A Domain-specific Multi-layer Convolutional Algorithm for Gastrointestinal Diagnostics

[1] Esha Saxena, [2] Suraiya Parveen, [3] Mohd. Abdul Ahad, [4] Meenakshi Yadav
[1, 2, 3] School of Engineering Sciences & Technology, Jamia Hamdard, 110062, New Delhi, India
[4] Galgotias College of Engineering and Technology, Gr. Noida, 201310, Uttar Pradesh, India

Corresponding Author: Esha Saxena

Abstract: The human body is a highly complex structure, and the presence of disorders further complicates its understanding. Gastrointestinal disorders are one of the issues that affect approximately 40% of the global population. In cases of intestinal disorders, diagnosis can sometimes be challenging, thereby necessitating the use of Wireless Capsule Endoscopy (WCE) for accurate internal observation. Different algorithms have boosted the use of AI and DL for medical imaging. In our research, we propose a DL-based algorithm for the identification and classification of GI tract issues from endoscopic images. The CNN is the base architecture with the highest accuracy rate for medical imaging. The system combines multiple layers into the algorithm using various image pre-processing techniques to enhance the accuracy rate. The algorithm is applied to eight different classes of intestinal diseases and achieves an accuracy of more than 90% for each class. Early detection is a blessing to the patient to giving them time to recover. We aim to narrate the potential of AI, DL techniques for endoscopic investigations and try to provide insights for future research directions through this paper.

Keywords: Deep learning, Endoscopic images, CNN, Gastroenterology,XAI

Introduction

The paper focuses on understanding the application of artificial intelligence methods in gastrointestinal (GI) disorders. None of the fields today is untouched by artificial intelligence (AI) and its capabilities. The GI tract is one of the areas where there is a significant need for AI to streamline the doctor's investigative work process. Medical imaging has been linked with AI tools for many years. According to the healthcare system, AI has revolutionized the world through techniques and intelligent systems designed to minimize decision-making time. Artificial intelligence has shown promising results in enhancing the detection process over endoscopic images. AI improved the diagnosis investigation for many diseases related to the heart, intestine, kidney, lungs, brain etc. ML and deep learning (DL) models provide the most powerful approaches for solutions to be comprehensive, supervised and time-efficient. Medical reports that cover image patterns have maximum data to feed AI models. Deep learning models showed tremendous results in addressing and solving challenges in image processing. Deep learning [1] is primarily based on neural networks. Convolutional neural networks (CNNs) are one of the leading DL algorithms for image analysis. Deep learning techniques demonstrated positive clinical results in multiple medical fields. Images are treated as live data for DL models. Deep learning faces many challenges for different tasks. Some most applied DL-based models are deep neural networks (DNN), convolutional neural networks (CNN), recurrent neural networks (RNN), and auto-encoders. AI models are capable to identify, categorize, and collect information from different fields like medical, education, etc. The DL models are classified under supervised, unsupervised and reinforcement learning.

Wireless capsule endoscopy (WCE) is a popular gastrointestinal investigation procedure. It's not just exploring the disease from the given symptoms; instead, take the overall digestive tract under observation. Endoscopic imaging [2] is a present-day, most advanced diagnostic technique that helps investigators to envision the inner organization of the desired organ. Our focus is on intestinal disorders [3], which is not easy to predict by general blood tests or verbal symptoms. The quality of images validates the analysis process. An endoscopic capsule is a tiny object embedded with a camera that can record 8-9 hours of video [4]. This is a painless procedure generally done under the observation of a gastroenterologist. The high-resolution endoscopes make the work easier and more efficient. Data extraction is done through standardized data collection techniques. Computer-aided machine is connected to the device by a wireless medium.

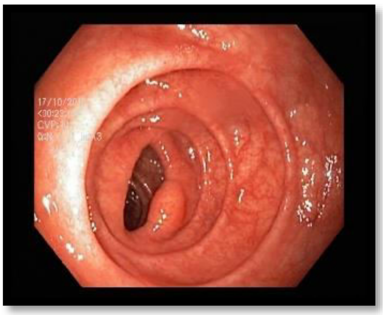

Fig. 1. Endoscopic image showing a gastrointestinal diseased region

Figure 1 represents an endoscopic image retrieved from the dataset taken by WCE (wireless capsule endoscopy). The shape of this image is (720×576×3), and it shows a diseased visual examination of an intestinal tract that has evidence of an abnormal condition. Endoscopic imaging is one of the foremost areas of investigation within gastroenterology.

Convolutional Neural Networks (CNNs) are a specific type of deep learning model that processes grid-like data, such as images. There are many types of neural networks, but CNNs are best suited for applications in health care. They use convolutional layers to automatically pull out important features like edges, corners, and textures from raw input data. By stacking multiple layers, [5] CNNs learn representations at different levels, from simple patterns to complex objects. Pooling layers reduce spatial dimensions and computational costs while keeping crucial information. Weight sharing in CNNs makes them efficient and less likely to overfit compared to traditional neural networks. CNNs are commonly used in computer vision tasks, including image classification, object detection, and facial recognition. Because of their high accuracy and strength, CNNs play a key role in many modern AI applications.

The primary contributions of this research are: (1) combining AI assistance with endoscopic images. The well-formed DL model collaborating with features of pre-processing techniques. (2) Comprehensive evaluation of gastrointestinal (GI) tract disorders. The new algorithm is designed to enhance the accuracy rate for endoscopic disease classification. (3) Overall, learning tools are AI-based and future additions are also discussed. In imaging studies, DL techniques, principally convolutional neural networks (CNNs), exhibited significant potential and result-oriented outcomes.

I. Literature Survey

Existing research papers studied from the years 2019 to 2025. Research papers from conferences and journals based on deep learning in healthcare are collected from different sources. The table represents the insights, methods used and limitations extracted from the papers.

Table I. Summary of existing work

Paper	Insights	Methods used	Limitations
[6]	Data augmentation addresses actual data challenges. Hyper parameter selection is crucial for CNN performance. Emphases given on the study of transfer learning techniques.	Explored DL approaches, CNN architectures and computational approaches.	Difficulties with deep learning and considering its alternatives.
[7]	DL models improve automation and accuracy in medical image classification. The hybrid model of CNN and auto encoder outperforms SVM, NB and CNN.	Feature selection methods used are Filter, wrapper, and hybrid approaches. CNN and Encoders are used for medical image classification.	Traditional methods lack in-area representation. Poor model hypothesis capacity on medical data.
[8]	Algorithm enhances images and overcomes traditional methods. Proposed algorithm resolves color distortion, and brightness issues in endoscopic images.	Image decomposition and base layers for image enhancement. Adaptive brightness correction. Algorithms compared with datasets.	Algorithm tested on few images. parameters and evaluation index settings need a detailed study.
[9]	Proposed cascaded model detects accurate gastric cancer images.	CNNmodels implemented like VGG-16, InceptionV3,	Limited image collection. Need of

		and ResNet-50. Segmentation using U-Net models for early gastric cancer. Grad-CAM used for visualization.	huge datasets for evaluation. Subjective evaluation of infected area.
[10]	New method for GIT disease recognition with deep CNN and geometric features. Proposed approach achieves high accuracy, improves classification performance, and reduces time.	Geometric feature extraction from lesions. DCNN for feature extraction.	Existing methods used hand-crafted features leading to wrong predictions. Challenges include lesion shape, size, color, texture, and irregularity.
[11]	DL systems for polyp detection with high performance. Use of faster R-CNN method with Inception Resnet. Proposed post-learning methods enhance detection and performance on videos.	Region-based CNN for polyp detection using Inception Resnet model. Image augmentation, false positive learning, and offline learning post methods.	Detection models show better performance metrics. Difficulty in comparing detection processing time due to different computer systems.

II. Algorithm Design and Pipeline

A. Dataset acquisition and preparation

The Endo-CNN follows the convolutional neural network architecture. The model used the prescribed dataset, which contains most of the possible diseased endoscopic images. The data pre-processing techniques [12] play a vital role in image analysis. A model needs to predict and learn features accurately from data and pre-processing helps in this context in enhancing accuracy. Many techniques like noise reduction, data augmentation, dimensionality reduction, contrast enhancement, image sharpening etc. can be applied by considering the image pattern to faster train the model and reduce over fitting.

Dataset: (www.kaggle.com).

Input: Raw endoscopic images I (Dyed-lifted polyps, Esophagitis, Dyed-resection margins, Normal pylorus, Normal cecum, Polyps, Normal z-line, and Ulcerative colitis) endoscopic images from 8 classes. The total images are 8,000 images.

- Batch size: {32 or 64}
- Image size: {720×576×3}
- Epochs: {30–50}

Output: Trained Endo-CNN model

B. Algorithm Development

Algorithm: Endo-CNN- A domain-specific multi-layer convolutional neural network.
Step 1: Data gathering
Step 2: Data preprocessing • Resize image I to size S $I \in R^{H \times W \times C}$ where H=224, W=224 and C=3. • Apply data augmentation on I using transformations (rotation, flip, zoom, etc.)
Step 3: Define architecture: • Convolutional Layers Apply Conv2D with 32 filters, kernel size 3×3, activation = ReLU Apply MaxPooling2D with pool size 2×2 Repeat Conv-Pool block to deepen feature learning: 2nd Conv2D layer → 32 filters, followed by MaxPooling2D • Flatten layer Flatten the output feature maps into a 1D vector $F \in R^{93312 \times 1}$ • Classification – Fully Connected Layers Apply a sequence of Dense layers: Dense(64), Activation: ReLU Dense(128), Activation: ReLU Apply Dropout Dense(256), Activation: ReLU Apply another Dropout Dense(512), Activation: ReLU • Output Layer Apply final Dense layer with 8 neurons (for 8 disease classes) Activation function: Softmax (for multiclass classification) Output: Probability vector $P = [p_1, p_2, \ldots, p_8]$ Each $p_i \in [0,1]$, and $\sum p_i = 1$
Step 4: Train Endo-CNN model • Initialize hyperparameters, epochs, and batch size. • Loss function: Categorical Cross-entropy • Optimizer: Adam Learning rate: α=0.001
Step 5: Evaluation & Validation using metrics • Evaluate the trained Endo-CNN on the Test set. • Compute: Accuracy, Confusion matrix, ROC-AUC, Precision, F1-scores End

The above algorithm is designed for an endoscopic image dataset (currently applied on the Kvasir dataset, having 8 diseased classes). Early convolutional layers help to detect basic visual patterns, such as edges (horizontal, vertical, diagonal), texture patterns, color gradients, contrast, brightness and shapes. While dense layers detect the presence of pathological regions, tissue classification and class-specific features.

Introducing non-linearity into a neural network with RELU (Rectified Linear Unit) as the activation function not only addresses vanishing gradients but also makes training faster by allowing only positive values to flow through the network. When applying MaxPooling2D to a feature map, you are performing a down-sampling process that extracts the maximum value from a region of the map defined by your pooling window. This helps to reduce the spatial size of the feature map and decreases the amount of computation for the network, while improving the robustness of features to small changes in location. Softmax is an activation function that converts the outputs of the last layer of a neural network into a probability distribution over multiple classes. You can examine how well a network is performing using a confusion matrix, which summarizes the results of a classification task by counting true positives, false positives, true negatives and false negatives. Each of these values can then be analyzed on its own, for a better understanding of how well the model is working by class. The ROC-AUC (Receiver Operating Characteristic - Area Under Curve) curves help you evaluate a given model's ability to differentiate between classes through the plotting of the true positive rate against the false positive rate. Precision is the ratio of correctly predicted positive samples to the total number of samples that were predicted to be positive, and is particularly important when the cost of falsely identifying a sample as positive is high. The F1-score combines precision and recall, providing an overall weighted measure based on both metrics [13].

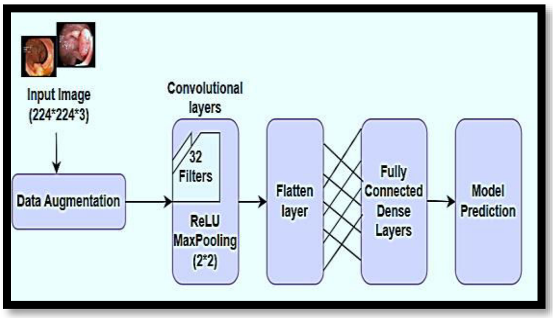

Fig. 2. The structural configuration of the Endo-CNN framework used in gastrointestinal diagnostics

Figure 2 depicts the proposed algorithm's architecture. The RGB images of 224 × 224 pixels are taken and augmented using various techniques to create additional versions of the original image and provide them to the model. The purpose of augmentation is to add variability and help reduce overfitting. Once the original and augmented images have been generated and passed to the convolutional layers, the various filters in these layers will learn different visual features from the images, such as edges, textures, and shapes. The next stage is to apply non-linear transformations using an activation function ReLU, allowing the model to learn complex patterns. Then, the model applies max pooling to reduce the number of data points and, therefore, the computational load while retaining the most prominent features. The next step in the process is to flatten the resulting feature maps into a 1D vector, so they can be used as input to a Fully Connected Layer (Dense Layer), which is responsible for integrating all of the learned features and ultimately providing classification. The classification is provided from the final layer of the CNN through the generation of a prediction.

III. Implementation and Result

The designed algorithm (Endo-CNN) is implemented using Python libraries and trained on a dataset (www.kaggle.com), having 8 classes of endoscopic images. The classes contain 8,000 images in total and after data augmentation; all the parameters are checked over 48,000 endoscopic images. Different data pre-processing techniques can be applied to enhance the image quality for better chances of correct identification of the disease. Further, the comparative study is performed by some pre-defined deep learning models with the designed model (Endo-CNN) to analyze the accuracy rate. The models are Res Net, Dense Net and VGG.

Res Net (Residual Network) and Dense Net (Densely Connected Convolution Networks) are DL architectures introduced to solve the complex problem of DNNs. Res Net uses residual blocks with skip connections that allow gradients to move more effortlessly during back propagation. Some of its most used variants are ResNet-50, ResNet-101, and ResNet-152. In Dense Net, each layer is linked to all other layers through feed-forward connections. It was developed to reduce the vanishing gradient problem. Some of its popular variants are DenseNet-121, DenseNet-169, and DenseNet-201, widely used in image classification and medical imaging. VGG (Visual Geometry Group) is a CNN architecture composed of convolutional layers with different filter sizes and strides. It's widely used for transfer learning; VGG-16 and VGG-19 are its popular variants.

System Flow Diagram

Deep learning architectures [14] are analyzed over the dataset and a comparative study is done. To analyze the model's performance average accuracy rate is depicted. Endoscopic images undergo data preprocessing to improve image quality and standardize inputs. The accuracy percentage helps in understanding the overall prediction rate of the model. DL models extract visual and contextual features from [15] endoscopic images.

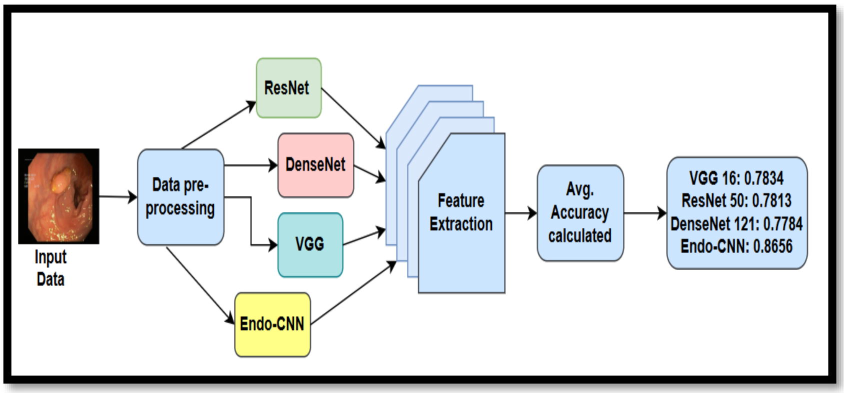

Fig. 3. Workflow of the proposed Endo-CNN-based diagnostic system

Figure 3 represents a comparative deep learning framework for endoscopic image analysis. The average accuracy for all the deep learning modelsis calculated and compared by considering all the dataset classes. The preprocessed data are then fed in parallel into multiple deep learning architectures, VGG, Res Net, Dense Net, and the proposed architecture Endo-CNN, each acting as a feature extractor. These models learn and extract high-level image features relevant to disease detection. The extracted features are evaluated, and the average classification accuracy is calculated to compare model performance. Finally, the results show that while VGG, Res Net, and Dense Net achieve competitive accuracy, the Endo-CNN model demonstrates the highest performance, indicating its superior suitability for the given medical imaging task.In multi-class classification, [16] there's no single False Positives/False Negatives per class, but each misclassified sample contributes a false positive (where the wrong class was predicted) and a false negative (where the true class was missed).

Using the formula: $$\text{Accuracy} = \frac{\text{correct prediction}}{\text{total samples}} \quad (1)$$

Table II. Performance evaluation and comparison of deep learning architectures and the proposed architecture

Dataset classes	Accuracy VGG 16	Accuracy Res Net 50	Accuracy Dense Net 121	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

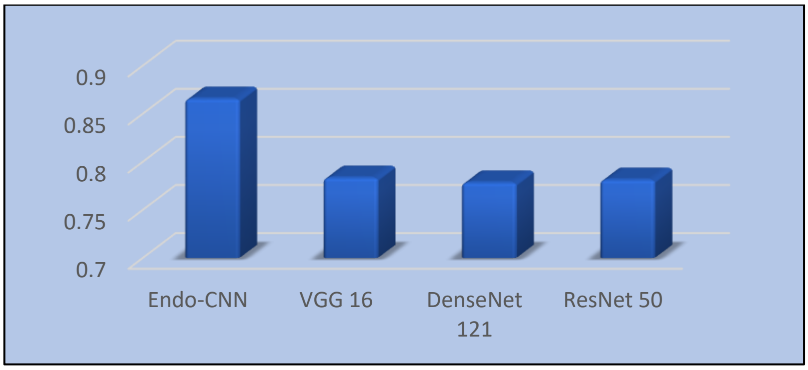

Fig. 4. Graphical representation of performance evaluation and comparison of deep learning architectures and the proposed architecture

The table above (2) and Figure 4 compare how well four different models from deep-learning (VGG-16, ResNet-50, DenseNet-121) do with classifying images of gastrointestinal endoscopy. Endo-CNN's performance consistently is high and shows either the best or near-highest classification accuracy for normal and abnormal categories of endoscopic images (particularly the following: Normal z-line: 0.9100, Ulcerative Colitis: 0.9000, Dyed Lifted Polyps: 0.8950), which means that Endo-CNN generalizes effectively across both normal and abnormal categories. ResNet50 has fairly stable classification accuracies across multiple class types, but is very poor at classifying specific features of learning to classify normal pylorus: 0.1250 indicating possible susceptibility of being influenced by class specific visual features. VGG-16 performed well on a number of classes (generally around 0.8750), but struggled to accurately classify Dyed Resection Margins (0.4275) & Esophagitis (0.7056) accurately. Although DenseNet-121 is generally consistent and just slightly lower than Endo-CNN yielding classification accuracy across all classes. Therefore, the overall results are suggesting that Endo-CNN will yield the best overall performance when using multi-class endoscopic images for classification purposes.

IV. Conclusion and Future Directions

In this work, we have proposed a multi-layered algorithm based on a CNN architecture [17, 18]. We sought to improve the convolutional model by incorporating additional features into a new algorithm, which was then evaluated using the endoscopic dataset. The model demonstrated promising performance, achieving an improved average score of 0.8656.

As per future improvements in the field of medical AI, especially for endoscopic image analysis, it requires explainability. Doctors need it for visual explanations, while AI researchers require it to understand the process and overcome the challenges that appear. It's very important to select the right XAI techniques [19] for a particular model. For example, if in a model researcher uses DL model without explainability → 70% accuracy in diagnosis, while in a second opinion, using DL model + Grad-CAM visualization → 85% accuracy is observed with better trust on the model. Different XAI methods [20] can be applied to the type of data (image, text, or video). Artificial intelligence can spontaneously detect various lesions, minimize the reading times of the endoscopic video and lower missed diagnoses. The proposed algorithm was intended to work in a real-time working environment for gastrointestinal endoscopic images, which will help doctors in detecting all possible disorders.

References

1. Zihan Nie, Muhao Xu, Zhiyong Wang, Xiaoqi Lu and Weiye Song (2024). A Review of Application of Deep Learning in Endoscopic Image Processing. J. Imaging10 (11), 275.
2. J. Mosleh Hmoud Al-Adhaileh, Ebrahim Mohammed Senan, Waselallah Alsaade, Theyazn H. H Aldhyani, Nizar Alsharif, Ahmed Abdullah Alqarni, M. Irfan Uddin, Mohammed Y. Alzahrani, Elham D. Alzain and Mukti E. Jadhav

(2021). Deep Learning Algorithms for Detection and Classification of Gastrointestinal Diseases. Complex System Modelling in Engineering under Industry 4.0, Hindawi Complexity.

3. Esha Saxena, Manoj Yadav, Meenakshi Yadav, and Preety Shoran (2022). Artificial Intelligence-based Diagnostic Analysis for Wireless Capsule Endoscopy in Obscure Bowel Disease Detection: A Potential. Proceedings of the 4th International Conference on Information Management & Machine Intelligence (ICIMMI '22). Association for Computing Machinery, Article 3.
4. Esha Saxena, Suraiya Parveen, Mohd. Abdul Ahad & Meenakshi Yadav (2024). Unveiling the potential of AI in Gastroenterology: Challenges and Opportunities” in International Conference on Deep Learning and Visual Artificial Intelligence (ICDLAI-24).
5. Chempavathy and S. K. Mouleeswaran (2025).Fusion-Driven Spatio-Temporal Deep Learning for IoT Intrusion Detection: Integrating CNN, TCN, and Capsule Net Features with LSTM. Engineering, Technology and Applied Science Research Vol.15, Issue 6.
6. Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaria, Mohammed A. Fadhel, Muthana Al-Amidie and Laith Farhan (2021).Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Alzubaidi et al. J Big Data8(1):53.
7. Bhanu Prakash Battul, Duraisamy Balaganesh (2020). Medical Image Data Classification Using Deep Learning Based Hybrid Model with CNN and Encoder. Revue D’intelligence Artificielle, Vol. 34, 645-652.
8. Dat Tien Nguyen, Min Beom Lee, Tuyen Danh Pham, Ganbayar Batchuluun, Muhammad arsalan and Kang Ryoung Park (2020). Enhanced Image-Based Endoscopic Pathological Site Classification Using an Ensemble of Deep Learning Models. Sensors 20(21), 5982.
9. B. Surya Prasath, Dang N. H. Thanh, Le Thi Thanh, N. Q. Sang, and S. Dvoenko (2020). “Human Visual System Consistent Model for Wireless Capsule Endoscopy Image Enhancement and Applications. Pattern Recognition and Image Analysis, Vol. 30, ISSN 1054-6618.
10. Park J, Hwang Y, Nam JH, Oh DJ, Kim KB, Song HJ (2020). Artificial intelligence that determines the clinical significance of capsule endoscopy images can increase the efficiency of reading. PLOS ONE 15(10).
11. Ibrahim Abdulrab Ahmed, Ebrahim Mohammed Senan and Hamzeh Salameh Ahmad Shatnawi (2023). Hybrid Models for Endoscopy Image Analysis for Early Detection of Gastrointestinal Diseases Based on Fused Features. Diagnostics13 (10):1758.
12. Subhashree Mohapatra, Pukhraj Singh Jeji, Girish Kumar Pati, Manohar Mishra, Tripti Swarnkar (2024). Comparative exploration of deep convolutional neural networks using real-time endoscopy images. Biomedical TechnologyVol. 8, 1-16.

13. Rajiv Avacharmal, Saigurudatta Pamulaparthyvenkata, Dileep Kumar Pandiya, Krishna Prakash R, Piyush Ranjan, Anandaganesh Balakrishnan and Maheshwaran (2024). Comprehensive investigation on Deep learning models: Applications, Advantages, and Challenges. 15th International Conference on Computing Communication and Networking Technologies (ICCCNT).
14. Ming-Hung Shen, Chi-Cheng Huang, Yu-Tsung Chen, Yi-Jian Tsai, Fou-Ming Liou, Shih-Chang Chang and Nam Nhut Phan (2023). Deep Learning Empowers Endoscopic Detection and Polyps Classification: A Multiple-Hospital Study. Diagnostics13 (8):1473.
15. V A Pavlov, E N Velichko, S V Zavjalov, I E Govorov, T M Pervunina and E V Komlichenko (2019). On specific features of the endoscopic image processing.IOP Conf. Series: Journal of Physics1236 012036.
16. Md Shakhawat Hossain, Md Mahmudur Rahman, Mm Mahbubul Syeed, Mohammad Faisal Uddin, Mahady Hasan, Md. Aulad Hossain, Amel Ksibi, Mona M. Jamjoom, and Zahid Ullah, And Md Abdus Samad (2023) .DeepPoly: Deep Learning based Polyps Segmentation and Classification for Autonomous Colonoscopy Examination. IEEE AccessVol. 11, pp. 101767-101777.
17. Mengfang Li, Yuanyuan Jiang, Yanzhou Zhang and Haisheng Zhu (2023). Medical image analysis using deep learning algorithms. Front. Public HealthVol.11.
18. Kopelman, Gal O, Jacob H, Siersema P, Cohen A, Eliakim R, Zaltshendler M and Zur (2019). Automated Polyp Detection System in Colonoscopy Using Deep Learning and Image Processing Techniques. Journal of Gastroenterology and Its Complications. Vol. 3.Issue 1, ISSN: 2575-5501.
19. Deepshikha Bhati, Fnu Neha and Md Amiruzzaman (2024). A Survey on Explainable Artificial Intelligence (XAI) Techniques for Visualizing Deep Learning Models in Medical Imaging. J. Imaging, 10, 239.
20. Doniyorjon Mukhtorov, Madinakhon Rakhmonova, Shakhnoza Muksimova and Young-Im Cho (2023). Endoscopic Image Classification Based on Explainable Deep Learning. Sensors23 (6), 3176.