Intestinal disorders categorization in endoscopic images using deep learning architectures

Esha Saxena[1], Suraiya Parveen[1], Mohd. Abdul Ahad[1], Meenakshi Yadav[2], Mohammad Anas[1]
[1]Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard (Government Aided Deemed to be University), New Delhi, India
[2]Galgotias College of Engineering and Technology, Greater Noida, India

Article Info

Article history:

Received Mar 5, 2025
Revised Feb 27, 2026
Accepted Apr 20, 2026

Keywords:

Convolutional neural network
Data preprocessing
Deep learning
Endoscopic image
Gastrointestinal tract

ABSTRACT

Gastroenterology is revolutionized by advancements in artificial intelligence (AI). As the gastrointestinal (GI) tract is consulted, globally 40% of the world's and 18% of the Indian population are affected. AI is a reliable sword for diagnosing issues related to the GI tract. The learning capabilities of deep learning (DL) techniques make it widely helpful in medical investigations. The variety of data available in the medical sector generates the need for an appropriate model for every problem domain. The purpose of this research is to explore the significance of medical image pre-processing and the implementation of pre-trained DL models on endoscopic images for the diagnosis of disease. Convolutional neural network (CNN)-based architectures have robust diagnostic potential for medical images. It can assist physicians as a tool for disease analysis, screening and help in investigating further needs. The paper also provides a comparative performance framework showing CNN architectures and preprocessing techniques for endoscopic images to highlight the key points important for investigating GI tract related diseases. The endoscopic images were trained over VGG-16, ResNet-50 and DenseNet-121, DL models. The result suggests that VGG-16 and ResNet-50 gave promising results with an accuracy maximum of 87.50%.

Corresponding Author:

Esha Saxena
Department of Computer Science and Engineering, School of Engineering Sciences and Technology
Jamia Hamdard (Government Aided Deemed to be University)
New Delhi, India
Email: eshasaxena.1987@gmail.com

1. INTRODUCTION

Endoscopic imaging and interpretation a key diagnostic evaluation processes, and particularly for diseases of the gastrointestinal (GI) tract. One of the current areas of concern in India is GI tract diseases. The trend in intestinal disorders is increasing due to unhealthy lifestyles. Human intestinal diseases could become difficult to evaluate and manage subsequently, based on the nature of the data collected [1]. This gown processes of investigation of aspects have progressed considerably within a short period owing to technology, and deep learning (DL) could form a considerable leap forward for the evaluation of GI diseases and help doctors and researchers to identify the disease. DL has established itself in healthcare in a predominant position, but there have been a number of more recent developments within DL and this has been an area of capture within the healthcare domain.

DL and healthcare use together can form an interesting and powerful relationship for interoperability, and it can serve the overall healthcare process well [2]. DL has impacted numerous areas in

United State (US), causing considerable anticipation in their working, but in healthcare, especially in medicine, DL has gained considerable attention and reported promising results. Artificial intelligence (AI) is seen as the fourth industrial revolution (4.0). Feature extraction and prediction throughout medicine, image recognition, robotics, cybersecurity, smart agriculture, business intelligence, genetics, games, and more recently, have significance in many areas and across fields. DL approaches can be in the form of models, including convolutional neural network (CNN), autoencoders (AE), deep belief networks (DBN), recurrent neural networks (RNNs), and hybrid deep neural network (DNN), to name a few. There are various studies and combinations of solutions can be proposed across supervised and unsupervised datasets. Each field will have research issues and fields according to the applications of DL. DL being applied in engineering, technology and similar areas of industry is dominating across the disciplinary fields.

DL has made its way too many domains such as medical studies, text recognition, and image classification. DL is based on neural networks [3]. The learning methods are supervised, unsupervised, and reinforcement learning. The architecture consists of various architectures such as CNN, AE, DBNs, stack autoencoders (SAEs), DNNs, RNNs, and recurrent Boltzmann machines (RBM) [4]. In DL models, large datasets work effectively for automated performance of the features extracted. AI has taken a considerable part of the research in the area of healthcare, especially regarding intelligent systems. The possibilities for better process workflows are made from the power of DL and medical techniques. New advances have made details faster and easier.

Supervised learning is a method where the model derives its output from previously labeled dataset, which is a more logical approach for problems, as it utilizes knowledge gained from an earlier labeled dataset. It encompasses some common algorithms, including CNNs, DNNs, RNNs, and feedforward neural networks (FNNs), all of which lend themselves well to image classification methods. Unsupervised learning emerges from unlabeled datasets, in addition to having no previously provided knowledge or labeled data. It incorporates live models named AE, Boltzmann machines, and generative adversarial networks (GANs), generally used in anomaly detection and dimensionality reduction type problems [5]. Semi-supervised learning typically occurs when datasets are small and provides a method to learn from semi-labeled datasets. It produces a lower requirement of labeled data and shows beneficial performance when labeled data is unavailable. The deep reinforcement learning method uses DL with reinforcement methods, leading to better decision-making for real-time use. To future increase accuracy, it is critical to observe the specified dataset to get the desired output. Recently, it has provided a second screen for doctors.

There are several neural networks that showed positive results for medical image classification. The neural network that is inspired by a human neuron is focused on passing information from neuron to neuron. It is a structured network of layers where information flows from the input layer to the output layer through hidden layers in between. CNN [6] is most popular DL model for image classification. It is best-performing model for image classification and recognition, particularly with 2D images [7]. CNN consists of convolutional, fully connected and pooling layers. It is sometimes challenging to identify abnormal features in the intestine because endoscopic quality may be limited by the narrow diameter of the tube, poor orientation, or sometimes low light due to the free capsule movement or bubbles in the intestine. CNN models can perform with sufficient performance for disease classification, including ulcers, polyps, celiac disease, and bleeding [8].

In still image classification, we can automatically determine whether an abnormal area is present using classification and feature extractions techniques [9]. Endoscopic image classification and feature extraction typically need specifically labeled datasets. The CNN systems have many different architectures for image classification problems. Table 1 lists some major CNN architectures [10], [11] with specific features. The DL models ResNet, VGG, and DenseNet are some of the most popular architectures in the computer vision space. They have been successfully modified to perform analyses of medical images, such as endoscopy, radiology, and histopathology datasets.

Table 1. CNN architectures

Models	Features
ResNet	The main concept is the residual connection (skip connection), which authorize the network to learn residual mappings instead of direct mappings. This alleviates the vanishing/exploding gradient problem and ultimately allows training of extremely deep models. As a result, ResNets rapidly became the backbone for tasks like classification, detection, and segmentation. The most used models are ResNet50, ResNet101 and ResNet152.
VGG	It's a deep CNN. VGG is named as the Visual Geometry Group. Adding depth increases accuracy significantly on ImageNet. Showed that using small filters in combination can approximate larger receptive fields with fewer parameters. It can explore complex patterns from an image with its two popular models, VGG-16 and VGG-19.
DenseNet	It's a densely connected convolutional network with a feed-forward CNN, in which every layer links with other layers. It consists of transition layers and dense blocks. The architecture uses bottleneck layers that help minimize parameters without reducing the features learned by the network. The most used models are DenseNet-121 and DenseNet-201.

2. GASTROENTEROLOGY

Gastroenterology is a branch that deals with research and treatment concerning the digestive system in humans. It is a broad area of study, especially in trying to identify abnormalities such as polyps or ulcers of the GI tract that may have otherwise been missed. Since the field requires heavy reliance upon observation of the GI tract, the incorporation of AI has led to improvements in the diagnosis [12]. AI has revolutionized the healthcare industry by introducing breakthrough technologies that have replaced labor-intensive traditional tasks. Large volumes of medical data, such as patient reports and medical imaging, were considered difficult to manage by practitioners. DL plays a critical role in identifying GI tract disorders and shortening the time between diagnosis and treatment. The effectiveness depends on the type and capacity of the dataset, where models can learn from complex patterns and relations. Their performance varies by various factors, including but not limited to, the availability of labelled data, computational resources, and the selection of the most appropriate model with accuracy in mind. CNNs, a class of DL architecture, have yielded very impressive results in endoscopic image analysis for the detection of lesions and tumors [13]. Figure 1 shows an endoscopic image taken from a video captured by wireless capsule endoscopy (WCE). The image shows shape: (259, 309, 3). It’s a normal view of an intestine without traces of any disease. It is an advanced medical technique to diagnose diseases occurring in internal organs that are tough to detect by other clinical methods or diagnostic tests. The term endoscope was first coined in the year 1850 [14]. This technology adopts either a wired flexible tube or a wireless capsule to capture internal images to assist doctors in spotting abnormal conditions.

WCE represents a major stride in the realm of endoscopy. It is a noninvasive, pain-free procedure wherein the patient swallows a small capsule accommodating a light source and camera to capture the images of the GI tract and other internal organs. The device can examine the digestive, urinary, reproductive, respiratory, and cardiovascular systems, providing a comprehensive internal view of the human body. Gastroenterology, related to the digestive system, covers the mouth, esophagus, stomach, intestines, rectum, and colon [15]. The main purposes of endoscopic examinations in gastroenterology are disease classification, biopsy, tumor detection, and surgical assessment [16]. Wired endoscopy and WCE both allow doctors to view the GI tract for the purpose of diagnosis and treatment planning. Developed in the year 2000, WCE gained popularity quite fast because of its simplicity, safety, and the possibility to obtain detailed diagnostic information. The capsule records high-quality video up to eight hours or more thus making it effective for the detection of intestinal abnormalities [17]. It has remained one of the most informative and patient-friendly means of internal disease investigation [18]. Endoscopic imaging remains a significant research priority in gastroenterology. Integrated with AI and leveraging large, high-quality datasets, endoscopic imaging systems are being designed to predict disease and identify conditions much earlier. It would improve diagnostic precision and contribute to better patient care.

Figure 1. Normal GI endoscopic image

3. RESULT AND DISCUSSION

This section measures evaluation metrics and finds the best-performing model among VGG-16, ResNet-50, and DenseNet. Eight to nine hours of endoscopic video recording were used to get the images.

3.1. Data gathering

The study's dataset was chosen from the openly available online resource Kaggle. A wide range of datasets are added to the website, which provides researchers with an extensive platform. 8,000 endoscopic images from eight different classes are included in the Kavasir-V2 dataset, which was used by us at www.kaggle.com/datasets/yasserhessein/the-kvasir-dataset, such as: dyed-lifted polyps, dyed-resection margins, normal cecum, normal z-line, esophagitis, normal pylorus, polyps, and ulcerative colitis.

3.2. Image preprocessing

Numerous sources of endoscopic images with varying endoscope models, dimensions, and image quality are available across different platforms, and pre-processing methods [19] have been applied in our research before the image classification procedure. An estimated 50% to 80% of the classification process is spent on preprocessing. It can be challenging for an observer to accurately identify the diseased area in endoscopic images because of problems with angle, curves, light, and non-uniformity. The majority of the image is 720 by 576 pixels in size. Images that are superfluous or extremely blurry are removed from the dataset prior to preprocessing. Ignoring such images could compromise accuracy. Some commonly applied preprocessing methods are: noise reduction, dimensionality reduction, contrast enhancement, data augmentation, image sharpening, and color space transformations. Figure 2 shows different image preprocessing techniques applied to a diseased endoscopic image taken from dataset. The different methods come with various identification marks, which help in identifying the issue associated with the image.

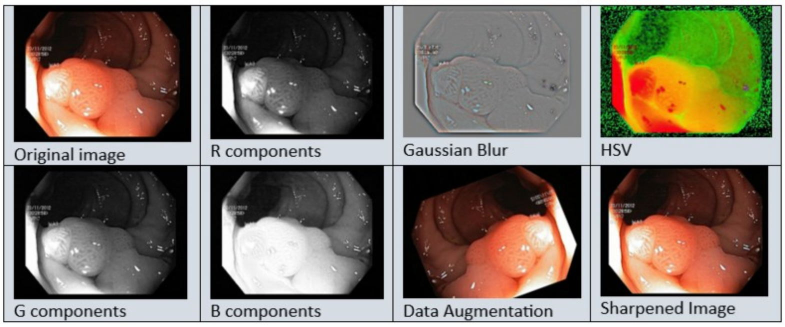

Figure 2. Image preprocessing techniques

3.3. Performance evaluation measures

Medical image analysis systems are commonly assessed with a set of widely accepted performance metrics based on an understanding of their full implications. The key metrics measured for its evaluation are accuracy, precision, sensitivity, specificity, and F1-score [20], [21]. Together, the metrics provide a rigorous framework for assessing and comparing DL models in medical diagnosis.

The mathematical formulations of these metrics are defined as (1) to (5).

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

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

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

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

$$F1\ score = 2 \times \frac{(Precision * Recall)}{(\mathrm{Precision+Recall})} \tag{5}$$

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.

3.4. Performance comparison

The Kvasir dataset is a publicly available dataset on Kaggle and is made up of a multi-class image dataset for endoscopic images. The Google Colab environment holds all of the code in the Python programming language. The models have been tested on endoscopic image classes consisting of polyps, esophagitis, ulcers, and normal intestinal views. The image classification represented is one part of the process of DL for an endoscopic dataset using different CNN networks (ResNet-50, VGG-16, and DenseNet).

i) Input data: the endoscopic image dataset represents raw images of patients undergoing actual endoscopic procedures (for disease diagnosis and tissue assessment).
ii) Data pre-processing: before training the data, cleansing and optimizing the raw images is done using various pre-processing methods like data augmentation, artificially expands the dataset (rotation, flipping, and cropping), noise reduction an image processing artifacts are removed to create clearer images, contrast enhancement enhances and defines key image features, dimensionality reduction

reduces the number of features and the complexity without losing important features, image sharpening defines edges and structures in images [22].

iii) Classification using DL: once the dataset is processed, the images are transferred to the architectures: ResNet-50 (uses residual connections to enable deep networks to be trained effectively), VGG-16 (classic deep network with 3×3 filters), and DenseNet (convolves every layer with prior and subsequent layers to improve feature reuse) for classification process [23].

iv) Performance comparison: outputs from 3 models will be compared based on accuracy/precision/recall, training time/computational efficiency. This identifies the model that performs the best in endoscopic image classification.

Figure 3 is a pictorial representation for an endoscopic dataset performing pre-processing techniques and its effect on model comparison analysis. Table 2 demonstrates the comparison study on pre-trained DL models that have good accuracy levels for endoscopic images, especially for polyps. The table contains readings calculating accuracy, precision, sensitivity (recall), specificity, and F1-score [24]. The augmentation technique doesn't have much effect on the accuracy rate. Based on the above data, we evaluated that the model to best work with the endoscopic dataset must have DNN layers [25], [26]. Figure 4 shows assessment of three DL models (VGG16, ResNet50, and DenseNet) in their classification of eight GI diseases reveals significant differences in performance.

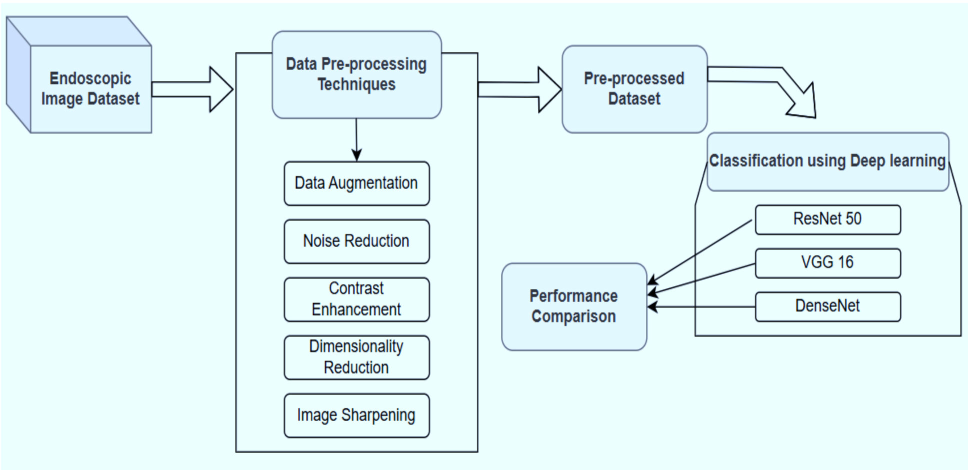

Figure 3. Data preprocessing and model evaluation workflow for the endoscopic dataset

Table 2. The performance comparison of the CNN architectures on endoscopic images

Disease classification	Model	Accuracy	Precision	Sensitivity (recall)	Specificity	F1-score
Dyed-lifted-polyps	VGG-16	0.8750	Nan	0.0000	1.0000	Nan
	ResNet-50	0.8750	Nan	0.0000	1.0000	Nan
	DenseNet	0.7987	0.1573	0.1400	0.8929	0.1481
Dyed-resection-margins	VGG-16	0.4275	0.1324	0.6450	0.3964	0.2198
	ResNet-50	0.8750	Nan	0.0000	1.0000	Nan
	DenseNet	0.7725	0.1204	0.1300	0.8643	0.1250
Esophagitis	VGG-16	0.7056	0.1444	0.2750	0.7671	0.1893
	ResNet-50	0.8750	Nan	0.0000	1.0000	Nan
	DenseNet	0.7881	0.0736	0.0600	0.8921	0.0661
Normal-cecum	VGG-16	0.8750	Nan	0.0000	1.0000	Nan
	ResNet-50	0.8750	Nan	0.0000	1.0000	Nan
	DenseNet	0.7694	0.0957	0.1000	0.8650	0.0978
Normal-pylorus	VGG-16	0.7594	0.1193	0.1450	0.8471	0.1309
	ResNet-50	0.1250	0.1250	1.0000	0.0000	0.2222
	DenseNet	0.7625	0.0794	0.0850	0.8593	0.0821
Normal-z-line	VGG-16	0.8750	Nan	0.0000	1.0000	Nan
	ResNet-50	0.8750	Nan	0.0000	1.0000	Nan
	DenseNet	0.7738	0.1478	0.1700	0.8600	0.1581
Polyps	VGG-16	0.8750	Nan	0.0000	1.0000	Nan
	ResNet-50	0.8750	Nan	0.0000	1.0000	Nan
	DenseNet	0.7900	0.1180	0.1050	0.8879	0.1111
Ulcerative-colitis	VGG-16	0.8750	0.5000	0.0050	0.9993	0.0099
	ResNet-50	0.8750	Nan	0.0000	1.0000	Nan
	DenseNet	0.7725	0.1132	0.1200	0.8657	0.1165

Accuracy: VGG-16 and ResNet-50 show consistent exceptional accuracies (often ~0.875) but in many instances have very low recall, indicating the accuracy is overstated due to class imbalance. Conversely, DenseNet, which showed slightly lower accuracy (~0.76–0.79), provides a more balanced output in every metric. Precision and recall (sensitivity): VGG16 and ResNet50 often report NaN values or near-zero values, indicating a poor ability to identify positives. DenseNet is modest in terms of precision values (~0.07–0.16) and recall values (~0.06–0.17), but does show more consistent detections of the positive class.

Specificity: while all models maintain a good deal of specificity (all >0.85 with many 1.0) and do a good job detecting negatives, specificity comes at the cost of recall for VGG-16 and ResNet-50. F1-score: DenseNet is the only model to provide meaningful F1-scores (0.07~0.16), which reflects the balance between precision and recall. VGG-16 and ResNet-50 often report close to non-existent or undefined values for F1, which validates their lack of sensitivity. Generally, if an issue is defined as low recall and low sensitivity (which it is in many cases), VGG-16 and ResNet-50 are highly functioning in terms of accuracy.

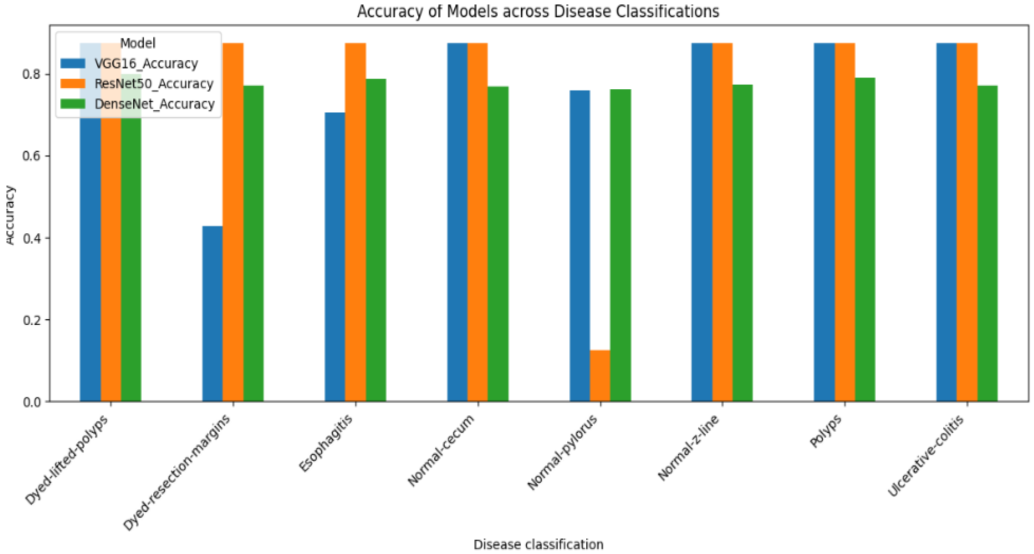

Figure 4. Accuracy variations in VGG-16, ResNet-50, and DenseNet

3.5. Plotting loss and accuracy value

The plotting train loss value indicates how suitably the model fits the training data, while the validation loss computes how well the model generalizes to unseen data. Both values can help detect overfitting and underfitting defects. Figures 5 to 7 represent graphs of train/validation loss vs train/validation accuracy for the three models (VGG-16, Resnet-50, and DenseNet).

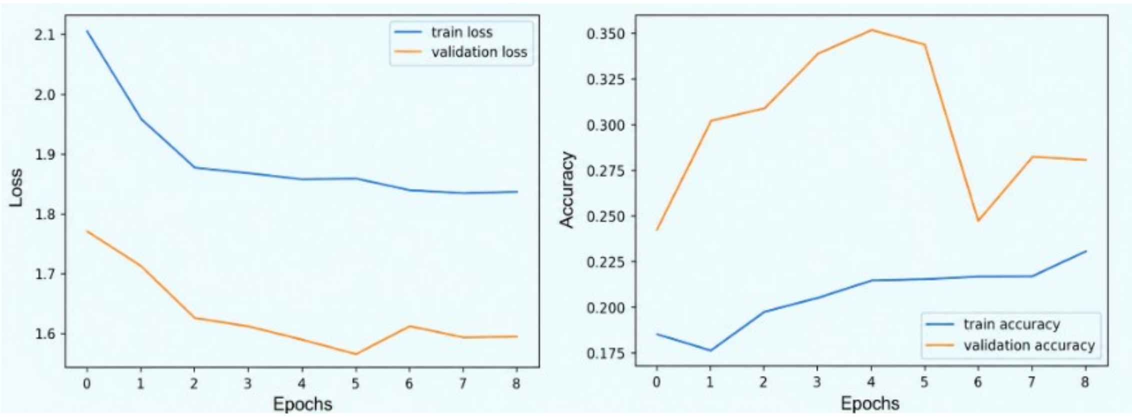

Figure 5. Train/validation loss and train/validation accuracy for VGG-16

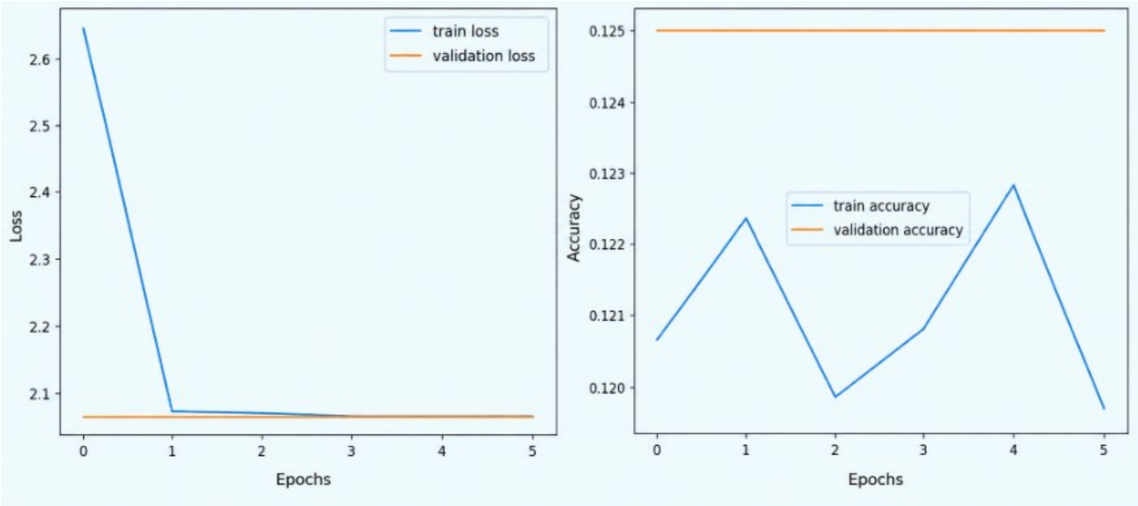

Figure 6. Train/validation loss and train/validation accuracy for ResNet 50

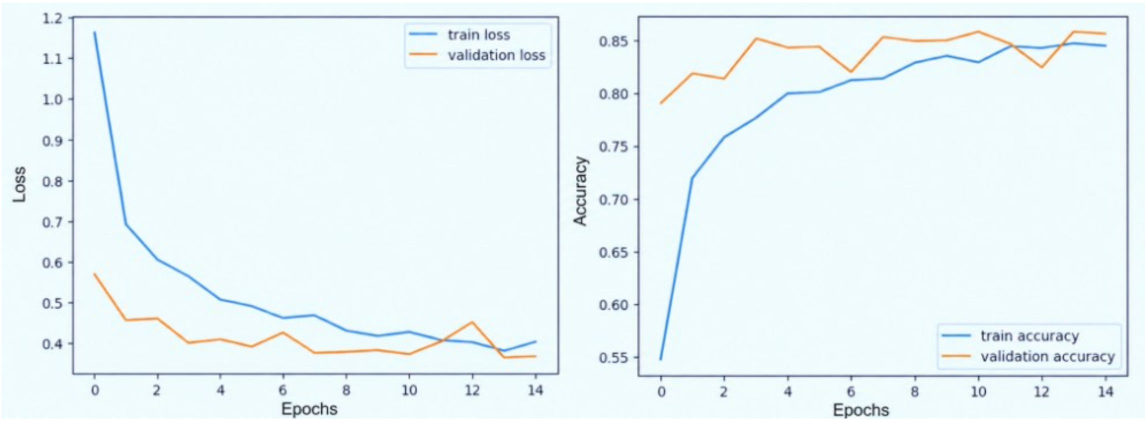

Figure 7. Train/validation loss and train/validation accuracy for DenseNet

Train accuracy and validation accuracy are required to measure performance of ML/DL models [27] during the training process, especially for classification purposes. VGG is good with feature extraction, while ResNet is efficient with large-scale image classification. DenseNet model lowers the redundancy and enhances gradient flow.

4. CONCLUSION

Because of lengthy manual methods, physicians have requested algorithmic approaches to efficiently identify and categorize illnesses. DL models are often effective with large datasets and assist physicians in reducing the amount of time for examination. GI disease identification is reliant on endoscopic image series. It is a pipeline where first, the dataset is pre-processed and then set to use a pre-trained DL model. This study compared three pre-trained models (VGG, ResNet and, DensNet) suitable for medical imaging, identifying VGG-16 and ResNet-50 as performing the best for endoscopic images. The comparative study can also help lead to a proposal for a new model for GI disease identification. CNN architectures showed promising outcomes with the endoscopic images, but results will differ based on the types of investigating images; that is the reason for applying data augmentation to examine each angle of an image. It was shown that pre-processing is beneficial to the classification process. As for future work, an advanced AI model can be constructed to assist researchers/doctors. Clinical practice is an ongoing work and success is dependent on embracing challenges.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Esha Saxena	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Suraiya Parveen		✓			✓	✓		✓	✓	✓		✓		
Mohd. Abdul Ahad	✓	✓				✓	✓		✓	✓	✓	✓	✓	
Meenakshi Yadav		✓		✓	✓	✓			✓	✓		✓	✓	
Mohammad Anas		✓	✓		✓	✓	✓			✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The author declares that there are no known conflicts of interest associated with this publication. There are no financial or personal relationships that could inappropriately influence or bias the content of this work.

INFORMED CONSENT

Not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

ETHICAL APPROVAL

Not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/yasserhessein/the-kvasir-dataset.

REFERENCES

[1] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
[2] J. Egger *et al.*, "Medical deep learning—a systematic meta-review," *Computer Methods and Programs in Biomedicine*, vol. 221, 2022, doi: 10.1016/j.cmpb.2022.106874.
[3] A. A. Isaza, L. M. Jiménez, and M. Z. Diaz, "An overview of deep learning in medical imaging," *Informatics in Medicine Unlocked*, vol. 26, 2021, doi: 10.1016/j.imu.2021.100723.
[4] B. P. Battula and D. Balaganesh, "Medical image data classification using deep learning based hybrid model with CNN and encoder," *Revue d'Intelligence Artificielle*, vol. 34, no. 5, pp. 645–652, 2020, doi: 10.18280/ria.340516.
[5] G. R. Kumar, J. Jeyasudha, G. Ravi, T. Jagadesh, and G. K. Madhura, "Wireless capsule endoscopy bleeding Images classification using CNN based model," in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, 2023, doi: 10.1109/EASCT59475.2023.10393338.
[6] D. T. Nguyen, M. B. Lee, T. D. Pham, G. Batchuluun, M. Arsalan, and K. R. Park, "Enhanced image-based endoscopic pathological site classification using an ensemble of deep learning models," *Sensors*, vol. 20, no. 21, pp. 1–24, 2020, doi: 10.3390/s20215982.
[7] S. Yaroshchak, M. Smaida, and Y. E. Barg, "Medical image enhancement based on convolutional denoising autoencoders and GMD model," in *MoMLeT+DS 2021: 3rd International Workshop on Modern Machine Learning Technologies and Data Science*, 2021, pp. 96–106.

[8] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, “Fundamental concepts of convolutional neural network,” in *Recent trends and advances in artificial intelligence and internet of things*, vol. 172, Cham, Switzerland: Springer, 2020, pp. 519–567, doi: 10.1007/978-3-030-32644-9_36.
[9] D. R. Sarvamangala and R. V. Kulkarni, “Convolutional neural networks in medical image understanding: a survey,” *Evolutionary Intelligence*, vol. 15, no. 1, 2022, doi: 10.1007/s12065-020-00540-3.
[10] L. Jiao and J. Zhao, “A survey on the new generation of deep learning in image processing,” *IEEE Access*, vol. 7, pp. 172231–172263, 2019, doi: 10.1109/ACCESS.2019.2956508.
[11] R. Noorda, A. Nevárez, A. Colomer, V. P. Beltrán, and V. Naranjo, “Automatic evaluation of degree of cleanliness in capsule endoscopy based on a novel CNN architecture,” *Scientific Reports*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-74668-8.
[12] W. Tan *et al.*, “An endoscope image enhancement algorithm based on image decomposition,” *Electronics*, vol. 11, no. 12, 2022, doi: 10.3390/electronics11121909.
[13] N. Ogata, K. Ohtsuka, M. Ogawa, Y. Maeda, F. Ishida, and S. E. Kudo, “Image-enhanced capsule endoscopy improves the identification of small intestinal lesions,” *Diagnostics*, vol. 11, no. 11, 2021, doi: 10.3390/diagnostics11112122.
[14] S. A. Albelwi, “Deep architecture based on DenseNet-121 model for weather image recognition,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 559–565, 2022.
[15] M. S. Ayyaz *et al.*, “Hybrid deep learning model for endoscopic lesion detection and classification using endoscopy videos,” *Diagnostics*, vol. 12, no. 1, 2022, doi: 10.3390/diagnostics12010043.
[16] G. Son *et al.*, “Small bowel detection for wireless capsule endoscopy using convolutional neural networks with temporal filtering,” *Diagnostics*, vol. 12, no. 8, 2022, doi: 10.3390/diagnostics12081858.
[17] V. B. S. Prasath, D. N. H. Thanh, L. T. Thanh, N. Q. San, and S. Dvoenko, “Human visual system consistent model for wireless capsule endoscopy image enhancement and applications,” *Pattern Recognition and Image Analysis*, vol. 30, no. 3, pp. 280–287, 2020, doi: 10.1134/S1054661820030219.
[18] B. Delagah and H. Hassanpour, “Feature extraction for polyp detection in wireless capsule endoscopy video frames,” *Journal of Healthcare Engineering*, vol 2023, no. 1, 2023, doi: 10.1155/2023/6076514.
[19] K. Maharana, S. Mondal, and B. Nemade, “A review: data pre-processing and data augmentation techniques,” *Global Transitions Proceedings*, vol. 3, no. 1, 2022, pp. 91–99, doi: 10.1016/j.gltp.2022.04.020.
[20] E. O. C. Rodriguez *et al.*, “Comparative study of convolutional neural network architectures for gastrointestinal lesions classification,” *PeerJ*, vol. 11, 2023, doi: 10.7717/peerj.14806.
[21] M. B. Haile, A. O. Salau, B. Enyew, and A. J. Belay, “Detection and classification of gastrointestinal disease using convolutional neural network and SVM,” *Cogent Engineering*, vol. 9, no. 1, 2022, doi: 10.1080/23311916.2022.2084878.
[22] J. Yogapriya, V. Chandran, M. G. Sumithra, P. Anitha, P. Jenopaul, and C. S. G. Dhas, “Gastrointestinal tract disease classification from wireless endoscopy images using pretrained deep learning model,” *Computational and Mathematical Methods in Medicine*, 2021, doi: 10.1155/2021/5940433.
[23] W. Sun, P. Li, Y. Liang, Y. Feng, and L. Zhao, “Detection of image artifacts using improved cascade region-based quality assessment of endoscopic images,” *Bioengineering*, vol. 10, no. 11, 2023, doi: 10.3390/bioengineering10111288.
[24] A. Teramoto, T. Shibata, H. Yamada, Y. Hirooka, K. Saito, and H. Fujita, “Detection and characterization of gastric cancer using cascade deep learning model in endoscopic images,” *Diagnostics*, vol. 12, no. 8, 2022, doi: 10.3390/diagnostics12081996.
[25] M. Sharif, M. A. Khan, M. Rashid, M. Yasmin, F. Afza, and U. J. Tanik, “Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images,” *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 33, no. 4, pp. 577–599, 2021, doi: 10.1080/0952813X.2019.1572657.
[26] D. Mukhtorov, M. Rakhmonova, S. Muksimova, and Y. I. Cho, “Endoscopic image classification based on explainable deep learning,” *Sensors*, vol. 23, no. 6, 2023, doi: 10.3390/s23063176.
[27] I. A. Ahmed, E. M. Senan, and H. S. A. Shatnawi, “Hybrid models for endoscopy image analysis for early detection of gastrointestinal diseases based on fused features,” *Diagnostics*, vol. 13, no. 10, 2023, doi: 10.3390/diagnostics13101758.

BIOGRAPHIES OF AUTHORS

Esha Saxena is a research scholar in the Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard (Government Aided Deemed to be University), New Delhi, India. She has done B.Tech. and M.Tech. in Computer Science and Engineering. She has a rich experience of teaching in the field of computer science and engineering for more than 13 years. She has many research papers published in national and international conferences. She has been granted two patents and published two books. Her area of interest includes machine learning and deep learning. She can be contacted at email: eshasaxena.1987@gmail.com.

Suraiya Parveen is currently working in the Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard (Government Aided Deemed to be University), New Delhi, India. She has a rich experience of more than 15 years in the field of computer science and engineering. She has her Bachelor of Engineering degree from Jamia Millia Islamia, New Delhi and her Master of Technology from GGS Indraprastha University, New Delhi. She has completed her Ph.D. in Computer Science and Engineering from Jamia Hamdard (Government Aided Deemed to be University), New Delhi. She has many research papers published in international conferences and journals. Her research interests include information retrieval, machine learning, and semantic technology. She can be contacted at email: suraiya@jamiahamdard.ac.in.

Mohd. Abdul Ahad is currently working in the Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard (Government Aided Deemed to be University), New Delhi, India. He has a rich experience of more than 15 years in the field of computer science and engineering. He obtained his Ph.D. degree in the field of big data architecture. His research areas include big data architecture, distributed computing, IoT, and sustainable computing. He has published several research papers in various international journals of repute. He has chaired several sessions in international conferences of Springer, and Elsevier. He is a certified Microsoft innovative educator and a certified Google educator. He is a life member of Indian Society of Technical Education (ISTE) as well as an active member of IEEE and ACM. He can be contacted at email: itsmeahad@gmail.com.

Meenakshi Yadav is currently working in the Department of Information Technology in Galgotias College of Engineering and Technology, affiliated with AKTU, Uttar Pradesh. She has been awarded a Ph.D. in CSE from Banasthali Vidyapith, Rajasthan. She has done B.E. (Hons.) from Rajasthan University and M.Tech. from Banasthali Vidyapith, Rajasthan. She has many research papers published in national and international journals/conferences. She has granted 3 patents and 1 book on AI. She is having 15 years of teaching experience and her current area of interest are wireless sensor networks, agriculture automation, and machine learning. She can be contacted at email: yadavmeenakshi@gmail.com.

Mohammad Anas is M.Tech. from the Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard (Government Aided Deemed to be University), New Delhi, India. He has published many papers in national and international conferences. His area of interest is LLM and neural networks. He can be contacted at email: anassiddiqui634@gmail.com.