validation loss starts increasing. The train accuracy learns from patterns in the training data and validation shows how a model can work with unseen data.

5.3 Comparative Analysis

The designed model called as Endo-CNN has been compared against DL models VGG 16, ResNet 50

and DenseNet on the same dataset to analyse their performances and differences among them; therefore, the overall model performances will be improved through creating an architecture with deeper feature representations and having multiple dense layers while properly utilizing regularization techniques. Dense layers can work well for capturing more complex decision boundaries. Kvasir dataset containing endoscopic images is balanced, hence to study performance one can study the combination of accuracy and F1 score.

Table 6 shows the new model Endo-CNN accuracy comparison with other pre-trained models and

it was found to have positive results for all the intestinal endoscopic images.

Table 5. Variations in the accuracy rate

Disease classification	Improved Accuracy
Dyed-lifted-polyps	0.8950
Dyed-resection-margins	0.8250
Esophagitis	0.8200
Normal-cecum	0.8700
Normal-pylorus	0.8650
Normal-z-line	0.9100
Polyps	0.8400
Ulcerative-colitis	0.9000

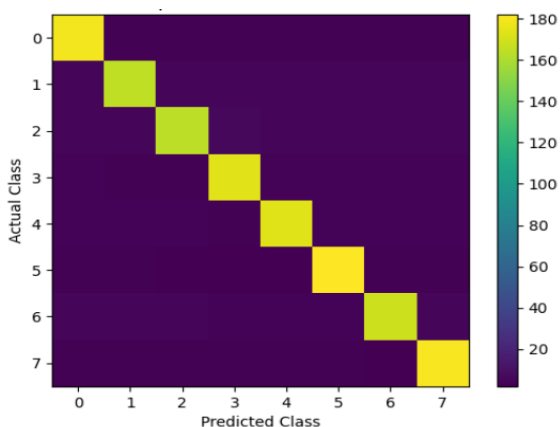


Figure 4. Confusion matrix

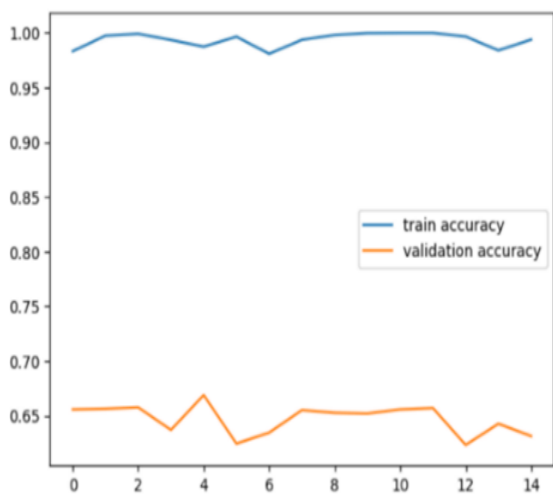
The Endo-CNN model was created to address the shortcomings of ordinary pretrained models like VGG16 and ResNet50, both of which have concerns with how well they can handle the unique issues of the gastrointestinal endoscopic images, such as low inter-

class variance, high intra-class variance, and visual artifacts. The architecture is collectively optimized for medical imaging using a variety of lightweight convolutional layers (6.1M parameters), specific feature-extraction filters (32, 64, 128), and regularization techniques (i.e., dropout, batch normalization) to combat overfitting. Extensive data augmentation was used to enhance generalization methods even more. The domain-specific changes made to this framework allowed Endo-CNN to maintain increasingly improved accuracy (up to 91%) and sensitivity, delivering an effective and trustworthy diagnostic aid for detection of gastrointestinal diseases.

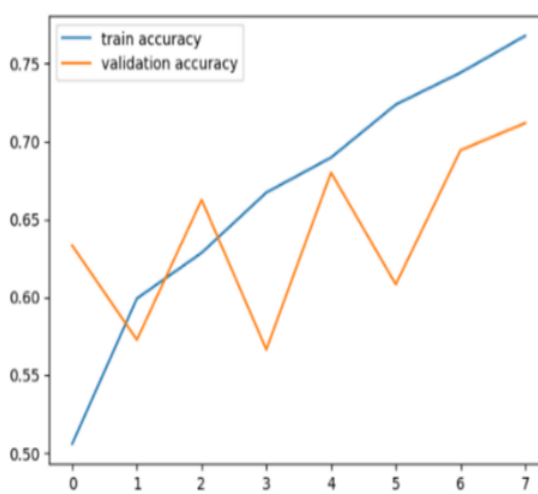
The graph shown in figure 7 demonstrates the comparative accuracy of the four deep learning models of VGG16, ResNet50, DenseNet, and the proposed Endo-CNN model across eight disease classification categories. Each bar represents the accuracy of a model for one specific disease class while showing differences in disease class predictiveness. Although the standard architectures (VGG16, ResNet50, and DenseNet) show varying degrees of strengths and weaknesses in classification prediction, the accuracy of the Endo-CNN remains higher or equal to others, especially for critical disease classes such as Dyed-lifted-polyps, Normal-z-line, and Ulcerative-colitis. This graph demonstrates that the Endo-CNN shows robustness in the prediction of several gastrointestinal disease patterns and may have implications for reliable diagnostic assistance capabilities in the clinical setting.

6. Endoscopic Imaging Challenges and Future Directions

Endoscopic imaging is challenging for researchers to use to investigate the intestinal lining.



(a)



(b)

Figure 5. Train and validation accuracy plot (a) before & (b) after cross-validation for the proposed Architecture

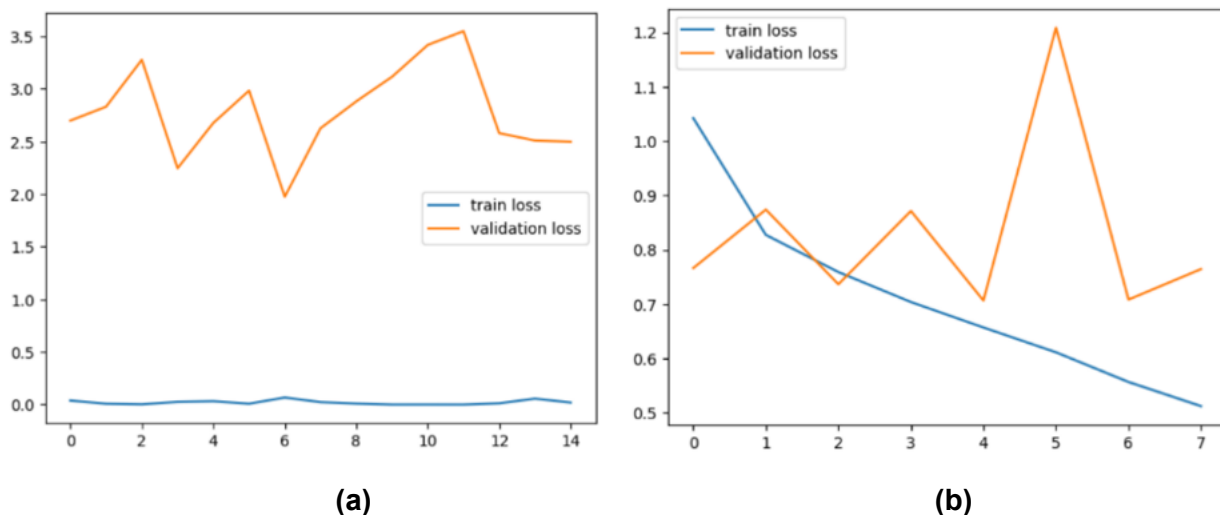


Figure 6. Train and validation loss plot (a) before & (b) after cross-validation for the proposed architecture

Table 6. Comparative study of the models

Disease classification	Accuracy (VGG 16)	Accuracy (ResNet 50)	Accuracy (DenseNet)	Accuracy (Endo-CNN)
Dyed-lifted-polyps	0.8750	0.8750	0.7987	0.8950
Dyed-resection-margins	0.4275	0.8750	0.7725	0.8250
Esophagitis	0.7056	0.8750	0.7881	0.8200
Normal-cecum	0.8750	0.8750	0.7694	0.8700
Normal-pylorus	0.7594	0.1250	0.7625	0.8650
Normal-z-line	0.8750	0.8750	0.7738	0.9100
Polyps	0.8750	0.8750	0.7900	0.8400
Ulcerative-colitis	0.8750	0.8750	0.7725	0.9000

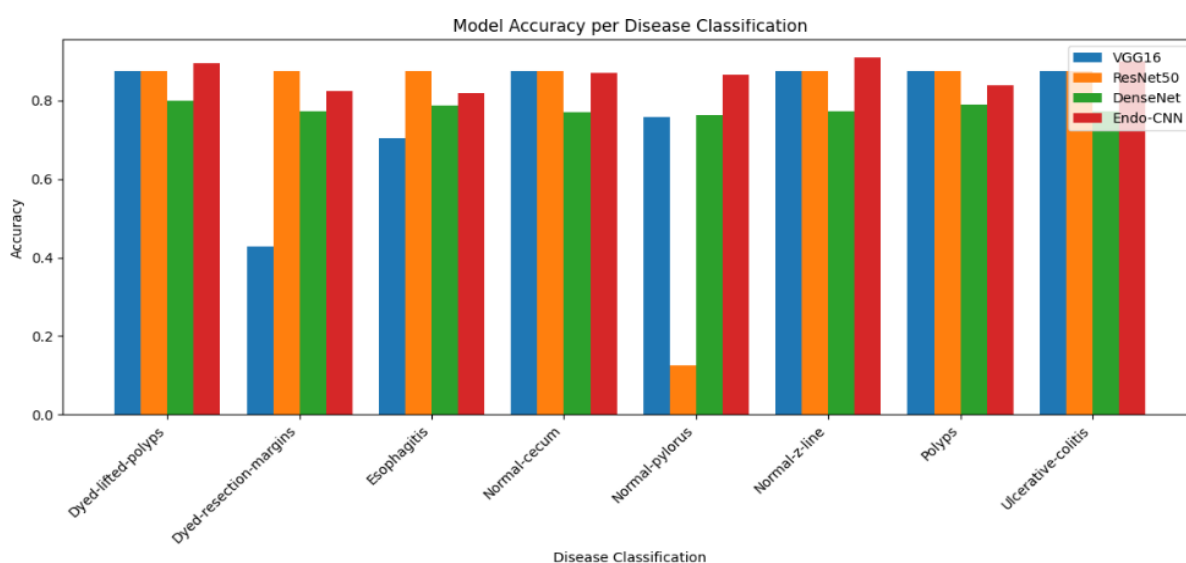


Figure 7. Graph showing variation in the accuracy for different DL models and the proposed model

There are many tasks to understand before proceeding with the research [32, 33]. (a) Availability of

large endoscopic image datasets and understanding of diseased image. (b) Light variations over different parts

of the organ. (c) Difficulty in capturing perfect images when bubbles or food traces appear during the procedure. (d) The narrow intestine creates difficulty in capturing images. Image resolution depends on camera quality. (e) Study different image modalities and different endoscopic images have different sample sizes. (f) Reduce the endoscopy time, optimizing hyperparameters and evaluating network. (g) performance Bridging the gap between doctors and clinical departments. (h) Standardized, Large-Scale Datasets: Building open access repositories with labeled endoscopic images.

Enhancements in future advancements of AI in endoscopy will include many different complementary areas that will support improved diagnostic performance and increased applicability in the clinic going forward. To reduce artifacts from image acquisition, ongoing or new techniques using AI-based preprocessing techniques will be implemented to eliminate air bubbles and food remnants and unevenly lit areas, improving image clarity and thus improving the accuracy of downstream models. Use of higher resolution imaging devices (new capabilities in next generation capsule endoscopes with better sensors) will also improve the extraction of features and viewing of lesions for downstream models. Moreover, techniques such as transfer learning and ensemble modelling will reduce the need for large, annotated datasets while improving models' robustness and generalization ability, allowing for better predictions with less reliance on large datasets. Additionally, real-time AI assistance systems are an important step forward; AI models will be able to assist clinicians in real-time during live endoscopic procedures, allowing for prompt clinical decisions during the procedure. Interdisciplinary collaborations between GI clinicians, AI researchers and Biomedical Engineers will be required to develop solutions that are both clinically applicable and technically feasible. [34] An improvement in image quality correlates directly with model training and performance. Optimizing for different types of data modalities will also improve the accuracy of model architecture. Finally, the use of Explainable Artificial Intelligence (XAI) techniques [35] will help overcome the "black box" issue of CNNs, providing clinicians access to visualization and optimization tools that support transparency and trust in clinical decisions.

7. Conclusion

Most CNN designs include two convolutional layers followed by max pooling, one flatten layer, one fully connected (dense) layer, and one softmax layer. As a result, architectures are usually relatively simple and computationally efficient. In contrast, the proposed design here has many more fully connected layers (64, 128, 256 and 512 neurons) and includes dropout regularization. Thus, a major difference between the two designs is the much greater capacity of the designed

system compared to a basic CNN. The overall parameter size represents one of the most important characteristics of the designed model. The more complex the model, the greater the amount of computational resources and training time needed for the model. The design of the model is biased toward the classification at the expense of parameter size and increased risk of overfitting due to, primarily, the additional flatten layers in the completely connected layer as a result of dropout. The proposed architecture can be scaled to a larger, more complex endoscopic image dataset. Combining different imaging methods, it may be possible to improve model accuracy. Technology and healthcare advancements travel a companion road. The process of development inherently attempts to make incremental improvements to existing models and new methodologies. This manuscript proposes an intestinal endoscopic images model (Endo-CNN) based on a basic convolutional neural network architecture. The construction of the architecture, illustrated pre-processing and deep neural network architecture was significant. The impetus for creating the proposed Endo-CNN architecture stems from the specific issues that arise in endoscopic image classification, which are not effectively solved with generic pre-trained models. The accuracy rate, reported as analyses compared to other DL models (VGG 16, ResNet 50 and DenseNet) achieved a good characterised accuracy rate of 0.9100.

References

- [1] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland, I. Balasingham, Automatic Colon Polyp Detection using Region based Deep CNN and Post Learning Approaches. *IEEE Access*, 6, (2018) 40950–40962. <https://doi.org/10.1109/access.2018.2856402>
- [2] E. Saxena, M. Yadav, M. Yadav, P. Shoran, Artificial Intelligence-based Diagnostic Analysis for Wireless Capsule Endoscopy in Obscure Bowel Disease Detection: a Potential. In *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*, (2022) 1–7. <https://doi.org/10.1145/3590837.3590840>
- [3] H. Malik, A. Naeem, A. Sadeghi-Niaraki, R.A. Naqvi, S. Lee, Multi-Classification Deep Learning models for Detection of Ulcerative Colitis, Polyps, and Dyed-Lifted Polyps using Wireless Capsule Endoscopy Images, *Complex & Intelligent Systems*, 10(2), (2023) 2477–2497. <https://doi.org/10.1007/s40747-023-01271-5>
- [4] M. Yamada, Y. Saito, H. Imaoka, M. Saiko, S. Yamada, H. Kondo, H. Takamaru, T. Sakamoto, J. Sese, A. Kuchiba, T. Shibata, R. Hamamoto, Development of a Real-Time Endoscopic Image Diagnosis Support System using Deep Learning Technology in Colonoscopy. *Scientific Reports*,

- 9(1), (2019) 14465.
<https://doi.org/10.1038/s41598-019-50567-5>
- [5] I. Sensoy, A review on the Food Digestion in the Digestive Tract and the used in Vitro Models. *Current Research in Food Science*, 4, (2021) 308–319.
<https://doi.org/10.1016/j.crfs.2021.04.004>
- [6] S. Sunitha, S. S. Sujatha, Combined feature learning and CNN for Polyp Detection in Wireless Capsule Endoscopy Images. *International Journal of Engineering Trends and Technology*, 69(6), (2021) 206–215.
<https://doi.org/10.14445/22315381/ijett-v69i6p230>
- [7] N. Karthika, B. Janet, H. Shukla, A Novel Deep Neural Network model for Image Classification. *International Journal of Engineering and Advanced Technology*, 8(6), (2019) 3241–3249.
<https://doi.org/10.35940/ijeat.f8832.088619>
- [8] A. Chahal, P. Gulia, Machine learning and deep Learning. *International Journal of Innovative Technology and Exploring Engineering*, 8(12), (2019) 4910–4914.
<https://doi.org/10.35940/ijitee.I3550.1081219>
- [9] I. H. Sarker, Deep learning: a Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions, *SN Computer Science*, 2(6), (2021) 420.
<https://doi.org/10.1007/s42979-021-00815-1>
- [10] E.H. Houssein, D. A. Abdelkareem, G. Hu, M.A. Hameed, I. A. Ibrahim, M. Younan, An effective Multiclass Skin Cancer Classification Approach based on Deep Convolutional Neural Network, *Cluster Computing*, 27(9), (2024) 12799–12819.
<https://doi.org/10.1007/s10586-024-04540-1>
- [11] T. Song, X. Yu, S. Yu, Z. Ren, Y. Qu, Feature Extraction Processing Method of Medical Image Fusion based on Neural Network Algorithm, *Complexity*, 2021(1), (2021).
<https://doi.org/10.1155/2021/7523513>
- [12] H.E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, T. Ganslandt, Transfer learning for Medical Image Classification: a Literature Review, *BMC Medical Imaging*, 22(1), (2022) 69. <https://doi.org/10.1186/s12880-022-00793-7>
- [13] M.M. Auzine, M.H. Khan, S. Baichoo, N.G. Sahib, P. Bissoonauth-Daiboo, X. Gao, Z. Heetun, Development of an Ensemble CNN Model with explainable AI for the Classification of Gastrointestinal Cancer, *PLoS ONE*, 19(6), (2024) e0305628.
<https://doi.org/10.1371/journal.pone.0305628>
- [14] J. Yogapriya, V. Chandran, M.G. Sumithra, P. Anitha, P. Jenopaul, C.S.G. Dhas, Gastrointestinal Tract Disease Classification from Wireless Endoscopy Images using Pretrained Deep Learning Model. *Computational and Mathematical Methods in Medicine*, (2021) 1–12.
<https://doi.org/10.1155/2021/5940433>
- [15] M.N. Noor, M. Nazir, S.A. Khan, O. Song, I. Ashraf, Efficient Gastrointestinal Disease Classification using Pretrained deep Convolutional Neural Network, *Electronics*, 12(7), (2023) 1557.
<https://doi.org/10.3390/electronics12071557>
- [16] S. Mohapatra, P.S. Jeji, G.K. Pati, M. Mishra, T. Swarnkar, Comparative Exploration of Deep Convolutional Neural Networks using Real-Time Endoscopy Images, *Biomedical Technology*, 8, (2024) 1–16.
<https://doi.org/10.1016/j.bmt.2024.09.003>
- [17] H. Taherdoost, Deep learning and Neural Networks: decision-making Implications. *Symmetry*, 15(9), (2023) 1723.
<https://doi.org/10.3390/sym15091723>
- [18] M. Bordbar, M.S. Helfroush, H. Danyali, F. Ejtehadi, Wireless Capsule Endoscopy Multiclass Classification using Three-Dimensional Deep Convolutional Neural Network Model, *BioMedical Engineering OnLine*, 22(1), (2023) 124.
<https://doi.org/10.1186/s12938-023-01186-9>
- [19] D. Marin-Santos, J.A. Contreras-Fernandez, I. Perez-Borrero, H. Pallares-Manrique, M.E. Gegundez-Arias, Automatic Detection of Crohn disease in Wireless Capsule Endoscopic Images using a Deep Convolutional Neural Network. *Applied Intelligence*, 53(10), (2022) 12632–12646. <https://doi.org/10.1007/s10489-022-04146-3>
- [20] J.A. Ayeni, Convolutional neural network (CNN): the Architecture and Applications. *Applied Journal of Physical Science*, 4(4), (2022)42–50.
<https://doi.org/10.31248/ajps2022.085>
- [21] F. Rustam, M.A. Siddique, H.U.R. Siddiqui, S. Ullah, A. Mehmood, I. Ashraf, G.S. Choi, Wireless Capsule Endoscopy Bleeding Images Classification using CNN based model. *IEEE Access*, 9, (2021) 33675–33688.
<https://doi.org/10.1109/access.2021.3061592>
- [22] M.A. Elmagzoub, S. Kaur, S. Gupta, A. Rajab, K.D. Rajab, M.S.A. Reshan, H. Alshahrani, A. Shaikh, Improving Endoscopic Image Analysis: attention Mechanism Integration in grid search fine-Tuned Transfer Learning model for multi-class Gastrointestinal Disease Classification. *IEEE Access*, 12, (2024) 80345–80358.
<https://doi.org/10.1109/access.2024.3408224>
- [23] J. Choi, K. Shin, J. Jung, H. Bae, D.H. Kim, J. Byeon, N. Kim, Convolutional neural network technology in endoscopic imaging: artificial intelligence for endoscopy. *Clinical Endoscopy*, 53(2), (2020) 117–126.
<https://doi.org/10.5946/ce.2020.054>

- [24] P. Kaur, R. Kumar, Performance analysis of convolutional neural network Architectures over Wireless Capsule Endoscopy Dataset, *Bulletin of Electrical Engineering and Informatics*, 13(1), (2023)312–319. <https://doi.org/10.11591/eei.v13i1.5858>
- [25] X. Su, Q. Liu, X. Gao, L. Ma, Evaluation of Deep Learning Methods for early Gastric Cancer Detection using Gastroscopic Images, *Technology and Health Care*, 31(S1), (2023) 313–322. <https://doi.org/10.3233/thc-236027>
- [26] O.F. Ahmad, L.B. Lovat, Artificial intelligence for Colorectal Polyp Detection: are we ready for prime time? *Journal of Medical Artificial Intelligence*, 2, (2019) 16. <https://doi.org/10.21037/jmai.2019.09.02>
- [27] B. Jain, P. Pawar, D. Gada, T. Patwa, P. Kanani, D. Patil, L. Kurup, Comprehensive analysis of Machine Learning and Deep Learning Models on Prompt Injection Classification using Natural Language Processing Techniques, *International Research Journal of Multidisciplinary Technovation*, (2025) 24–37. <https://doi.org/10.54392/irjmt2523>
- [28] S. M. Ganie, P. K. D. Pramanik, A comparative Analysis of Boosting Algorithms for Chronic Liver Disease Prediction, *Healthcare Analytics*, 5,(2024)100313. <https://doi.org/10.1016/j.health.2024.100313>
- [29] T. J. Bradshaw, Z. Huemann, J. Hu, A. Rahmim, A guide to Cross-Validation for Artificial Intelligence in Medical Imaging, *Radiology: Artificial Intelligence*, 5(4), (2023) e220232. <https://doi.org/10.1148/ryai.220232>
- [30] C. Shorten, T. M. Khoshgoftaar, A survey on Image Data Augmentation for Deep Learning, *Journal of Big Data*, 6(1), (2019). <https://doi.org/10.1186/s40537-019-0197-0>
- [31] M. Yamada, R. Shino, H. Kondo, S. Yamada, H. Takamaru, T. Sakamoto, P. Bhandari, H. Imaoka, A. Kuchiba, T. Shibata, Y. Saito, R. Hamamoto, Robust Automated Prediction of the Revised Vienna classification in Colonoscopy using Deep Learning: Development and Initial External Validation, *Journal of Gastroenterology*, 57(11), (2022) 879–889. <https://doi.org/10.1007/s00535-022-01908-1>
- [32] E. Saxena, S. Parveen, M. A. Ahad, M. Yadav, Unveiling the potential of AI in Gastroenterology: Challenges and Opportunities, In *Algorithms for Intelligent Systems*, (2024) 103–114. https://doi.org/10.1007/978-981-97-4533-3_8
- [33] S. Ali, Where do we stand in AI for Endoscopic Image Analysis? Deciphering Gaps and Future directions, *NPJ Digital Medicine*, 5(1), (2022) 184. <https://doi.org/10.1038/s41746-022-00733-3>
- [34] M.M. Auzine, M. H. Khan, S. Baichoo, N. G. Sahib, P. Bissoonauth-Daiboo, X. Gao, Z. Heetun, Development of an Ensemble CNN Model with Explainable AI for the Classification of Gastrointestinal Cancer, *PLoS ONE*, 19(6), (2024)e0305628. <https://doi.org/10.1371/journal.pone.0305628>
- [35] D. Mukhtorov, M. Rakhmonova, S. Muksimova, Y. Cho, Endoscopic Image Classification based on Explainable Deep Learning, *Sensors*, 23(6), (2023) 3176. <https://doi.org/10.3390/s23063176>

Acknowledgement

The authors would like to thank the School of Engineering Sciences and Technology, Jamia Hamdard (Hamdard University), New Delhi, India, for supporting this research.

Authors Contribution Statement

Esha Saxena: Methodology, Conceptualization, Writing - Original Draft. Suraiya Parveen: Methodology, Investigation, Writing - Review & Editing. Mohd. Abdul Ahad: Investigation, Formal analysis, Writing - Review & Editing. Meenakshi Yadav: Methodology, Formal Analysis, Writing - Review & Editing. All the authors have read and agreed to the published version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Has this article screened for similarity?

Yes

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

About the License

© The Author(s) 2026. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License.