# SHAP-Guided Gastrointestinal Disease Classification with Lightweight Parallel Depthwise Separable CNN and Ridge Regression ELM

**Md. Nahiduzzaman[1], Md. Faysal Ahamed[2], Norah Saleh Alghamdi[3, *], S. M. Riazul Islam[4, *]**

[1] *Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh.* (nahiduzzaman@ece.ruet.ac.bd)

[2] *Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh.* (faysalahamedjishan@gmail.com)

[3] *Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia.* (nosalghamdi@pnu.edu.sa)

[4] *Department of Computing Science, University of Aberdeen, Aberdeen, UK* (riazul.islam@abdn.ac.uk)

**Corresponding Author:**

Norah Saleh Alghamdi

*Princess Nourah bint Abdulrahman University*

*Riyadh, 11671, Saudi Arabia*

Email: nosalghamdi@pnu.edu.sa

S. M. Riazul Islam

*University of Aberdeen*

*Aberdeen, UK.*

Email: riazul.islam@abdn.ac.uk

## *Abstract*

Gastrointestinal diseases pose a global health challenge, necessitating prompt detection and precise categorization for effective treatment. For the first time, a study investigated 24 distinct gastrointestinal (GI) problems across two testing trials involving 24 and 13 different GI tract diseases. This research introduces a novel Lightweight Parallel Depth-Wise Separable Convolutional Neural Network (LPDS-CNN), along with a Ridge Regression Extreme Learning Machine (RRELM) classifier, for the accurate identification of images from the GI endoscopy dataset. A hybrid pre-processing technique was developed to enhance image quality and minimize noise, combining artefact removal, contrast-limited adaptive histogram equalization (CLAHE), sharpening, and Gaussian filtering. The LPDS-CNN effectively captures discriminative features, retaining a mere 0.498 million parameters with nine layers, significantly reducing complexity during computations. Impressively, the proposed framework delivers remarkable performance on various metrics. In the first trial (24 classes), average precision, recall, f1, accuracy, and ROC-AUC scores stand at 83.42±0.27%, 68.08±0.311%, 72.63±0.275%, 89.13%, and 98.11% respectively. In the second trial (13 classes), scores are even higher, with 91.08±0.062%, 88.15±0.092%, 89.54±0.066%, 92.15%, and 98.26%. The proposed framework is exceptionally efficient, with an average training and testing time of 0.0192 and 0.002 seconds, respectively. Comparative analysis with state-of-the-art (SOTA) transfer learning (TL) methods validates the model's real-time analytical prowess. Additionally, the integration of SHAP (Shapley Additive Explanations) enhances interpretability, offering valuable insights for confident real-world GI diagnosis. This comprehensive approach shows the potential to improve diagnosis and enable earlier real-time treatment worldwide.

**Keywords**

GastroVision, Lightweight Parallel Depth-wise Separable Convolutional Neural Networks (LPDS-CNN), Ridge Regression Extreme Learning Machines (RRELM), Contrast-Limited Adaptive Histogram Equalization (CLAHE), Gastrointestinal diseases, Multiclass, Classification, and Shapley Additive Explanations (SHAP).

## 1. Introduction

The gastrointestinal (GI) tract is prone to various diseases directly impacting the digestive system's performance. Medical imaging is crucial in identifying and evaluating different diseases, providing valuable insights, and assisting in precise evaluations for appropriate treatments. When dealing with huge volumes of imaging data, endoscopists and medical professionals encounter a challenge that can elevate the probability of inaccurate medical inspections. There is a high demand for medical professionals, which can be expensive, error-prone, and time-consuming. Additionally, it is challenging for rural areas to keep up with the need for extra medical technologists [1].

Digestive diseases, including polyps, esophageal, colon cancer, and ulcerative colitis, have emerged as significant causes of death in recent times. Among those possibilities, colorectal cancer stands out as the most commonly occurring type of cancer, potentially leading to mortality. GI cancer causes 26% of new cases and is responsible for 35% of cancer-related fatalities worldwide [2]**.** Approximately 4.8 million cases of GI cancer were diagnosed in 2018, leading to the tragic

loss of 3.4 million lives [3]. Colon cancer, comprising 1.93 million new cases in 2015, ranks as the third most prevalent type of GI cancer, behind pancreatic cancer (466,003 deaths) and liver cancer (905,677 new cases in 2015), and stomach cancer (1.09 million new cases in 2015) and esophageal cancer (604,100 new cases in 2015) [4]. According to experts, there is expected to be a significant increase in cancer cases by 2040, with a projected surge of 58%. Additionally, there will be a corresponding rise in cancer-related deaths, estimated to increase by 73% [2]. Early diagnosis of these diseases is crucial for mitigating risks and facilitating successful treatment. Endoscopy or esophagogastroduodenoscopy (EGD), colonoscopy, and capsule endoscopy imaging techniques such as endoscopic ultrasound (EUS), CT scan, magnetic resonance imaging (MRI), and positron emission tomography (PET) scan are utilized to detect GI diseases. EGD and colonoscopy are widely acknowledged as the best and most accurate methods for examining the upper and lower bowling part to detect abnormalities and potential health issues.

A regular endoscopic screening misses many lesions due to the existence of excretion and the intricate organ structure. There is a cause for concern as many polyps are not detected during bowel cleansing for cancer or precursor lesion diagnosis. Rates of undetected polyps range from 21.4% to 26.8% [5]. Additionally, identifying these growths can be challenging since they can share similarities across different categories, posing a significant diagnostic challenge. Wireless capsule endoscopy (WCE) is a modern approach for visualizing the inside of the GI tract. Previously, individuals had to ingest a capsule with an inbuilt camera during routine endoscopies, which captured multiple pictures as it traversed through the GI tract. The lesion is tracked by compiling all the photos into a single video frame. However, the primary issue is that classifying many diseases takes too much time.

Many researchers have devoted their time to developing an image diagnosis system for accurately classifying precursor lesions for colorectal diseases. Fundamental processes in these methods involve increasing contrast and eliminating noise, segmenting an infected region by extracting relevant features, and finally classifying it [6]. Lower classification accuracy can lead to misclassification of diseases. Addressing the numerous challenges and limitations faced by GI images, significant emphasis is required on enhancing image quality, mainly due to restrictions in magnitude and intensity that adversely affect the overall image clarity. We have successfully developed a lightweight system for classifying 24 types of gastrointestinal diseases for the first time using automated DL methods. Our strategy involved implementing novel procedures for image pre-processing and automated multiclassification. The primary contribution of this research is outlined as follows:

- A hybrid pre-processing method has been developed to enhance the quality of captured images from multi-class GastroVision dataset by reducing noise and improving contrast. This technique has proven to be highly effective in improving the performance of DL models across a wide range of multi-class datasets.
- This is the first work, where a lightweight parallel depth-wise separable CNN (LPDS-CNN) has been designed to enhance the accuracy of detecting 24 types of gastrointestinal infected regions while reducing computational requirements. By extracting important

features from the enhanced images, this model is able to classify them with better precision and accuracy.

- A hybrid ridge regression extreme learning machine (RRELM) classifier has been developed to improve the classification performance through better feature extraction and learning.
- Extensive comparisons with the latest SOTA models demonstrate that the proposed framework outperforms them regarding classification accuracy, model parameters, and classification layer complexity.
- The framework's interpretability is highlighted using SHAP (Shapley Additive Explanations), which offers valuable insights into the model's decision-making process. This greater transparency instills confidence and trust in the framework's real time practical diagnostic capabilities.
- Two separate testing phases involving 24 and 13 classes have been presented to assess the model's versatility without augmentation procedures at the model's worst-case scenarios.

Section 2 of this research provides a complete overview of the previous related works in this field. Section 3 introduces the suggested methodology, encompassing an overall framework, dataset description, image pre-processing, feature extraction, and performance metrics. In Section 4, comprehensive classification results are presented, accompanied by a thorough discussion. Section 5 offers the final remarks and summary of this research.

## 2. Related Works

Numerous research studies have been carried out on comparable tract areas [7–13]. But very little work has been done recently on the multi-class classification of the GI system. The popularity of using medical imaging data for GI classification research is increasing. The approach used in this study adheres to the customary framework for machine learning (ML) [14–16] and deep learning (DL) [6,17–20].

Researchers have utilized various abnormalities such as cancer, blood, polyps, ulcer lesions, Dyed-lifted-polyps, and Ileocecal to identify issues in endoscopy capsule frames. However, these systems have focused more on extracting the maximum number of features rather than ensuring a balanced real-time benchmark. To prepare the data, morphological operations, and statistical analysis were performed. There were limited multilevel and binary level (healthy or ill) classification techniques, where images were analyzed pixel-by-pixel and region-by-region. Many machine learning researchers employed support vector machine (SVM) in classification tasks. Numerous research studies on GI disorders, including polyps, colon-related diseases, and capsule endoscopy, have been published in medicine [14–19].

Yeh et al. [21], introduced an approach for identifying hemorrhaging and ulcers in images obtained by WCE. The conventional method of manually analyzing around 60,000 images per examination, which typically requires about 2 hours from experts, highlighted the need to create a computer-assisted system for processing images. Using color characteristics to evaluate the condition of the intestinal tract has produced promising results in detecting areas of bleeding and ulcers. In their study, Lee et al. [14], employed DL models like Inception-V3, VGG16, and ResNet-50 to classify

GI images as usual or having ulcers. ResNet-50 demonstrated superior performance among these deep networks, surpassing the other models when employing this approach. The computer-aided detection system for ulcer detection developed by Yuan et al. [22]. The method adequately found ulcer areas by using a multi-level super-pixel representation. The system operated Locality-constrained Linear Coding (LLC) and saliency max-pooling to store visual features. Pogorelov et al. [23], consisted of images obtained from the GI tract. The KVASIR dataset was organized into two image categories related to endoscopic polyp removal, three critical anatomical landmarks and three therapeutically relevant observations. Furthermore, implementing the VGGNet model has been proposed for detecting GI ulcers. This approach involves examining images captured during regular endoscopy procedures. The VGGNet model has demonstrated an impressive detection accuracy of 90.8% when tested on a dataset of 5,360 WCE images containing ulcers and 440 healthy images. WCENet, an advanced deep CNN created by Jain et al. [24], was the most effective tool for detecting and locating anomalies in images obtained from WCE. With an accuracy rate of 98% on the KID dataset, WCENet surpassed other models in the field and had enormous potential for clinical use. Using the Grad-CAM++ technique and a customized SegNet model, WCENet effectively categorizes images into four groups (normal, vascular, polyp, and inflammatory), while accurately delineating abnormal regions. It was an invaluable tool for the diagnosis of GI anomalies. The study conducted by Lan et al. [25], proposed a hybrid approach that incorporates unsupervised deep learning techniques to summarize videos using the weakly supervised cross-modal embedding framework. They utilized various networks like long short-term memory (LSTM) and autoencoder to enable healthcare experts to WCE videos thoroughly. In [26], a state-of-the-art classification system for GI diseases was introduced. ResNet-50 and ResNet-152 models achieved a remarkable classification performance of 96.43%. In another study [27], discrete wavelength transformation (WT) and CNN techniques were employed to classify polyp and esophagitis classes. This approach was applied to the Kvasir dataset, producing an outstanding accuracy rate of 96.65%.

This is the first-ever research of 24 multi-class classifications, significantly advancing the classification of GI tract diseases. The significance and originality of this study's contribution are highlighted by the fact that previous research in this area required significantly more attention, highlighting the importance of the study's contribution. Through the inventive design of LPDS-CNN and a hybrid RRELM classifier, this work not only sets a new benchmark for accuracy but also establishes a foundation for future research into complex GI tract disease classification algorithms. However, there was a significant problem with the computing requirements for this research. The model required many parameters and layers, resulting in more extended processing periods, making it challenging to implement the classification model efficiently. Moreover, some advanced methods necessitate high-quality images for accurate classification, demanding real-world applications, particularly those involving embedded systems [5], [6], [14]–[20], [7]–[13]. Enhancing existing models by reducing parameter counts, shortening processing times, and improving classification performance is crucial to encourage the widespread adoption of GI disease classification models. This study identifies these obstacles and suggests a novel and efficient solution.

## 3. Methodology

### 3.1 Proposed Framework

A framework utilizing DL techniques has been implemented to tackle the challenges of identifying GI diseases. Figure 1 provides the detailed working steps of this research. First, we developed a novel approach to automatically classifying disease in human GI organs, as no such study existed previously. After collecting and analyzing the images, the labelled data was split into twenty-four categories. The training, and testing ratio 80:20 of the samples was conducted. The images undergo reshaping and pre-processing to ensure normalization and enhancement. An advanced novel neural network architecture called the LPDS-CNN and parallel CNN (P-CNN) were utilized to extract essential image features. Additionally, two classifiers, Pseudo Extreme Learning Machine (PELM) and RRELM, were developed to distinguish class identification performances. An explanatory strategy based on the SHAP method was employed to gain valuable insights into the model's decision-making process.

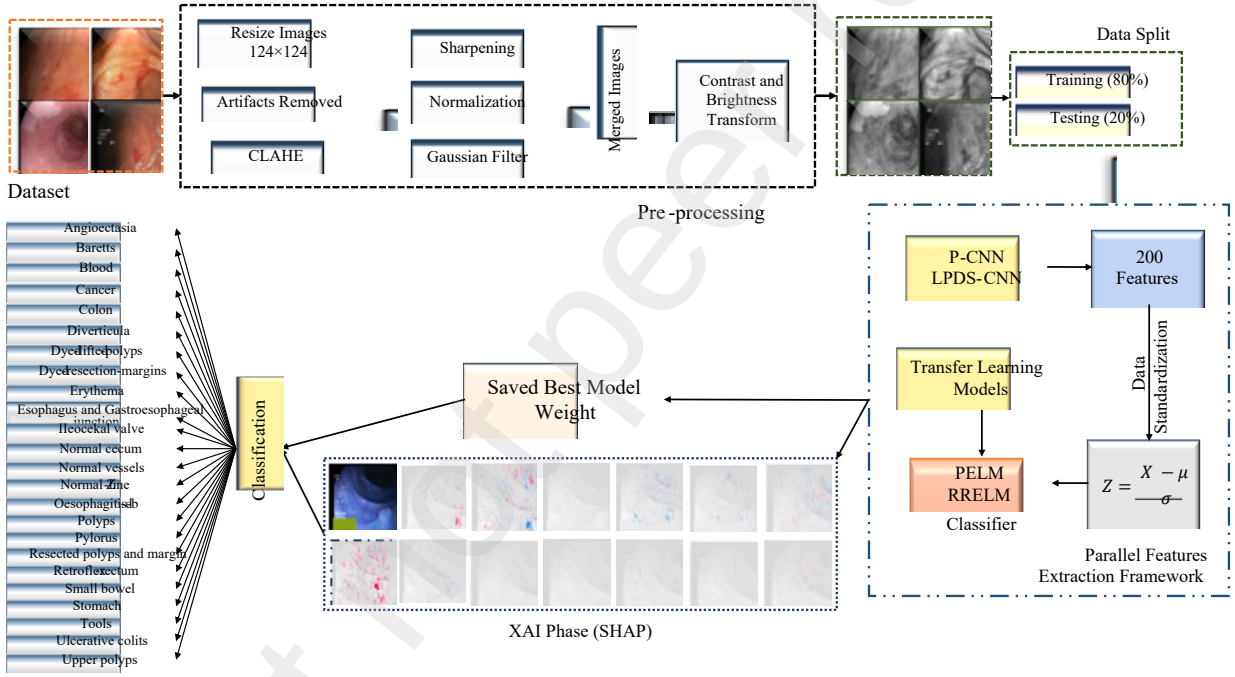



***Figure 1: Proposed research framework on multi-class endoscopic image classification.***

### 3.2 Dataset Description

The identification of GI diseases was significantly limited due to issues with the datasets. However, Jha et al. [28] presented a new multiclass endoscopy dataset called GastroVision. This dataset was obtained from various gastrointestinal organs and available at here [29]. During an endoscopy, the upper gastrointestinal tract comprised the esophagus, stomach, and small intestine. This classification was meticulously subdivided into anatomical landmarks and pathological findings, encompassing diverse groups. The subsequent information shaded light on these subcategories of the GI tract. Colonoscopy is a procedure used to inspect the lower gastrointestinal tract, which includes the colon, rectum, and terminal ileum (the final part of the small intestine). This data collection comprised 8,000 images, which are classified into 24 unique labels. The

images are separated into three groups based on the digestive tract - upper GI, lower GI, and a combination. Table 1 and Figure 2 provide a comprehensive overview of the dataset and sample images. The dataset images were taken from two separate locations: Baerum Hospital in Vestre Viken Hospital Trust, Norway and the Karolinska University Hospital in Stockholm, Sweden.

***Table 1. Overall datasets on both twenty-four classes and thirteen classes.***

| Testing phase | Disease Types | Training | Testing |
|---|---|---|---|
| Trial 1: 24-classes | Angioectasia | 13 | 3 |
| | Baretts | 31 | 8 |
| | Blood | 138 | 32 |
| | Cancer | 103 | 17 |
| | Colon | 1042 | 258 |
| | Diverticula | 13 | 2 |
| | Dyed-lifted-polyps | 125 | 38 |
| | Dyed-resection-margins | 198 | 50 |
| | Erythema | 8 | 4 |
| | Esophagus and Gastroesophageal junction | 437 | 118 |
| | IIeocekal valve | 151 | 35 |
| | Normal cecum | 31 | 10 |
| | Normal vessels | 277 | 67 |
| | Normal Z-line | 47 | 11 |
| | Oesophagitis-b-d | 14 | 2 |
| | Polyps | 642 | 171 |
| | Pylorus | 33 | 8 |
| | Resected polyps and margin | 101 | 17 |
| | Retroflex-rectum | 46 | 12 |
| | Small bowel | 834 | 206 |
| | Stomach | 1055 | 264 |
| | Tools | 1006 | 254 |
| | Ulcerative colits | 6 | 3 |
| | Upper polyps | 49 | 10 |
| | **Total** | **6400** | **1600** |
| Trial 2: 13-classes | Blood | 140 | 30 |
| | Cancer | 96 | 24 |
| | Colon | 1030 | 270 |
| | Dyed Lifted Polyps | 130 | 33 |
| | Dyed-resection-margins | 199 | 49 |
| | Esophagus and Gastroesophageal junction | 437 | 118 |
| | IIeocekal valve | 149 | 37 |
| | Normal vessels | 262 | 82 |
| | Polyps | 671 | 142 |
| | Resected polyps and margin | 92 | 26 |
| | Small bowel | 830 | 210 |

| | | | |
|---|---|---|---|
| | Stomach | 1074 | 245 |
| | Tools | 998 | 262 |
| | **Total** | **6108** | **1528** |

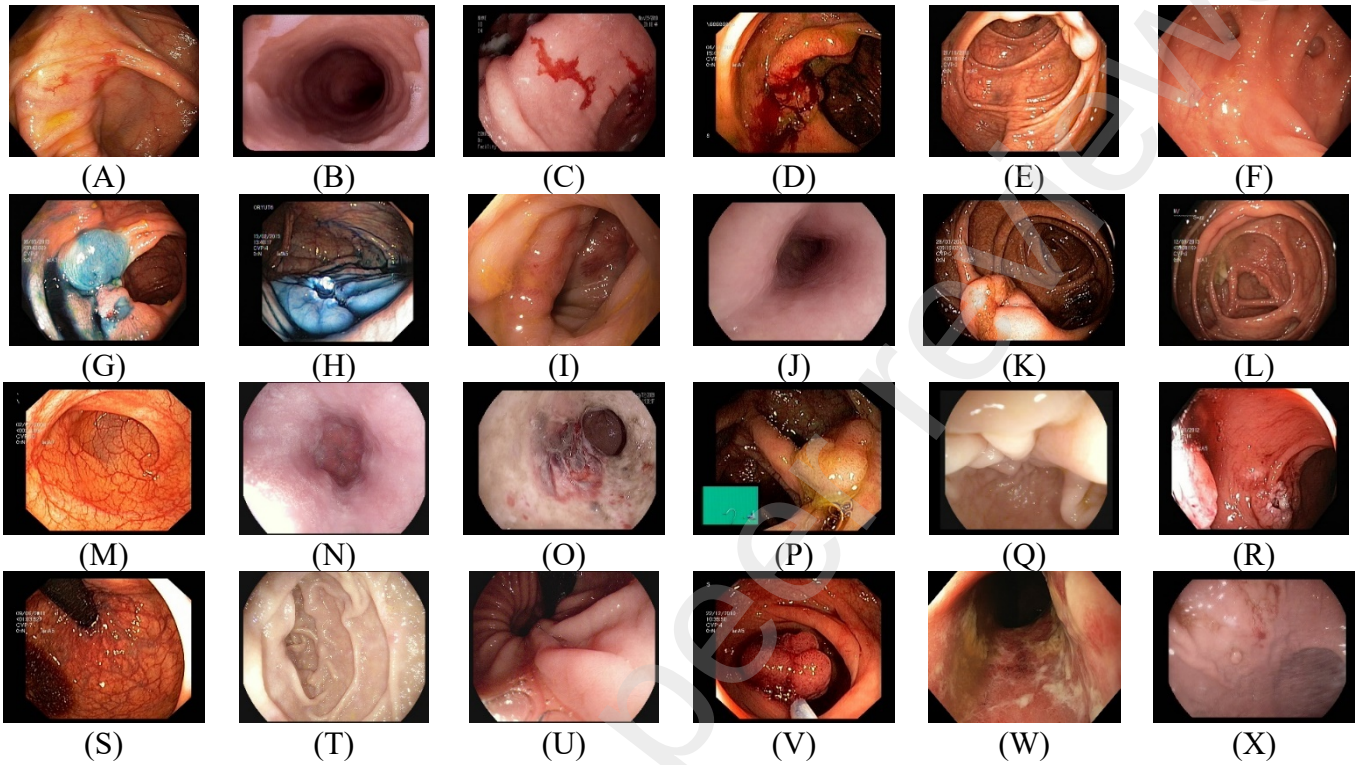
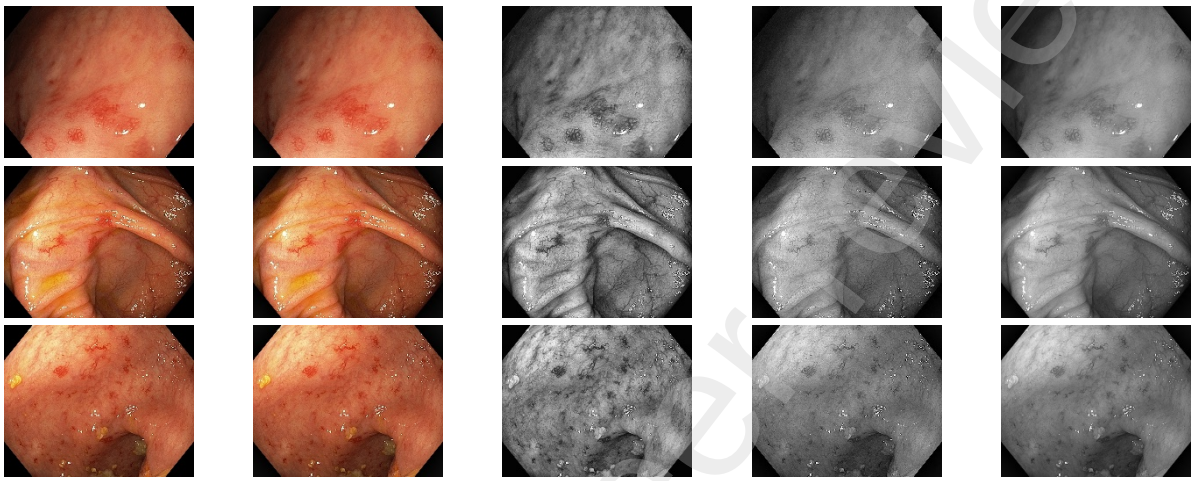



***Figure 2: GastroVision dataset include (A) Angio ectasia, (B) Baretts, (C) Blood, (D) Cancer, (E) Colon, (F) Diverticula, (G) Dyed-lifted-polyps, (H) Dyed-resection-margins, (I) Erythema, (J) Esophagus and Gastroesophageal junction, (K) IIeocekal valve, (L) Normal cecum, (M) Normal vessels, (N) Normal Z-line, (O) Oesophagitis-b-d, (P) Polyps, (Q) Pylorus, (R) Resected polyps and margin, (S) Retroflex-rectum, (T) Small bowel, (U) Stomach, (V) Tools, (W) Ulcerative colits, and (X) Upper polyps multi-classes.***

### 3.3 Image Preprocessing

Image processing plays a crucial role in GI multi-class endoscopy classification. Our proposed hybrid technique, comprising four distinct image processing methods, which is vital in enhancing model performance by facilitating better feature extraction in the context of GI endoscopy image classification. Firstly, the inpainting technique efficiently removes unwanted artefacts from the images, creating a clean and artefact-free dataset. This pre-processing step ensures that the model focuses solely on the relevant gastrointestinal features without being misled by distracting elements. CLAHE vastly improves image contrast and the clarity of critical features. The model can extract more discriminative features with higher accuracy because of CLAHE's ability to equalize local regions, bringing attention to subtle structures and crucial locations. By using the sharpening filter, the aesthetic qualities of the image can be enhanced, resulting in improved details such as patterns and edges. Incorporating a Gaussian blur reduces image noise, resulting in more refined representations. The Gaussian blur guarantees that the model emphasizes authentic

anatomical characteristics while minimizing the intrusion of extraneous noise by reducing undesirable variations and shortcomings. In conclusion, our hybrid image processing method enhanced endoscopic images by lowering noise, enhancing contrast and edges, and facilitating improved feature extraction. It includes two more key processes, normalization, and image down sampling, to resize the images to a standardized 124×124 pixels and the aforementioned preconditioning procedures. These processes are crucial in elevating model efficiency and feature extraction. Figure 3 shows the processed representations of the original samples individually.



(A) (B) (C) (D) (E)

***Figure 3: Pre-processed images include (A) original images, (B) artifacts removed, (C) CLAHE, (D) sharpening, and (E) gaussian filter.***

### 3.4 Deep Learning Model

After researching the strengths and weaknesses of many TL models, most TL models have unnecessarily huge parameters, layers, and sizes, dramatically increasing the processing power requirements. We developed a lightweight parallel depth-wise separable convolutional neural network (LPDS-CNN) to overcome these obstacles. This novel model is intended to reduce the computing burden by having fewer parameters, fewer layers, and more negligible overhead. In the following sections, the comprehensive summary of the LPDS-CNN model is presented.

#### 3.4.1 Feature Extraction

The primary objective was to create a CNN model that efficiently extracts vital features with minimum necessary parameters and network layers. This method was essential to guarantee effective performance while maintaining its use in practical settings. It is crucial to achieve the most effective arrangement of layers in a CNN model. More parameters and layers may prevent the model from achieving unique features, leading to performance limitations. On the other hand, an excess of parameters and layers can result in overfitting, causing extended processing times and requiring more computational resources. Therefore, finding the perfect balance is essential to ensure accurate feature extraction and practical implementation.

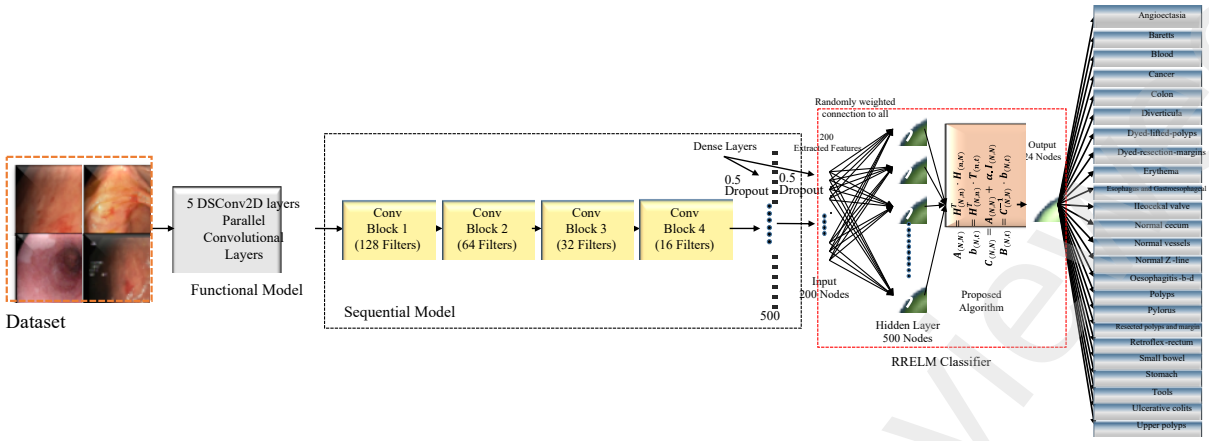



**Figure 4: Proposed LPDS-CNN-RRELM architecture.**

We developed a CNN model that extracts the features efficiently and practical size. To accomplish this objective, we designed a simplified LPDS-CNN model that effectively managed the complexity of layers and adheres to parameter constraints. Figure 4 visualizes the suggested LPDS-CNN model architecture. We improved the accessibility of the CNN model through a refined strategy for extracting significant features. The model has nine convolution layers (CL) and two fully connected layers (FC) to achieve optimal balance. We used five parallel CL instead of just one CL to extract the most essential features (see figure 5). Contrariwise, employing five consecutive CLs would increase model complexity by increasing the number of layers (depth). To address this issue, we employed to run the initial five CLs concurrently, and their selection was determined by trial and error.

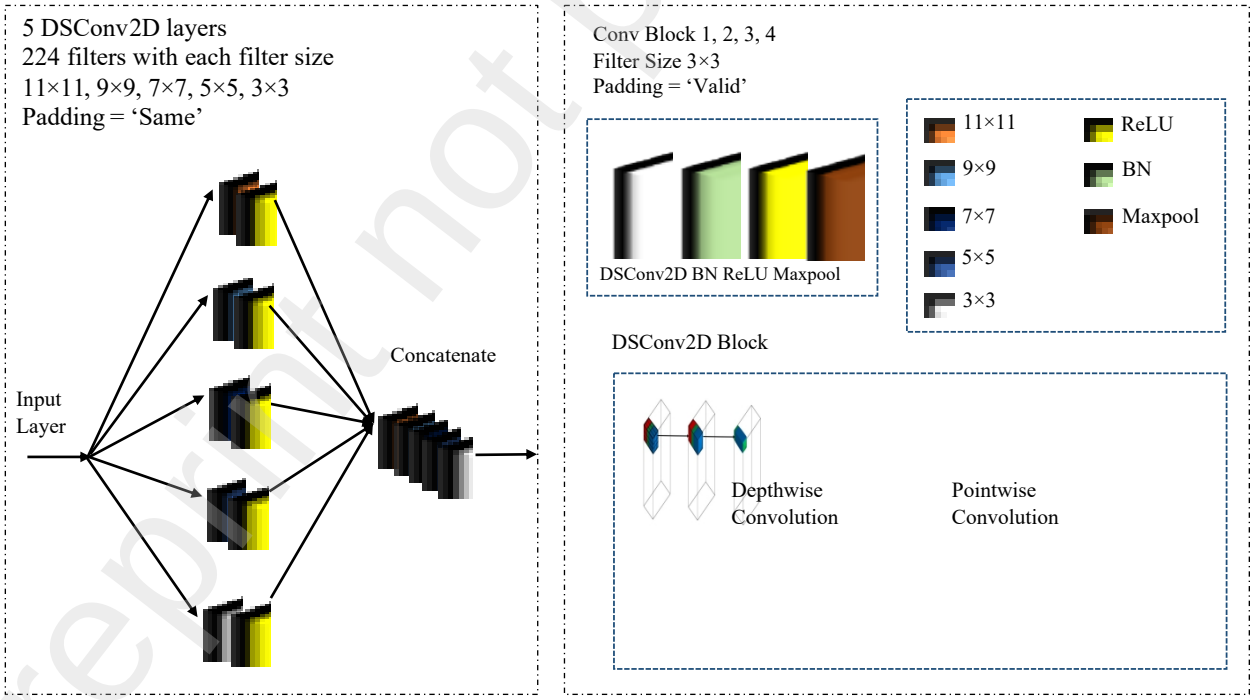



**Figure 5: Detail overview of the convolution block.**

The kernel sizes used were 11×11, 9×9, 7×7, 5×5, and 3×3, respectively. In our study, we have followed Krizhevsky et al.'s method of selecting the kernel size [30]. They utilized larger kernel sizes like 11×11 and accomplished satisfactory performance in classification. We analyzed and fused different kernels, considering their varying sizes, to identify essential features and achieve optimal classification performance. This is because different kernels generate unique feature maps. For optimal results when extracting critical data from the border element of GI images, keeping the padding size consistent for the first five CLs is necessary. The feature maps acquired from these concurrent CLs must be merged and seamlessly passed on to a sequential CL without error.

The CNN architecture was enhanced by incorporating depth-wise separable convolution (DSC), a technique that partitions the conventional convolution operation into two distinct steps: depth-wise and pointwise. A compact kernel for a specific segment of DSC is employed on an infusion feature map, generating a new feature map output with an equivalent number of channels. A new feature map with fewer channels is generated by applying a 1×1 convolutional kernel to each channel independently during the pointwise convolution. This statement highlights the paramount importance of DSC. The significant reduction in computational complexity directly results from this tightening of the settings. In the concluding stage, we integrated three CLs, along with the utilization of BN and max-pooling (2×2 kernel size). These CLs filters were as 128, 64, 32 and 16, correspondingly. Each filter was equipped with 3×3 kernels and configured to utilize VALID padding. To optimize the efficacy of the model, BN is incorporated. This technique effectively re-establishes the mean and standard deviation of the inputs for every layer, leading to enhanced speed and stability during model execution. All CLs used the Rectified Linear Unit (ReLU) activation function. In addition to the inclusion of two FC layers, dropout was implemented as a means to mitigate overfitting and enhance the efficiency of the training procedure. During each training iteration, a dropout technique was employed where 50% of all nodes were randomly deactivated. This approach aimed to enhance generalization and accelerate convergence. In the final FC layer, 200 features were extracted and for increasing the classification performance we used Ridge Regression Extreme Learning Machine (RRELM) classifier instead of SoftMax function. The model was trained using a loss function based on the sparse categorical loss function. For this operation, we chose to use a 32-batch-size ADAM optimizer. The model was trained for 200 epochs at a learning rate 0.001 (decided by trial and error). The details model summary is presented at Table 2.

***Table 2. Lightweight parallel depth-wise separable convolutional neural network (LPDS-CNN) model summary.***

| Layer Type | Output Shape | Params |
|---|---|---|
| Model input | (None, 124, 124, 3) | 0 |
| model (Functional) | (None, 124, 124, 1280) | 5975 |
| Separable conv2d 5 | (None, 122, 122, 128) | 175488 |
| Batch normalization | (None, 122, 122, 128) | 512 |
| Activation | (None, 122, 122, 128) | 0 |
| Max pooling 2d | (None, 61, 61, 128) | 0 |
| Separable conv2d 6 | (None, 59, 59, 64) | 9408 |
| Batch normalization 1 | (None, 59, 59, 64) | 256 |
| Activation 1 | (None, 59, 59, 64) | 0 |
| Max pooling2d 1 | (None, 29, 29, 64) | 0 |

| | | |
|---|---|---|
| Separable conv2d 7 | (None, 27, 27, 32) | 2656 |
| Batch normalization 2 | (None, 27, 27, 32) | 128 |
| Activation 2 | (None, 27, 27, 32) | 0 |
| Max pooling2d 2 | (None, 13, 13, 32) | 0 |
| Last conv | (None, 11, 11, 16) | 816 |
| Batch normalization 3 | (None, 11, 11, 16) | 64 |
| Activation 3 | (None, 11, 11, 16) | 0 |
| Max pooling2d 3 | (None, 5, 5, 16) | 0 |
| Dropout | (None, 5, 5, 16) | 0 |
| Flatten | (None, 400) | 0 |
| Dense | (None, 500) | 200500 |
| Batch normalization 4 | (None, 500) | 2000 |
| Dropout 1 | (None, 500) | 0 |
| DenseLastGV | (None, 200) | 100200 |
| **Total Parameters:** | 498,003 | |
| **Trainable Parameters:** | 496,523 | |
| **Non-Trainable Parameters:** | 1,480 | |

### 3.4.2 Transfer Learning Models

Transfer learning models like DenseNet169 [31], DenseNet201 [32], InceptionResNetV2 [33], MobileNetV3Small [34], ResNet50V2 [35], VGG19 [36], and Xception [37] are indispensable in improving the detection of GI diseases across multiple classes. These models are pre-trained on large datasets, allowing them to extract meaningful image features with remarkable efficiency. By fine-tuning these models on limited data for the target task, they can effectively capture complex patterns and subtle details related to GI diseases. Over 14 million classifications across 1,000 categories (ImageNet) were used to train the pre-trained models. We combined the training of TL models with the RRELM classifier proposed in order to achieve precise classification results. As there are no prior studies on this dataset, we compared the novel LPDS-CNN model to TL approaches in terms of classification results and processing resources. This comparison includes performance, model parameters, layer sizes, as well as training and testing time. Following the initialization of these models, the final layers were changed by adding two FC layers with 500 and 200 nodes, respectively, to aid in detecting GI disorders. Figure 6, presents the overall illustration on TL with RRELM classifier.

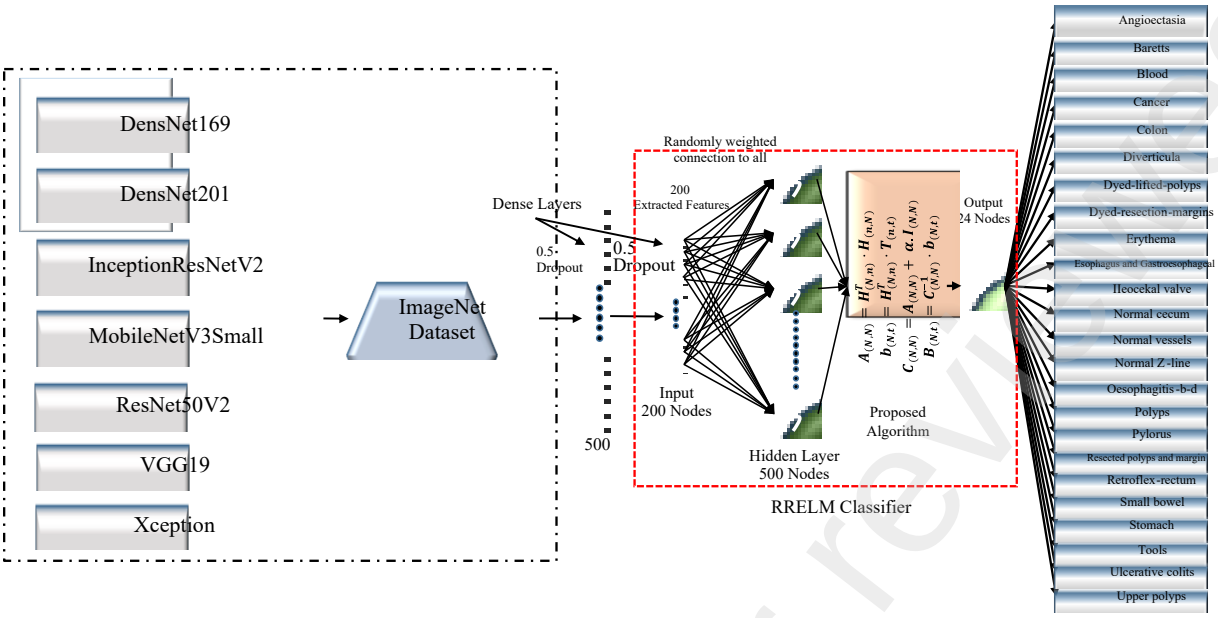



***Figure 6: The modified transfer learning architecture to classify GI diseases.***

The Visual Geometry Group (VGG) was a CNN architecture with several convolutional layers and filters. After each CL, an MP layer and a ReLU function aid in feature extraction. The last FC layer uses a SoftMax function with three FC layers.

A CNN architecture known as DenseNet has a unique feature called a dense connectivity pattern. All layers of this architecture received feature maps from the previous layers, which results in the smooth flow of information and establishes highly connected pathways throughout the network. This dense connectivity helps capture and utilize rich feature representations, leading to effective feature learning and improved performance. Different variants of DenseNet, including DenseNet-121, DenseNet-169, DenseNet-201, and DenseNet-264, differ in the number of layers. For instance, DenseNet-121 has 121 layers, while DenseNet-169 has 169 layers.

In 2016, Google created Xception, which builds upon the Inception architecture [37]. It utilizes DSCs, which filter each channel of the input feature map independently, and pointwise convolutions. This approach significantly reduces computational requirements and memory usage while maintaining accuracy. Due to its efficiency, Xception is often used for various computer vision tasks, particularly in cases where computational resources are limited.

The InceptionResNetV2 combines Inception, and ResNet designs that effectively extract features by utilizing inception modules and residual connections. On the other hand, the MobileNetV3Small is part of the MobileNet family that focuses on lightweight operations, making it ideal for real-time applications on devices with limited resources. The MobileNetV3Small is designed to balance model size, speed, and accuracy, while the InceptionResNetV2 is best suited for handling complex visual patterns due to its deep architecture and powerful feature learning.

### 3.5 Ridge Regression Extreme Learning Machine (RRELM)

The ELM (Extreme Learning Machine), introduced by Huang et al. [38], represents a remarkable paradigm shift in feature classification. It adopts a forward feed network approach based on supervised learning, constituting a groundbreaking innovation. Leveraging the capabilities of neural networks (NN), the ELM completely dispenses with the need for backpropagation, leading

to an astonishing thousand-fold increase in training speed. This novel approach has revolutionized the field of feature classification.

This new development has given the model exceptional abilities in classification and generalization. Specifically, the pseudo ELM (PELM) has shown remarkable proficiency in large-scale multi-class classification and outperformed latest ML models [39,40]. The PELM's significance lies in its innovative and flexible initialization of parameters between the input and hidden layers, which only uses one hidden layer. The parameters that connect the hidden and output layers are determined by using the pseudoinverse technique. However, this study introduced a new level of sophistication by replacing the pseudoinverse method with the ridge regression technique. This improvement significantly enhances the model's ability to learn features effectively and to regulate them, thus increasing its potential for generalization and achieving unprecedented accuracy compared to the PELM. The model's architecture consists of 200 nodes in the input layer and a formidable ensemble of 500 nodes in the hidden layer. Finally, twenty-four nodes created with the RRELM algorithm were crucial in classifying different types of samples from GI tract images in the output layer. RRELM classifier architecture is shown in Figure 5.

The explanation of RRELM algorithm is given below:

**Algorithm 1:** Proposed RRELM algorithm for multi-class classification.

$$X_{(n,m)} = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} & \cdots & x_{(1,m)} \\ x_{(2,1)} & x_{(2,2)} & \cdots & x_{(2,m)} \\ x_{(3,1)} & x_{(3,2)} & \cdots & x_{(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(n,1)} & x_{(n,2)} & \cdots & x_{(n,m)} \end{bmatrix} \quad Y_{(n,t)} = \begin{bmatrix} y_{(1,1)} & y_{(1,2)} & \cdots & y_{(1,t)} \\ y_{(2,1)} & y_{(2,2)} & \cdots & y_{(2,t)} \\ y_{(3,1)} & y_{(3,2)} & \cdots & y_{(3,t)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{(n,1)} & y_{(n,2)} & \cdots & y_{(n,t)} \end{bmatrix}$$

$$W_{(m,N)} = \begin{bmatrix} w_{(1,1)} & w_{(1,2)} & \cdots & w_{(1,N)} \\ w_{(2,1)} & w_{(2,2)} & \cdots & w_{(2,N)} \\ w_{(3,1)} & w_{(3,2)} & \cdots & w_{(3,N)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{(m,1)} & w_{(m,2)} & \cdots & w_{(m,N)} \end{bmatrix} \quad H_{(n,N)} = \begin{bmatrix} h_{(1,1)} & h_{(1,2)} & \cdots & h_{(1,N)} \\ h_{(2,1)} & h_{(2,2)} & \cdots & h_{(2,N)} \\ h_{(3,1)} & h_{(3,2)} & \cdots & h_{(3,N)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{(n,1)} & h_{(n,2)} & \cdots & h_{(n,N)} \end{bmatrix}$$

The input matrix is denoted as X, and the output matrix is represented as Y.

1. $W_{(m,N)}$ is denoted as input weight and $B_{(1,N)}$ is denoted as bias matrices.

$$B_{(1,N)} = \begin{bmatrix} b_{(1,1)} & b_{(1,2)} & \cdots & b_{(1,N)} \end{bmatrix}$$

2. The second step is to find the output $H_{(n,N)}$ of the hidden layer.

$$H_{(n,N)} = G(X_{(n,m)} \cdot W_{(m,N)} + B_{(1,N)})$$

Where, G denotes an activation function in this context.

3. Determine the output weight matrix $\beta_{(N,t)}$ using pseudoinverse

$$\beta_{(N,t)} = H^{\dagger}_{(N,n)} \cdot T_{(n,t)}$$

**In this proposed hybrid Ridge regression, the pseudoinverse has been replaced by these equations:**

$$\boldsymbol{A_{(N,N)} = H^{T}_{(N,n)} \cdot H_{(n,N)}}$$

$$\boldsymbol{b_{(N,t)} = H^{T}_{(N,n)} \cdot T_{(n,t)}}$$

$$\boldsymbol{C_{(N,N)} = A_{(N,N)} + \alpha . I_{(N,N)}}$$

$$\boldsymbol{B_{(N,t)} = C^{-1}_{(N,N)} \cdot b_{(N,t)}}$$

**Where, α denotes regularization parameters.**

4. Generate prediction $\beta_{(N,t)}$.

The model exuded certainty that it would generate exact final predictions. RRELM, a method that incorporated Ridge Regression into the ELM framework in a seamless approach, crossed a perfect equilibrium between potent feature learning and regularization. As a result, the model's ability to generalize and decipher complex patterns in the data improved, boosting its predictive power. Figure 5 illustrates the RRELM architecture together with the proposed LPDS-CNN model for multiclass classification.

### 3.6 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) is essential in accurately diagnosing gastrointestinal diseases across 24 different classes. For the first time, we have utilized SHAP (Shapley Additive Explanations) to address deep learning models' "black box" nature, often making them less practical. We aimed to make the inner workings of the LPDS-CNN model more transparent and interpretable by using SHAP. By integrating the LPDS-CNN model and SHAP for disease classification, endoscopists can now make more informed decisions in identifying diseases more efficiently and confidently [41]. It also aids the medical professionals to validate the model's predictions, detect errors, and mitigate biases to ensure accurate and fair diagnoses. This breakthrough opens up new possibilities for improved disease management strategies and more effective gastrointestinal disease classification treatments, spanning different classes. Shapley values used in the study to quantify the significance of individual pixels were shown to follow a distinct pattern. Red pixels contribute positively to correct class identification, whereas blue pixels have a negative effect, decreasing the probability of an accurate categorization [38]. These Shapley values were calculated using Equation (1).

$$\emptyset_k = \sum_{M \subseteq N \setminus k} \frac{M|!(A - |M| - 1)!}{A!} [f_x(M \cup k) - f_x(M)] \quad (1)$$

Where $f_x$ represents the change in output that is caused by the Shapely values of a particular feature, $k$. Subset $M$ consists of all features that are part of feature $N$, with the exception of feature

$k.\frac{M|!(A-|M|-1)!}{A!}$ denotes the weighted factor of the subset $M$ permutations. Equation (2) leads to the predicted outcome, which is represented by the symbol $f_x\,(M)$,

$$f_x\,(M) = P[f(x)|x_M] \tag{2}$$

The SHAP method involves the substitution of every initial identifiable ($x_k$) with a binary value ( $b'_k$) that denotes the presence or absence of $x_k$, as illustrated in Equation (3).

$$l\,(b') = \emptyset_0 + \sum_{k=1}^{A} \emptyset_k b'_k \tag{3}$$

In the proposed framework $f(x)$, the bias is denoted by $\emptyset_0$, and the feature's contribution is represented by $\emptyset_k b'_k$, where $l\,(b')$ represents the substitute model for the framework. The contribution of feature $k$ to the final outcome and the role of $\emptyset_k$ are essential aspects that aid in understanding the underlying mechanisms of the model.

### 3.7 Classification Matrices and Loss Function

In order to evaluate the classification performance of the RRELM model, we employed confusion matrices (CMs), which are widely recognized for their effectiveness in evaluation. We calculated several metrics using the equations, including accuracy, precision, recall, f1-score, and area under the curve (AUC) [43].

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{4}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{5}$$

$$Recall = \frac{T_P}{T_N + F_P} \tag{6}$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{7}$$

$$AUC = \frac{1}{2}\left(\frac{T_P}{T_P + F_N} + \frac{T_N}{T_N + F_P}\right) \tag{8}$$

Where, $T_P$ = True positive, $T_N$ = True negative, $F_P$ = False positive, and $F_N$ = False negative.

The cross-entropy formula compares the integer-based actual class label to the probability distribution projected by the model. It measures the difference between the actual and predicted labels to minimize cross-entropy loss as much as feasible. The sparse categorical cross-entropy loss is frequently used in DL instances such as image classification, mainly when multiple classes are involved [44]. The formula is:

$$L_{ce} = -\sum_{i=1}^{n} y_r \times \log(y_p) \quad (9)$$

Where, $n$ denotes the class number, truth label is defined as $y_r$, and $y_p$ as the probability.

## 4. Results and Discussions

This study considered twenty-four classes (trial-1) and thirteen classes (trial-2) classifications with the P-CNN and LPDS-CNN models to assess the performance.

### 4.1 Trial 1: Results for twenty-four class classification

#### 4.1.1 P-CNN-PELM and P-CNN-RRELM

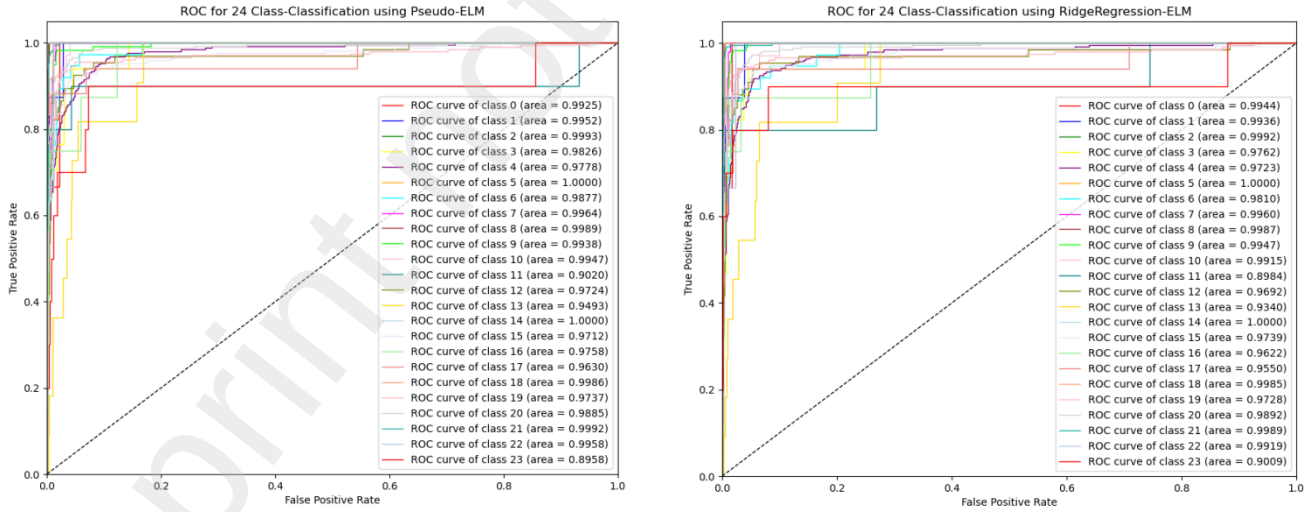
At first, we used the P-CNN model and trained it with a dataset of 6,400 images representing 24 different disease categories. These categories are depicted briefly in Figure 2 and listed in Table 1. The P-CNN (without DSC) was tested individually with a dataset of 1,600 test images. We conducted class-specific performance evaluations using both PELM and RRELM to assess the performance of the P-CNN. The results of these evaluations are presented in Table 3. The P-CNN-PELM had an average test precision of 64.54±0.36%, recall of 55.46±0.36%, and f1-score of 55.71±0.36%. The accuracy and AUC were 85.75% and 97.67%, respectively. In contrast, the average precision, recall, and f1-score using RRELM classifier on this model were 71.04±0.35% (9.15% improvement), 56.54±0.37% (1.91% improvement), and 58.33±0.33% (4.49% improvement), respectively. The average scores for accuracy and AUC were 85.81% (0.07% increase) and 97.93% (0.26% increase), respectively. P-CNN-RRELM can enhance traditional PELM and decrease the rate of misclassification. The class-wise ROC curves were visualized in Figure 7.

**Table 3: Twenty-four class performances by using P-CNN-ELM and P-CNN-RRELM architecture.**

| Class Name | P-CNN-PELM | | | P-CNN-RRELM | | |
|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | Precision | Recall | F1 |
| Angioectasia (0) | 0.00 | 0.00 | 0.00 | 1.00 | 0.33 | 0.50 |
| Baretts (1) | 0.50 | 0.25 | 0.33 | 1.00 | 0.38 | 0.55 |
| Blood (2) | 0.83 | 0.94 | 0.88 | 0.83 | 0.91 | 0.87 |
| Cancer (3) | 0.73 | 0.47 | 0.57 | 0.89 | 0.47 | 0.62 |
| Colon (4) | 0.80 | 0.87 | 0.83 | 0.81 | 0.87 | 0.84 |
| Diverticula (5) | 0.50 | 1.00 | 0.67 | 0.50 | 1.00 | 0.67 |
| Dyed-lifted-polyps (6) | 0.77 | 0.53 | 0.62 | 0.75 | 0.55 | 0.64 |
| Dyed-resection-margins (7) | 0.73 | 0.88 | 0.80 | 0.74 | 0.86 | 0.80 |
| Erythema (8) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Esophagus and Gastroesophageal junction (9) | 0.84 | 0.97 | 0.90 | 0.83 | 0.97 | 0.90 |
| Ileocekal valve (10) | 0.81 | 0.74 | 0.78 | 0.75 | 0.72 | 0.72 |
| Normal cecum (11) | 1.00 | 0.10 | 0.18 | 1.00 | 0.10 | 0.18 |
| Normal vessels (12) | 0.75 | 0.70 | 0.72 | 0.76 | 0.72 | 0.74 |
| Normal Z-line (13) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Oesophagitis-b-d (14) | 1.00 | 0.50 | 0.67 | 1.00 | 0.50 | 0.67 |
| Polyps (15) | 0.83 | 0.84 | 0.84 | 0.84 | 0.86 | 0.85 |
| Pylorus (16) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Resected polyps and margin (17) | 0.93 | 0.76 | 0.84 | 0.92 | 0.65 | 0.76 |
| Retroflex-rectum (18) | 0.71 | 0.83 | 0.77 | 0.69 | 0.75 | 0.72 |
| Small bowel (19) | 0.90 | 0.92 | 0.91 | 0.89 | 0.91 | 0.90 |
| Stomach (20) | 0.89 | 0.92 | 0.90 | 0.89 | 0.92 | 0.91 |
| Tools (21) | 0.97 | 0.99 | 0.98 | 0.96 | 1.00 | 0.98 |
| Ulcerative colits (22) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Upper polyps (23) | 1.00 | 0.10 | 0.18 | 1.00 | 0.10 | 0.18 |
| **Average (μ) ± SD (σ) (%)** | 64.54±0.36 | 55.46±0.36 | 55.71±0.36 | **71.04±0.35** | **56.54±0.37** | **58.33±0.33** |
| **Accuracy (%)** | 85.75 | | | **85.81** | | |
| **AUC (%)** | 97.67 | | | **97.93** | | |

*Bold values indicate best results.

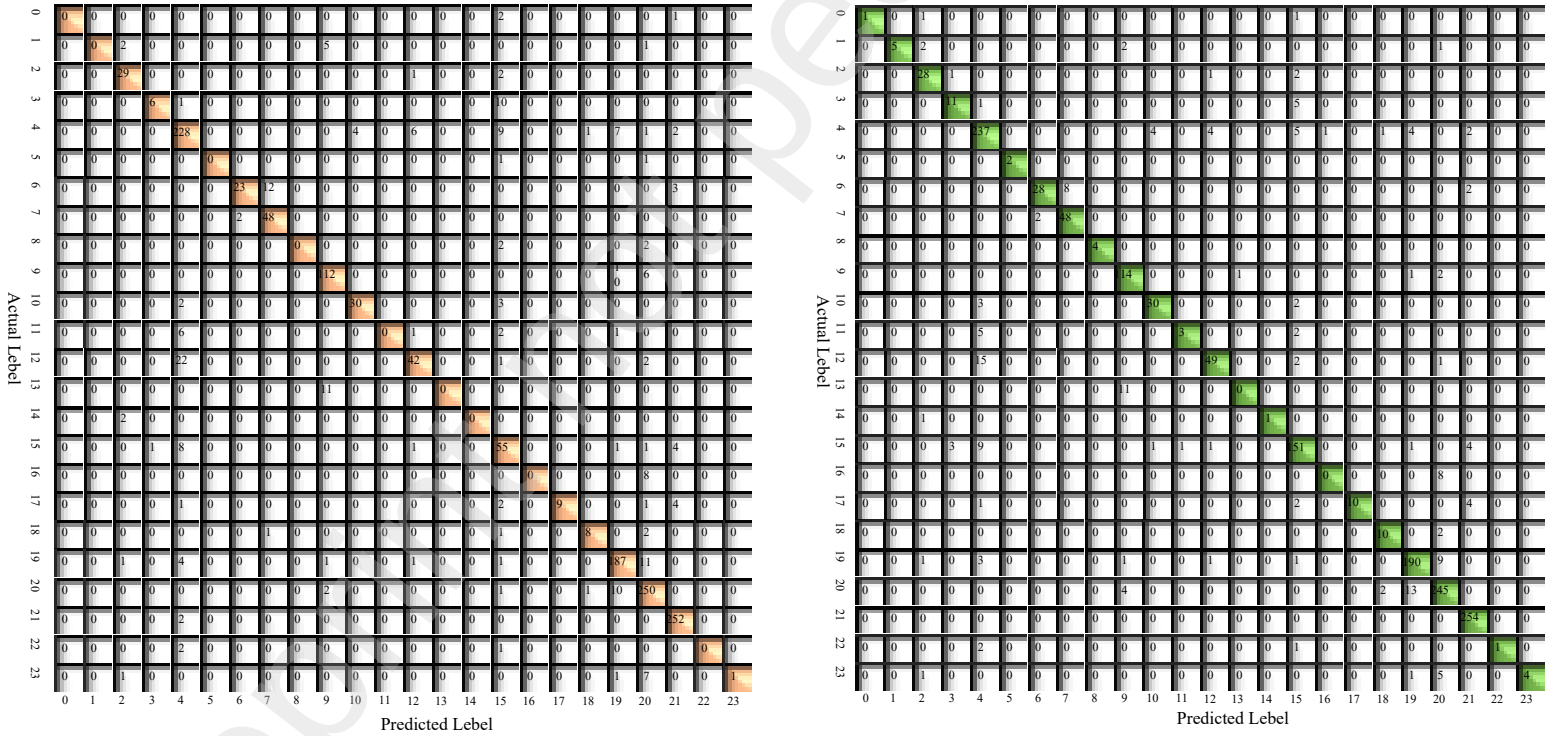



(A) (B)

***Figure 7. Class-wise ROCs on (A) P-CNN-PELM and (E) P-CNN-RRELM for twenty-four class classification.***

### 4.1.2 LPDS-CNN-PELM and LPDS-CNN-RRELM (proposed)

Table 2 demonstrates the significant performance disparities between PELM and RRELM within the proposed lightweight parallel depth-wise separable CNN (LPDS-CNN) framework. Twenty-four distinct GI disease datasets were used to validate the accuracy of the models and ensure their dependability. Figures 8A and 8B provide vital insights into the classification outcomes by displaying the confusion matrices for both classifiers. RRELM substantially decreased misclassification rates for Angioectasia (0), Baretts (1), Diverticula (5), and Ulcerative colits (22). The matrices compute class-specific precision, recall, and f1-scores. RRELM obtained an impressive average precision of 83.42±0.270% across multiple categories, whereas PELM achieved 54.46±0.434% (34.71% gain). In addition, RRELM attained 68.08±0.311% recall and 72.63±0.275% f1-score, respectively where 31.39% and 29.2% were the respective improvement from PELM. The accuracy score increased from 86.25% to 89.13%. The results show that RRELM effectively separates the multiclass from the samples. RRELM has the maximum AUC at 98.11%, exceeding PELM's AUC of 97.64%. These results provide solid evidence for the applicability of the proposed RRELM method. It helps medical professionals accurately diagnose multi-scale images, demonstrating significant performance. Finally, this analysis presents a novel lightweight model using an RRELM classifier that beats the conventional method on other models regarding performance and accuracy.




(A) (B)

***Figure 8. Confusion matrices of LPDS-CNN model with (A) PELM and (B) RRELM for twenty-four class classification.***

Figure 9 presents a comprehensive analysis of the LPDS-CNN-ELM and LPDS-CNN-RRELM models' performance in accurately differentiating 24 different gastrointestinal disease classes. The

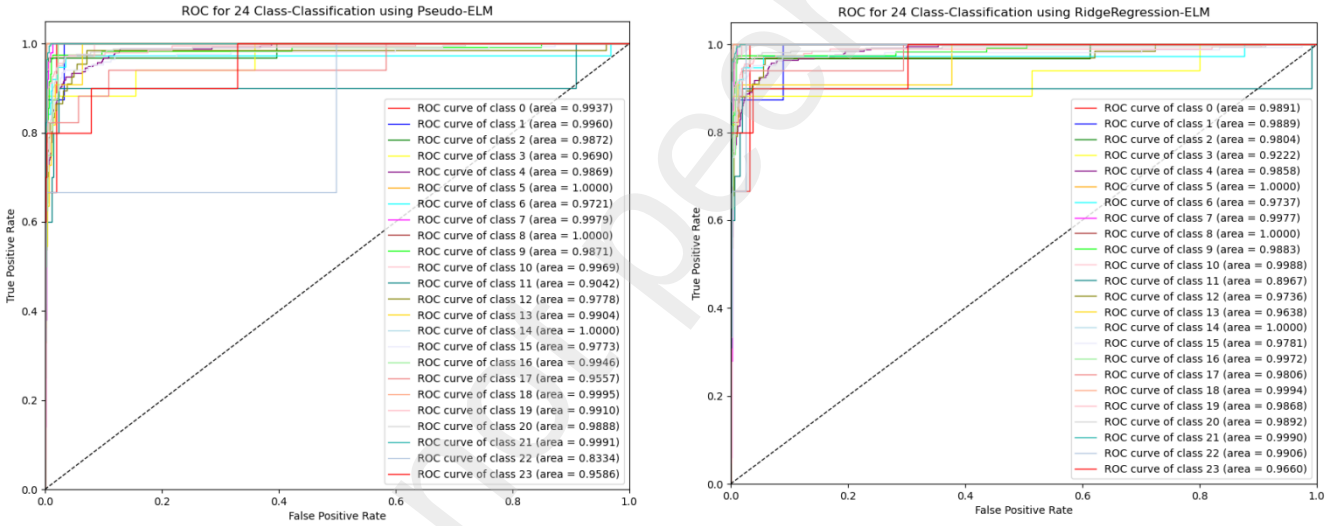
ROC curve for each category reveals that both models performed exceptionally well, with ROC values exceeding 97.74% for maximum classes. The average ROC score of 98.11% for the proposed RRELM integrated with the LPDS-CNN model highlights the model's outstanding accuracy and efficiency in classifying GI diseases. The results indicate that the LPDS-CNN-RRELM model is highly dependable and effective in its classification performance. It is worth noting that the proposed model achieved a 100% accuracy rate in identifying Diverticula (5), Erythema (8), and Oesophagitis-b-d (14), as demonstrated by the ROC values. This analysis showed that the proposed system accurately identifies these specific diseases. Additionally, the ROC curve significantly improved the detection of individuals with Ulcerative colitis (22) who belong to a lower scored class. This ROC improvement was seen as the detection rate increased from 83.34% to an impressive 99.06%. It is important to emphasize that our LPDS-CNN-RRELM method outperformed all other methods in class-wise detection. The dataset was not balanced due to a lack of available class data for each category in the GastroVision dataset. One of our key objectives when developing the model was to find a way to handle this data discrepancy efficiently. We ensured that each class had an equal say in the overall score, even if it could have been possible to get a more even distribution of samples throughout the analysis that were taken into classification. AUC scores varied from 0.9222 to 1.0000 at LPDS-CNN-RRELM across all classes, proving that better AUC results obtained even with imbalanced data sets (see Figure 9B). The model's recall rate for Diverticula (5) and Erythema (8) was 100%, indicating that the proposed framework can reliably differentiate all other gastrointestinal disorders. Our proposed framework has the potential to lessen significantly medical professional's liability and increase their assurance in correctly identifying such diseases.

**Table 4: Class-wise performance using LPDS-CNN-PELM and LPDS-CNN-RRELM.**

| Class Name | LPDS-CNN-ELM | | | LPDS-CNN-RRELM | | |
|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | Precision | Recall | F1 |
| Angioectasia (0) | 0.00 | 0.00 | 0.00 | 1.00 | 0.33 | 0.50 |
| Baretts (1) | 0.00 | 0.00 | 0.00 | 1.00 | 0.62 | 0.77 |
| Blood (2) | 0.83 | 0.91 | 0.87 | 0.90 | 0.88 | 0.89 |
| Cancer (3) | 0.86 | 0.35 | 0.50 | 0.73 | 0.65 | 0.69 |
| Colon (4) | 0.83 | 0.88 | 0.85 | 0.86 | 0.92 | 0.89 |
| Diverticula (5) | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| Dyed-lifted-polyps (6) | 0.92 | 0.61 | 0.73 | 0.93 | 0.74 | 0.82 |
| Dyed-resection-margins (7) | 0.79 | 0.96 | 0.86 | 0.86 | 0.96 | 0.91 |
| Erythema (8) | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| Esophagus and Gastroesophageal junction (9) | 0.85 | 0.95 | 0.90 | 0.86 | 0.97 | 0.91 |
| IIeocekal valve (10) | 0.88 | 0.86 | 0.87 | 0.86 | 0.86 | 0.86 |
| Normal cecum (11) | 0.00 | 0.00 | 0.00 | 0.75 | 0.30 | 0.43 |
| Normal vessels (12) | 0.81 | 0.63 | 0.71 | 0.88 | 0.73 | 0.80 |
| Normal Z-line (13) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Oesophagitis-b-d (14) | 0.00 | 0.00 | 0.00 | 1.00 | 0.50 | 0.67 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Polyps (15) | 0.80 | 0.91 | 0.85 | 0.87 | 0.88 | 0.88 |
| Pylorus (16) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Resected polyps and margin (17) | 1.00 | 0.53 | 0.69 | 1.00 | 0.59 | 0.74 |
| Retroflex-rectum (18) | 0.80 | 0.67 | 0.73 | 0.77 | 0.83 | 0.80 |
| Small bowel (19) | 0.90 | 0.91 | 0.91 | 0.90 | 0.92 | 0.91 |
| Stomach (20) | 0.85 | 0.95 | 0.90 | 0.90 | 0.93 | 0.91 |
| Tools (21) | 0.95 | 0.99 | 0.97 | 0.95 | 1.00 | 0.98 |
| Ulcerative colits (22) | 0.00 | 0.00 | 0.00 | 1.00 | 0.33 | 0.50 |
| Upper polyps (23) | 1.00 | 0.10 | 1.00 | 1.00 | 0.40 | 0.57 |
| **Average (μ) ± SD (σ) (%)** | 54.46±0.434 | 46.71±0.422 | 51.42±0.419 | **83.42±0.270** | **68.08±0.311** | **72.63±0.275** |
| **Accuracy (%)** | 86.25 | | | **89.13** | | |
| **AUC (%)** | 97.74 | | | **98.11** | | |

*Bold values indicate best results.




(A) (B)

***Figure 9. Class-wise ROCs on (A) LPDS-CNN-PELM and (E) LPDS-CNN-RRELM for twenty-four class classification.***

## 4.2 Trial 2: Results for thirteen class classification

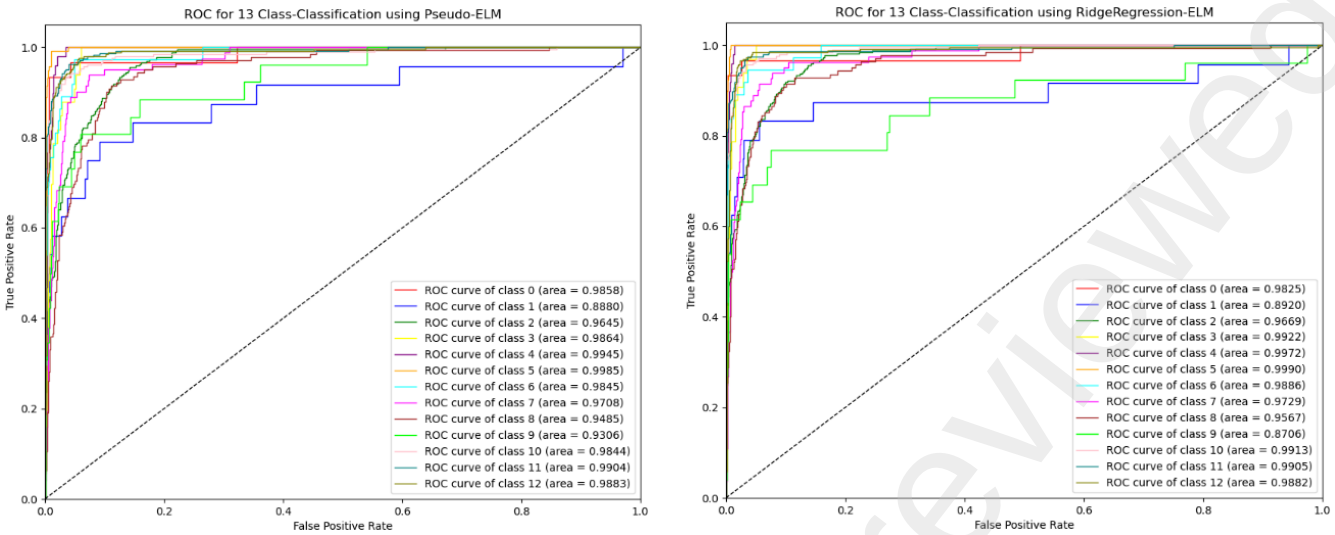
### 4.2.1 P-CNN-PELM and P-CNN-RRELM

For all thirteen classes, we used the same multiclass classification strategy. To ensure accuracy, we divided the data into categories based on the sample size of the training set. However, the model needed support in identifying the actual class label when there were insufficient samples. Therefore, we excluded several classes with 50 or fewer training samples from this investigation. The eliminated classes were Angioectasia, Baretts, Diverticula, Erythema, Normal Cecum, Normal Z-Line, Oesophagitis, B.D., Pylorus, Retroflex Rectus, Ulcerative Colitis, and Upper Polyps (see **Table 1**).

In trial 2, the P-CNN model was trained and tested on 6,108 and 1,528 images, respectively. The classification accuracy of the models is displayed in Table 5. Precision, recall, and f1-scores for the P-CNN-ELM model were 75.46±0.252%, 62.69±0.355%, and 63.23±0.316%, respectively. The effectiveness of the model was enhanced by employing RRELM. According to **Table 5**, the "Resected polyps and margin (9)" class was misclassified by P-CNN-PELM but correctly detected by P-CNN-RRELM. The P-CNN-RRELM model increased test accuracy by approximately 3.21% (from 81.02% to 83.71%). It earned the highest scores of 80.92±0.091% (6.75% improved), 75.15±0.205% (16.58% improved), and 76.77±0.156% (17.64% improved) in precision, recall, and f1-score, respectively. In the thirteen-class analysis, the AUC score of the RRELM classifier is reduced (97.04% to 96.83%). Class-wise ROC curve is visualized frequently on the P-CNN model for both classifiers in **Figure 10**.

**Table 5: Thirteen class performances by using P-CNN-PELM and PCNN-RRELM.**

| Class Name | P-CNN-PELM | | | P-CNN-RRELM | | |
|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **Precision** | **Recall** | **F1** |
| Blood (0) | 0.84 | 0.70 | 0.76 | 0.87 | 0.90 | 0.89 |
| Cancer (1) | 1.00 | 0.08 | 0.15 | 0.69 | 0.46 | 0.55 |
| Colon (2) | 0.71 | 0.86 | 0.78 | 0.77 | 0.82 | 0.80 |
| Dyed Lifted Polyps (3) | 0.80 | 0.12 | 0.21 | 0.71 | 0.61 | 0.66 |
| Dyed-resection-margins (4) | 0.68 | 0.92 | 0.78 | 0.78 | 0.94 | 0.85 |
| Esophagus and Gastroesophageal junction (5) | 0.94 | 0.96 | 0.95 | 0.94 | 0.96 | 0.95 |
| IIeocekal valve (6) | 0.78 | 0.57 | 0.66 | 0.84 | 0.70 | 0.76 |
| Normal vessels (7) | 0.76 | 0.45 | 0.56 | 0.74 | 0.61 | 0.67 |
| Polyps (8) | 0.61 | 0.73 | 0.66 | 0.67 | 0.73 | 0.70 |
| Resected polyps and margin (9) | 0.00 | 0.00 | 0.00 | 0.80 | 0.31 | 0.44 |
| Small bowel (10) | 0.93 | 0.88 | 0.90 | 0.92 | 0.87 | 0.89 |
| Stomach (11) | 0.88 | 0.93 | 0.90 | 0.88 | 0.93 | 0.90 |
| Tools (12) | 0.88 | 0.95 | 0.91 | 0.91 | 0.93 | 0.92 |
| **Average (μ) ± SD (σ) (%)** | 75.46±0.252 | 62.69±0.355 | 63.23±0.316 | **80.92±0.091** | **75.15±0.205** | **76.77±0.156** |
| **Accuracy (%)** | 81.02 | | | **83.71** | | |
| **AUC (%)** | **97.04** | | | 96.83 | | |

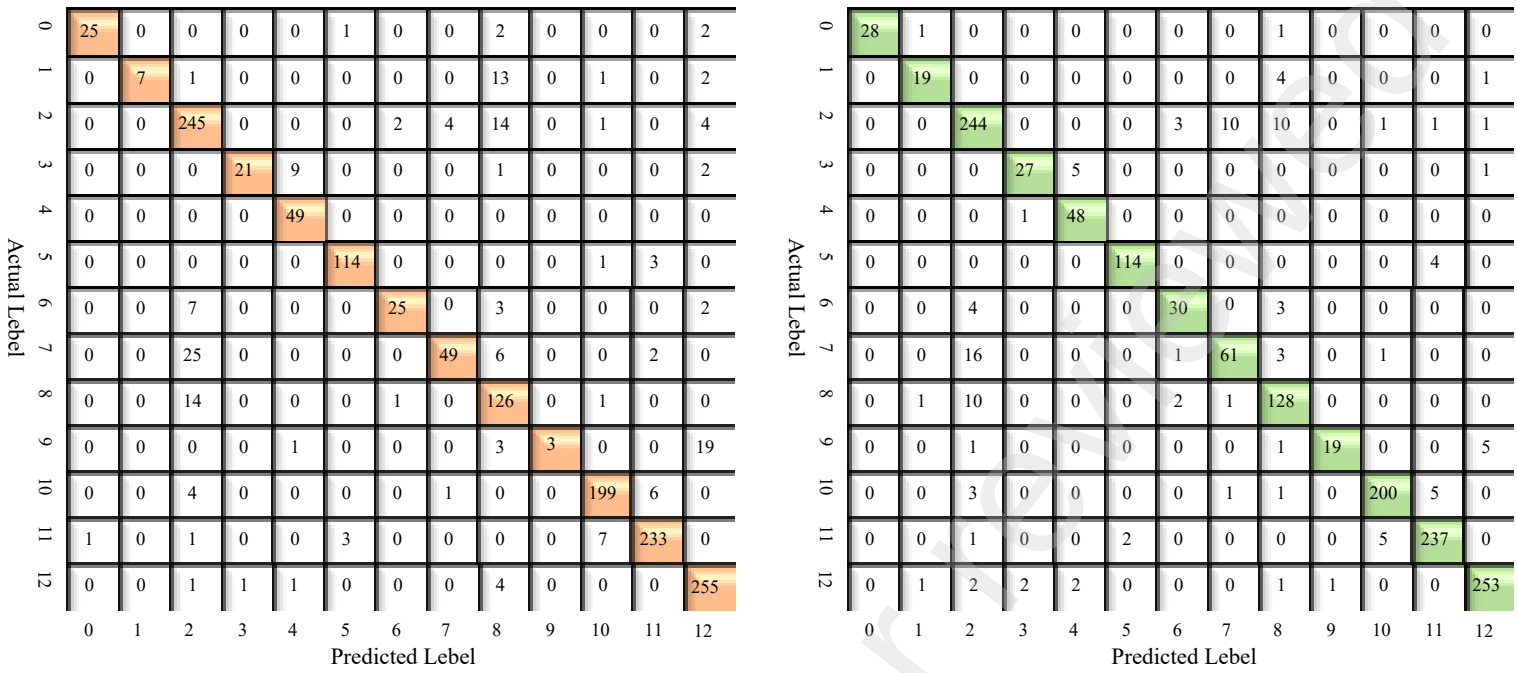



(A) (B)

***Figure 10. Class-wise ROCs on (A) P-CNN-PELM and (E) P-CNN-RRELM for thirteen-class classification.***

### 4.2.2 LPDS-CNN-PELM and LPDS-CNN-RRELM

Figure 11 holds crucial information regarding the classifiers' performance. The confusion matrices for both classifiers are depicted. It is evident that the RRELM classifier outperformed in numerous categories, such as Blood (0), Cancer (1), Colon (2), Dyed Lifted Polyps (3), Dyed-resection-margins (4), IIeocekal valve (6), Normal vessels (7), Polyps (8), Resected polyps and margin (9), and Tools (12). The precision, recall, and f1-scores were well-balanced and exceeded 80% for each category. The presented results strongly support the effectiveness of the RRELM method in accurately diagnosing multi-scale medical images. The proposed method significantly outperforms conventional models, indicating its superiority. Further analysis reveals a new, lightweight model utilizing an RRELM classifier, surpassing other models in performance and accuracy. These findings provide a solid basis for applying the RRELM method in medical image diagnosis. Across all 13 categories, PELM and RRELM achieved an impressive average precision of 91.08±0.062%. RRELM also attained a recall of 88.15±0.092% and an f1-score of 89.54±0.066%. Compared to PELM, RRELM showed improvements of 14.49% and 12.72%, respectively. The accuracy score also improved from 88.42% to 92.15%, representing a 4.05% increase. Figure 12 comprehensively analyzes the LPDS-CNN-PELM and LPDS-CNN-RRELM models' capacity to differentiate between 13 distinct GI diseases. The ROC curves for each category demonstrate the outstanding performance of both models, with ROC values exceeding 98% for most classes. In class 1, the PELM AUC score was 86.29%, but the RRELM classifier found an exceptional score of 95.17% (9.33% higher). Even though the GastroVision dataset was not balanced, each class identification record is satisfactory on the proposed LPDS-CNN-RRELM architecture. That highlights the significance of the proposed system in clinical practice.



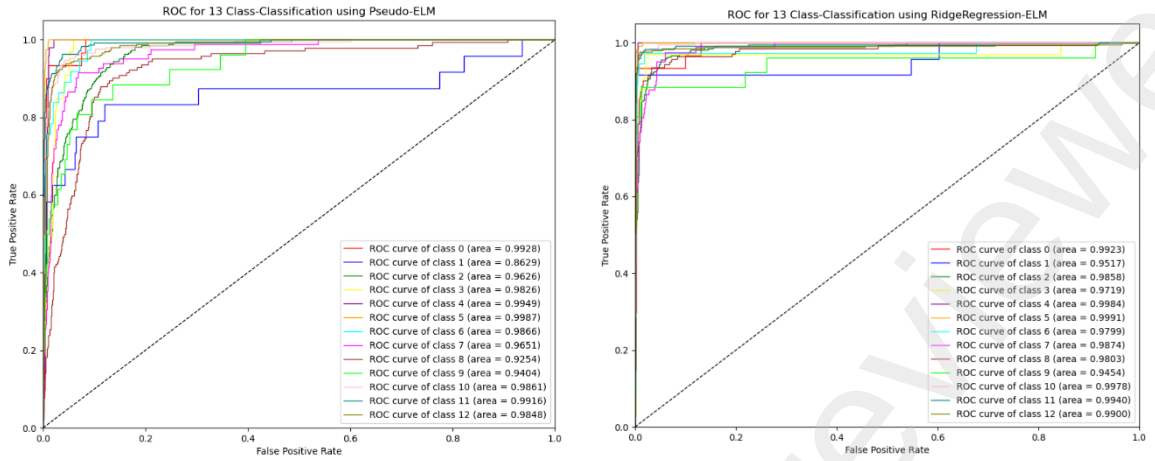

(A) (B)

***Figure 11. Confusion matrices of LPDS-CNN model with (A) PELM and (B) RRELM for thirteen classes.***

**Table 6: Class-wise performance using LPDS-CNN-PELM and LPDS-CNN-RRELM.**

| Class Name | LPDS-CNN-PELM | | | LPDS-CNN-RRELM | | |
|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **Precision** | **Recall** | **F1** |
| Blood (0) | 0.96 | 0.83 | 0.89 | 1.00 | 0.93 | 0.97 |
| Cancer (1) | 1.00 | 0.29 | 0.45 | 0.86 | 0.79 | 0.83 |
| Colon (2) | 0.82 | 0.91 | 0.86 | 0.87 | 0.9 | 0.89 |
| Dyed Lifted Polyps (3) | 0.95 | 0.64 | 0.76 | 0.9 | 0.82 | 0.86 |
| Dyed-resection-margins (4) | 0.82 | 1.00 | 0.9 | 0.87 | 0.98 | 0.92 |
| Esophagus and Gastroesophageal junction (5) | 0.97 | 0.97 | 0.97 | 0.98 | 0.97 | 0.97 |
| IIeocekal valve (6) | 0.89 | 0.68 | 0.77 | 0.83 | 0.81 | 0.82 |
| Normal vessels (7) | 0.91 | 0.6 | 0.72 | 0.84 | 0.74 | 0.79 |
| Polyps (8) | 0.73 | 0.89 | 0.8 | 0.84 | 0.9 | 0.87 |
| Resected polyps and margin (9) | 1.00 | 0.12 | 0.21 | 0.95 | 0.73 | 0.83 |
| Small bowel (10) | 0.95 | 0.95 | 0.95 | 0.97 | 0.95 | 0.96 |
| Stomach (11) | 0.95 | 0.95 | 0.95 | 0.96 | 0.97 | 0.96 |
| Tools (12) | 0.89 | 0.97 | 0.93 | 0.97 | 0.97 | 0.97 |
| **Average (μ) ± SD (σ) (%)** | 91.08±0.080 | 75.38±0.279 | 78.15±0.221 | **91.08±0.062** | **88.15±0.092** | **89.54±0.066** |
| **Accuracy (%)** | 88.42 | | | **92.15** | | |
| **AUC (%)** | **98.36** | | | 98.26 | | |

*Bold values indicate best results.




(A) (B)

***Figure 12. Class-wise ROCs on (A) LPDS-CNN-PELM and (E) LPDS-CNN-RRELM for fourteen-class classification.***

### 4.3 Performances comparisons among proposed model with other SOTA TL Models

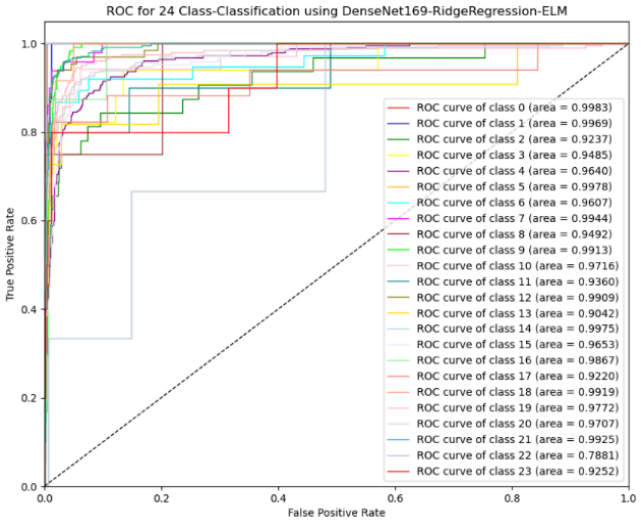
Table 7 presents the performance metrics comparisons of LPDS-CNN with other SOTA TL models with RRELM classifier in terms of overall 24 multiclass classification results. ResNet50V2, as previously stated, achieved the best classification results among the other five TL models for all twenty-four class. The mean precision, recall, f1-score, accuracy, and AUC scores were 67.25±0.403%, 54.71±0.386%, 57.92±0.378%, 86.89%, and 95.98%, respectively. Lowest model performance was achieved at MobileNetV3Small TL model. In case of DenseNet201 model achieved highest ROC 98% among the other TL models. But DenseNet169 achieved 96.01% (2.03% improvement). As shown in Table 13, the suggested LPDSCNN-RRELM achieved a reasonable accuracy of 89.13% for twenty-four class classification, approximately 2.51% higher than the ResNet50V2. Except for that, the AUC was 98.11%, which was higher than DenseNet201 (95.98%). In Figure 13, the TL model's ROC curve is presented for all twenty-four classes. Figure 14 provides a bar chart visualization of the overall model performances.

**Table 7. Twenty-four class classification performance of TL models and proposed model.**

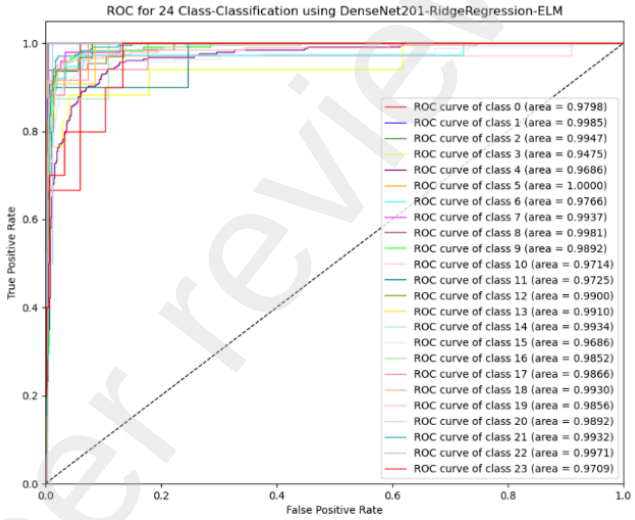
| Model Name | mP (%) | mR (%) | mF1 (%) | Accuracy (%) | AUC (%) |
|---|---|---|---|---|---|
| DenseNet169-RRELM | 46.95±0.445 | 36.83±0.407 | 38.29±0.396 | 81.44 | 96.01 |
| DenseNet201-RRELM | 56.58±0.451 | 45.13±0.421 | 47.17±0.418 | 86.93 | 98.00 |
| InceptionResNetV2-RRELM | 46.71±0.384 | 36.75±0.363 | 38.83±0.352 | 75.50 | 95.33 |
| MobileNetV3Small-RRELM | 16.58±0.245 | 15±0.28 | 14.33±0.246 | 48.63 | 72.38 |
| ResNet50V2-RRELM | 67.25±0.403 | 54.71±0.386 | 57.92±0.378 | 86.89 | 95.98 |
| VGG19-RRELM | 57.29±0.419 | 50.04±0.397 | 52.25±0.396 | 82.38 | 96.40 |
| Xception-RRELM | 50.46±0.44 | 37.17±0.404 | 39±0.889 | 83.56 | 95.64 |

| | | | | | |
|---|---|---|---|---|---|
| PCNN-PELM | 64.54±0.36 | 55.46±0.36 | 55.71±0.36 | 85.75 | 97.67 |
| LPDSCNN-PELM | 54.46±0.434 | 46.71±0.422 | 51.42±0.419 | 86.25 | 97.74 |
| PCNN-RRELM | 71.04±0.35 | 56.54±0.37 | 58.33±0.33 | 85.81 | 97.93 |
| LPDSCNN-RRELM | **83.42±0.270** | **68.08±0.311** | **72.63±0.275** | **89.13** | **98.11** |

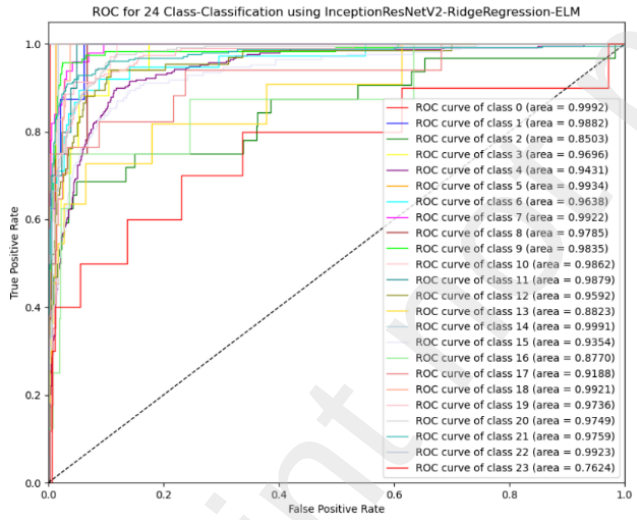
*Bold values indicate best results.



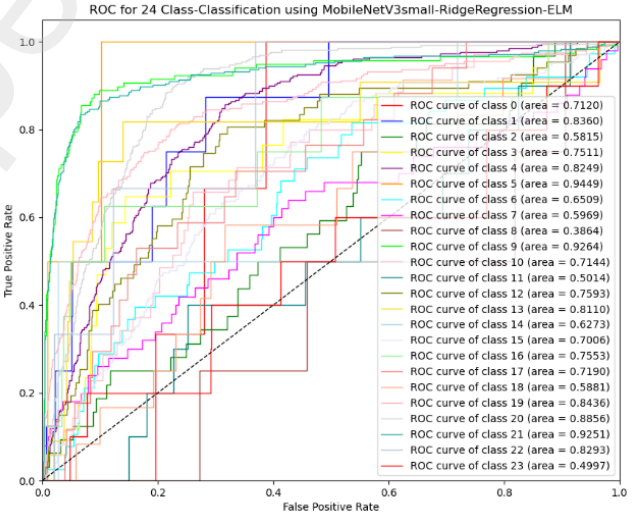

(A)



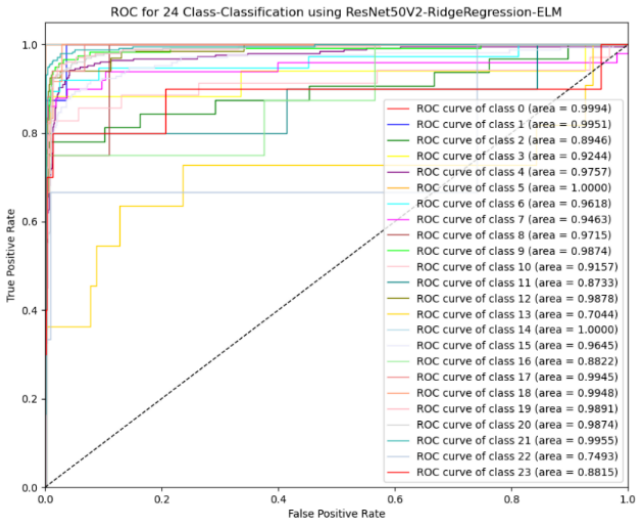

(B)



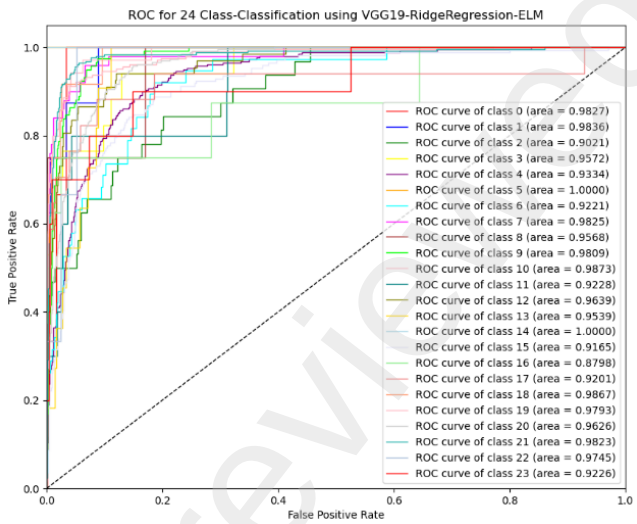

(C)



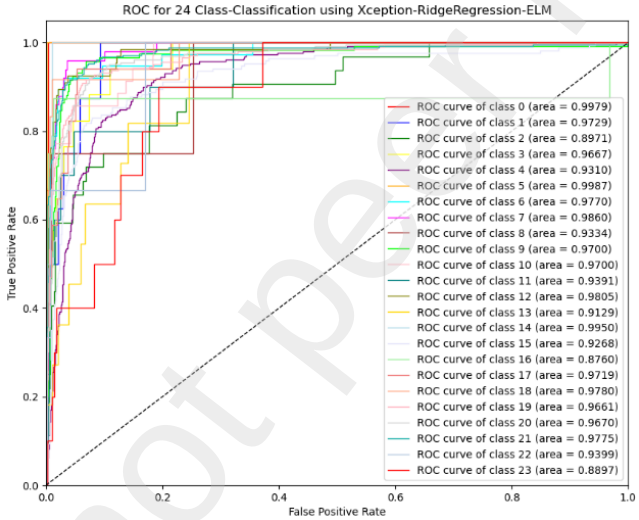

(D)




(E)




(F)




(G)

**Figure 13. Class-wise ROCs on (A) DensNet169, (B) DensNET201, (C) InceptionResNetV2, (D) MobileNetV3Small, (E) ResNet50V2, (F) VGG19, and (G) Xception with RRELM for twenty-four class classification.**

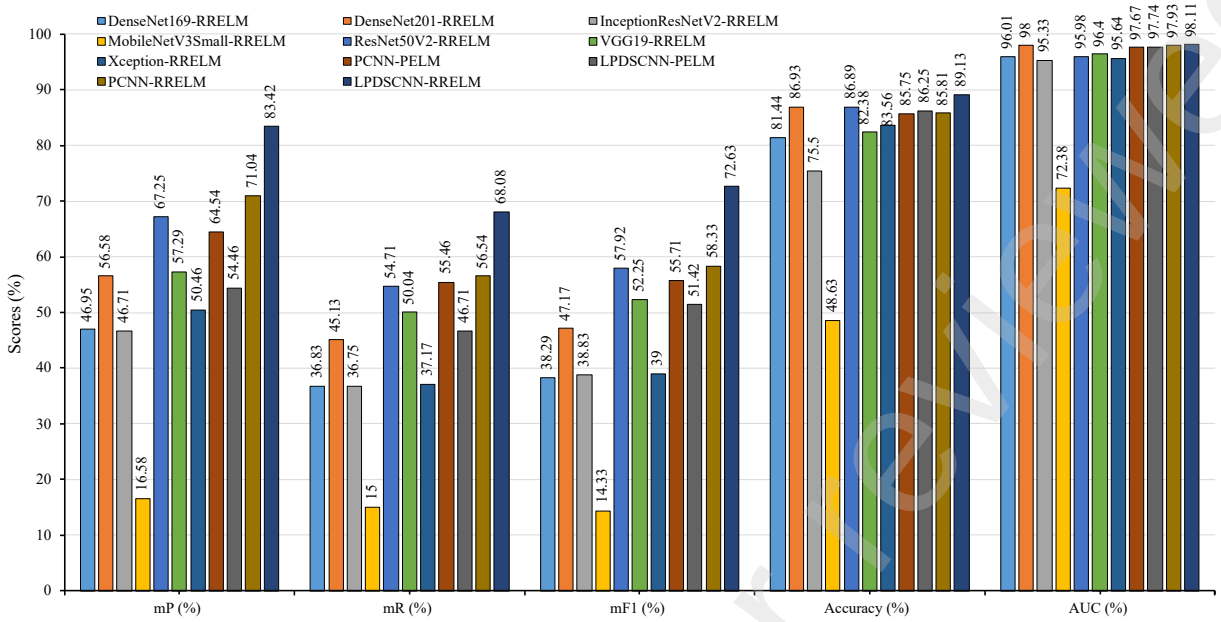



***Figure 14. Proposed and TL models performances on (A) DenseNet169, (B) DenseNET201, (C) InceptionResNetV2, (D) MobileNetV3Small, (E) ResNet50V2, (F) VGG19, (G) Xception, (H) P-CNN, and (I) LPDS-CNN with RRELM for twenty-four class classification.***

## 4.4 Computational time and resource comparison

The computational resources of LPDS-CNN, P-CNN, and other SOTA TL models are compared in Table 8 based on their parameters, layers, sizes, training time, and testing time. The LPDS-CNN model outperformed all other models regarding classification precision, training and testing duration, and model complexity, making it the most effective model. Figure 15 provides a distinct representation of the total amount of effort and resources required. The study results indicate that the LPDS-CNN-RRELM method is reliable and can effectively differentiate between various GI diseases.

The InceptionResNetV2 model has the most parameters of all the analyzed GI class datasets, with 55.57 million, 7,83 layers, and an overall size of 223 MB. Due to its compact size of 5.9 MB, 0.498 million parameters, and 9 CL, the LPDS-CNN model is highly efficient in loading and processing. This model has only 0.49 million trainable parameters, approximately 111.59 times less than the InceptionResNetV2 model and 42.07 times less than the P-CNN model, which has 20.95 million trainable parameters. Additionally, LPDS-CNN obtained a more balanced and less consecutive processing time, 0.0192 seconds for training and 0.002 seconds for testing.

Nevertheless, computational time could be more satisfactory for other TL models due to model complexity issues and performance factors. Despite having fewer layers and smaller sizes, the LPDS-CNN model can still precisely detect gastrointestinal diseases (from the result section) using fewer resources. The LPDCNN-RRELM model is an exceptional approach that delivers high performance while maintaining a small and resource-efficient design, making it suitable for various clinical-level applications.

***Table 8. Computational resources and time comparison for multiclass classifications.***

| Performance Criteria | P-CNN | LPDS-CNN | DenseNet 169 | DenseNet 201 | Inception ResNetV2 | MobileNet V3Small | ResNet 50V2 | VGG19 | Xception |
|---|---|---|---|---|---|---|---|---|---|
| Total Parameters (Million) | 20.95 | **0.498** | 15.64 | 21.78 | 55.57 | 4.81 | 30.12 | 20.95 | 27.42 |
| Trainable Parameters (Million) | 20.93 | **0.496** | 2.99 | 3.46 | 1.23 | 3.28 | 6.55 | 9.21 | 6.55 |
| Number of Layers | **9** | **9** | 598 | 710 | 783 | 246 | 193 | 25 | 36 |
| Size (Megabytes) | 24.1 | **5.9** | 84 | 111 | 223 | 44 | 165 | 87 | 24.1 |
| Training Time (24-class) | 0.0748 | 0.0192 | 0.0130 | **0.0089** | 0.0628 | 0.0528 | 0.0099 | 0.0538 | 0.0100 |
| Testing Time (24-class) | 0.0100 | **0.0020** | **0.0020** | 0.0030 | 0.0090 | 0.0080 | 0.0030 | 0.0080 | 0.0030 |

*Bold values indicate best results.

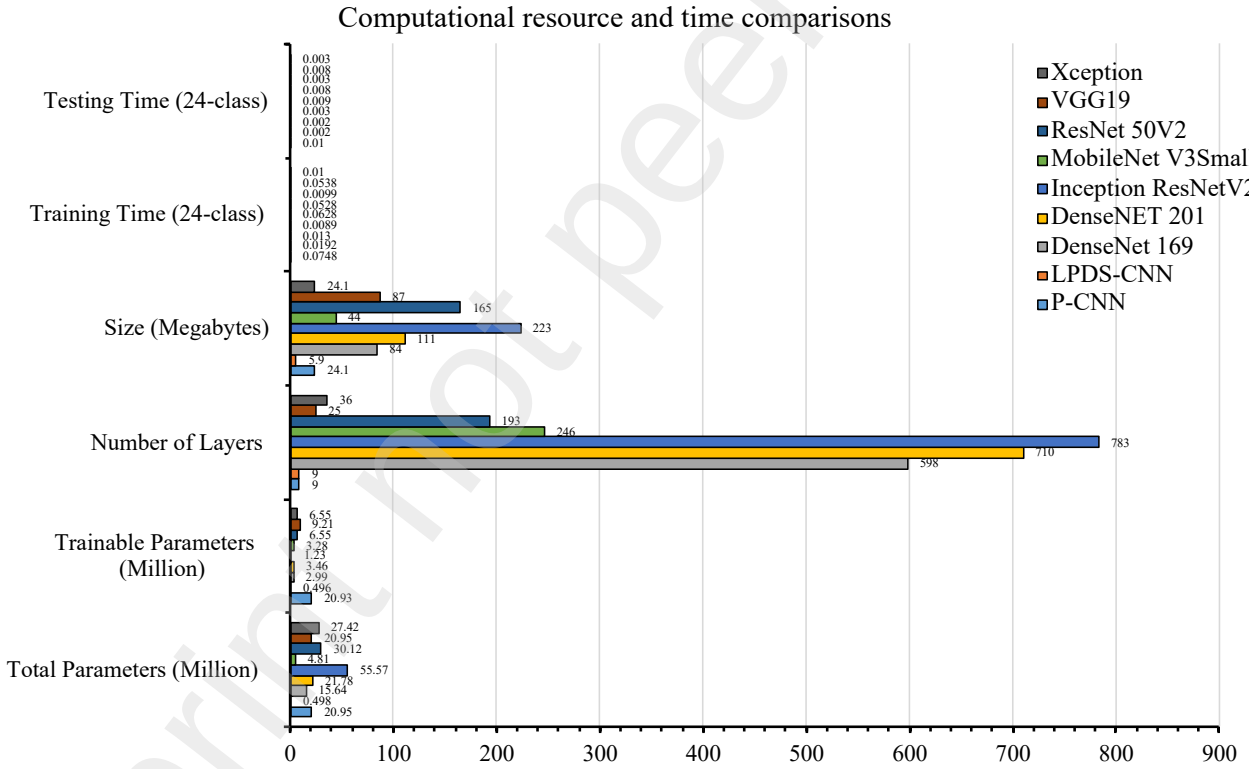



**Figure 15. Computational resource and time comparisons among the proposed LPDS-CNN-RRELM and other SOTA TL models.**

## 5. Interpretability LPDS-CNN-RRELM using SHAP

Methodical consideration of all possible combinations of gastrointestinal features led to the design of Shapley values, resulting in pixelated representations [45]**.** A clear trend emerged from the research, with red pixels showing powerful performance in accurately diagnosing specific class

identification. Conversely, the presence of blue pixels indicated a higher probability of being distant from the target class. In Figure 16, the SHAP results provided medical experts with explanation images for different classifications, including all the GI 13 cases. It is important to note that the SHAP explanation graphics have faint grey backgrounds merged with the original input images. In the top row, red pixels in the SHAP explanation image indicated the presence of Dyed-resection-margins (4). Conversely, the absence of blue pixels and fewer red pixels effectively eliminated other class categories. Remarkably, the second row exhibited a distinct pattern, whereby red pixels in the SHAP explanation images indicated the Esophagus and Gastroesophageal junction (5) class. On the other hand, a surplus of red pixels in the SHAP explanation image accurately identified the image as belonging to that class. In the third row, the SHAP explanation image's intense concentration of red pixels represents the Normal vessels (7) class. However, the presence of blue pixels in the SHAP explanation images for other classes confirmed the absence of probability. In the SHAP explanation image, the presence of red pixels in the fourth row indicated significant evidence of Polyps (8) GI class disease. Furthermore, other rows ($5^{th}$, $6^{th}$, and $7^{th}$) also accurately identified disease classes by highlighting red pixels in specific regions. These visual SHAP explanations confirmed the model's results, providing doctors with a more comprehensive understanding of the different strains of classification.

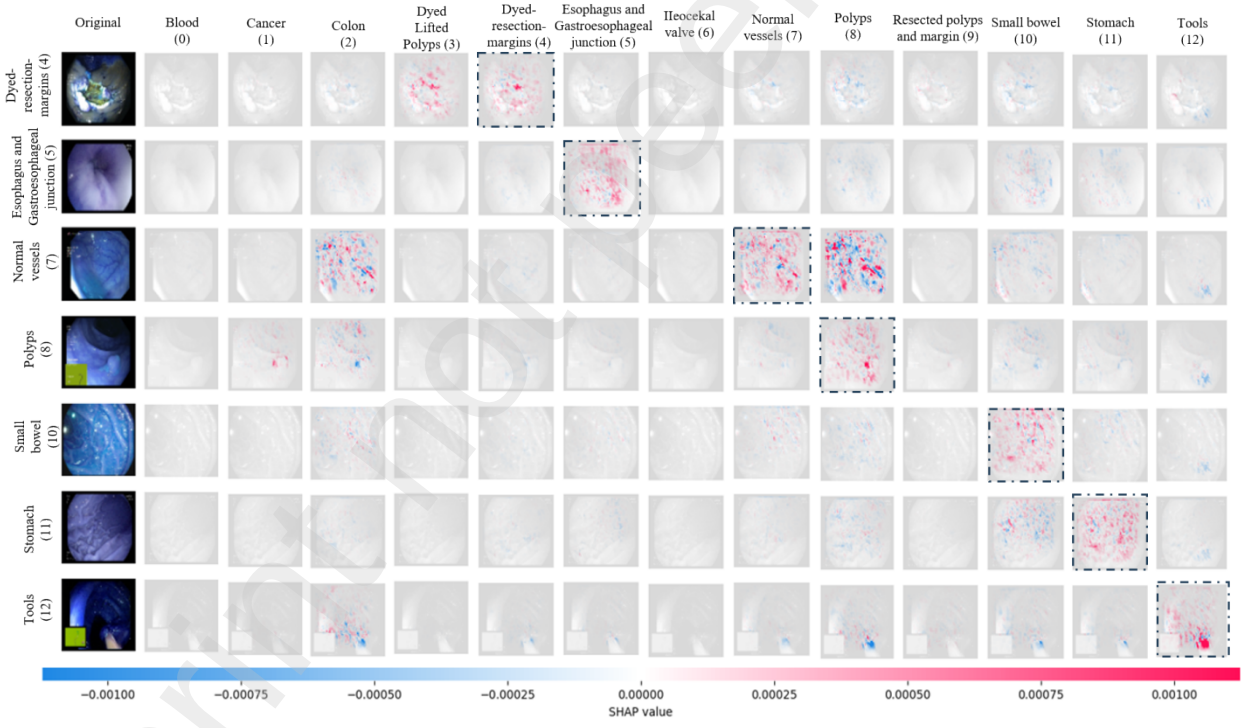



***Figure 16. SHAP explanation images for LPDS-CNN-RRELM.***

## 6. Conclusion

Using a combination of the LPDS-CNN model and the RRELM classifier, the presented research introduced a novel method for reliably classifying various GI diseases. Data augmentation and balance class images were not utilized to evaluate the worst-case performance of the model. Incorporating a significant image preprocessing method known as CLAHE, inpainting for artefact removal, sharpening, and Gaussian blur techniques makes it possible to improve the appearance

and accessibility of essential features while reducing unwanted noise. With only 0.498 parameters and nine layers, the proposed LPDS-CNN architecture efficiently accumulates infected-specific patterns with minimal computational burden. The classification performance is significantly enhanced by incorporating the hybrid ridge regression ELM model, which replaces pseudoinverse with ridge regression. This cutting-edge method obtains remarkable accuracy levels of 89.13% for 24 class classifications and 92.15 % for 13 class classifications. Precision, recall, and f1-scores values of 83.420.270%, 68.080.311%, and 72.630.275% with an outstanding AUC score of 98.11%, respectively, demonstrate that the LPDS-CNN-RRELM model is extraordinarily reliable in recognizing various cases. This significant advance in accuracy demonstrates the instrument's potential as an effective method for simultaneously identifying various patients. In addition, the model's minimal computing resource requirements enable seamless implementation on low-powered devices, thereby increasing its accessibility and applicability in real-world healthcare settings. SHAP's incorporation greatly benefits medical doctors and endoscopists because it facilitates a reliable interpretation of the model's results. The openness and interpretability of SHAP increase confidence in categorization results, resulting in improved decisions and more targeted treatments.

In conclusion, the proposed LPDS-CNN-RRELM method enhances the classification accuracy of GI diseases significantly and is easy to implement and interpret. This study proves that early detection and successful treatment of upper, lower, and both tract diseases can improve patients' health and save lives.

**Acknowledgement:** This work was supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R40), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The work was also supported by the University of Aberdeen, UK through its SCNU Joint Institute of AI and Data Science.

**Data Availability:** Data will be made available on request.

## References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries., CA. Cancer J. Clin. 71 (2021) 209–249. https://doi.org/10.3322/caac.21660.

[2] M. Arnold, C.C. Abnet, R.E. Neale, J. Vignat, E.L. Giovannucci, K.A. McGlynn, F. Bray, Global Burden of 5 Major Types of Gastrointestinal Cancer., Gastroenterology. 159 (2020) 335-349.e15. https://doi.org/10.1053/j.gastro.2020.02.068.

[3] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA. Cancer J. Clin. 68 (2018) 394–424. https://doi.org/10.3322/caac.21492.

[4] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2015, CA. Cancer J. Clin. 65 (2015)

5–29. https://doi.org/10.3322/CAAC.21254.
[5] N.H. Kim, Y.S. Jung, W.S. Jeong, H.J. Yang, S.K. Park, K. Choi, D. Il Park, Miss rate of colorectal neoplastic polyps and risk factors for missed polyps in consecutive colonoscopies, Intest. Res. 15 (2017) 411–418. https://doi.org/10.5217/IR.2017.15.3.411.
[6] M.A. Khan, M.S. Sarfraz, M. Alhaisoni, A.A. Albesher, S. Wang, I. Ashraf, StomachNet: Optimal Deep Learning Features Fusion for Stomach Abnormalities Classification, IEEE Access. 8 (2020) 197969–197981. https://doi.org/10.1109/ACCESS.2020.3034217.
[7] R. Saraiva, M. Perkusich, L. Silva, H. Almeida, C. Siebra, A. Perkusich, Early diagnosis of gastrointestinal cancer by using case-based and rule-based reasoning, Expert Syst. Appl. 61 (2016) 192–202. https://doi.org/https://doi.org/10.1016/j.eswa.2016.05.026.
[8] P. Aruna, N. Puviarasan, B. Palaniappan, Diagnosis of gastrointestinal disorders using DIAGNET, Expert Syst. Appl. 32 (2007) 329–335. https://doi.org/https://doi.org/10.1016/j.eswa.2005.11.039.
[9] C. Dimoulas, G. Kalliris, G. Papanikolaou, V. Petridis, A. Kalampakas, Bowel-sound pattern analysis using wavelets and neural networks with application to long-term, unsupervised, gastrointestinal motility monitoring, Expert Syst. Appl. 34 (2008) 26–41. https://doi.org/https://doi.org/10.1016/j.eswa.2006.08.014.
[10] M.M. Awais, S.K. Awan, Gastro-intestinal tract inspired computational model for myocardial infarction diagnosis, Expert Syst. Appl. 38 (2011) 5633–5641. https://doi.org/https://doi.org/10.1016/j.eswa.2010.10.072.
[11] B. Li, M.Q.-H. Meng, Automatic polyp detection for wireless capsule endoscopy images, Expert Syst. Appl. 39 (2012) 10952–10958. https://doi.org/https://doi.org/10.1016/j.eswa.2012.03.029.
[12] J. Song, M. Patel, A. Girgensohn, C. Kim, Combining deep learning with geometric features for image-based localization in the Gastrointestinal tract, Expert Syst. Appl. 185 (2021) 115631. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115631.
[13] L.H. Eadie, C.B. Reid, A.J. Fitzgerald, V.P. Wallace, Optimizing multi-dimensional terahertz imaging analysis for colon cancer diagnosis, Expert Syst. Appl. 40 (2013) 2043–2050. https://doi.org/https://doi.org/10.1016/j.eswa.2012.10.019.
[14] J.H. Lee, Y.J. Kim, Y.W. Kim, S. Park, Y. i. Choi, Y.J. Kim, D.K. Park, K.G. Kim, J.W. Chung, Spotting malignancies from gastric endoscopic images using deep learning, Surg. Endosc. 33 (2019) 3790–3797. https://doi.org/10.1007/S00464-019-06677-2/METRICS.
[15] B. Li, M.Q.-H. Meng, Computer-aided detection of bleeding regions for capsule endoscopy images., IEEE Trans. Biomed. Eng. 56 (2009) 1032–1039. https://doi.org/10.1109/TBME.2008.2010526.
[16] F. Noya, M.A. Álvarez-González, R. Benítez, Automated angiodysplasia detection from wireless capsule endoscopy, in: 2017 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2017: pp. 3158–3161. https://doi.org/10.1109/EMBC.2017.8037527.
[17] D.K. Iakovidis, A. Koulaouzidis, Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software, Gastrointest. Endosc. 80 (2014) 877–883. https://doi.org/10.1016/j.gie.2014.06.026.
[18] M.A. Usman, G.B. Satrya, M.R. Usman, S.Y. Shin, Detection of small colon bleeding in wireless capsule endoscopy videos., Comput. Med. Imaging Graph. Off. J. Comput. Med. Imaging Soc. 54 (2016) 16–26. https://doi.org/10.1016/j.compmedimag.2016.09.005.
[19] B. Li, M.Q.-H. Meng, Texture analysis for ulcer detection in capsule endoscopy images, Image Vis. Comput. 27 (2009) 1336–1342.

https://doi.org/https://doi.org/10.1016/j.imavis.2008.12.003.
[20] G. Pan, G. Yan, X. Qiu, J. Cui, Bleeding Detection in Wireless Capsule Endoscopy Based on Probabilistic Neural Network, J. Med. Syst. 35 (2011) 1477–1484. https://doi.org/10.1007/s10916-009-9424-0.
[21] J.-Y. Yeh, T.-H. Wu, W.-J. Tsai, J.-Y. Yeh, T.-H. Wu, W.-J. Tsai, Bleeding and Ulcer Detection Using Wireless Capsule Endoscopy Images, J. Softw. Eng. Appl. 7 (2014) 422–432. https://doi.org/10.4236/JSEA.2014.75039.
[22] Y. Yuan, J. Wang, B. Li, M.Q.-H. Meng, Saliency Based Ulcer Detection for Wireless Capsule Endoscopy Diagnosis, IEEE Trans. Med. Imaging. 34 (2015) 2046–2057. https://doi.org/10.1109/TMI.2015.2418534.
[23] K. Pogorelov, K.R. Randel, C. Griwodz, S.L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P.T. Schmidt, M. Riegler, P. Halvorsen, KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection, in: Proc. 8th ACM Multimed. Syst. Conf., Association for Computing Machinery, New York, NY, USA, 2017: pp. 164–169. https://doi.org/10.1145/3083187.3083212.
[24] S. Jain, A. Seal, A. Ojha, A. Yazidi, J. Bures, I. Tacheci, O. Krejcar, A deep CNN model for anomaly detection and localization in wireless capsule endoscopy images, Comput. Biol. Med. 137 (2021) 104789. https://doi.org/https://doi.org/10.1016/j.compbiomed.2021.104789.
[25] L. Lan, C. Ye, Recurrent generative adversarial networks for unsupervised WCE video summarization, Knowledge-Based Syst. 222 (2021) 106971. https://doi.org/https://doi.org/10.1016/j.knosys.2021.106971.
[26] M.N.A.M.I.A.T.K. Mousa Alhajlah Muhammad Nouman Noor, Gastrointestinal Diseases Classification Using Deep Transfer Learning and Features Optimization, Comput. Mater. \& Contin. 75 (2023) 2227–2245. https://doi.org/10.32604/cmc.2023.031890.
[27] S. Mohapatra, G. Kumar Pati, M. Mishra, T. Swarnkar, Gastrointestinal abnormality detection and classification using empirical wavelet transform and deep convolutional neural network from endoscopic images, Ain Shams Eng. J. 14 (2023) 101942. https://doi.org/https://doi.org/10.1016/j.asej.2022.101942.
[28] D. Jha, V. Sharma, N. Dasu, N.K. Tomar, S. Hicks, M.K. Bhuyan, P.K. Das, M.A. Riegler, P. Halvorsen, T. de Lange, U. Bagci, GastroVision: A Multi-class Endoscopy Image Dataset for Computer Aided Gastrointestinal Disease Detection, (2023). http://arxiv.org/abs/2307.08140.
[29] OSF | GastroVision, (n.d.). https://osf.io/84e7f/ (accessed August 27, 2023).
[30] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Commun. ACM. 60 (2017) 84–90. https://doi.org/10.1145/3065386.
[31] R. Tri Wahyuningrum, A. Kusumaningsih, W. Putra Rajeb, I.K. Eddy Purnama, Classification of Corn Leaf Disease Using the Optimized DenseNet-169 Model, in: Proc. 2021 9th Int. Conf. Inf. Technol. IoT Smart City, 2021: pp. 67–73.
[32] C. Zhao, R. Shuai, L. Ma, W. Liu, D. Hu, M. Wu, Dermoscopy image classification based on StyleGAN and DenseNet201, Ieee Access. 9 (2021) 8659–8679.
[33] Y. Bhatia, A. Bajpayee, D. Raghuvanshi, H. Mittal, Image captioning using Google's inception-resnet-v2 and recurrent neural network, in: 2019 Twelfth Int. Conf. Contemp. Comput., 2019: pp. 1–6.

[34] S. Maheta, Manisha, Deep Learning-Based Cancelable Biometric Recognition Using MobileNetV3Small Model, (2023) 347–356. https://doi.org/10.1007/978-981-99-1203-2_29/COVER.
[35] M. Rahimzadeh, A. Attar, A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2, Informatics Med. Unlocked. 19 (2020) 100360. https://doi.org/https://doi.org/10.1016/j.imu.2020.100360.
[36] T.R.G. V. Sudha, A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning, Comput. Mater. \& Contin. 66 (2021) 827–842. https://doi.org/10.32604/cmc.2020.012008.
[37] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Institute of Electrical and Electronics Engineers Inc., 2017: pp. 1251–1258. https://doi.org/10.1109/CVPR.2017.195.
[38] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, in: 2017 IEEE Conf. Comput. Vis. Pattern Recognit., 2017: pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243.
[39] M. Nahiduzzaman, M. Robiul Islam, M. Omaer Faruq Goni, M. Shamim Anower, M. Ahsan, J. Haider, M. Kowalski, Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier, Expert Syst. Appl. 217 (2023) 119557. https://doi.org/https://doi.org/10.1016/j.eswa.2023.119557.
[40] M. Nahiduzzaman, M.O.F. Goni, R. Hassan, M.R. Islam, M.K. Syfullah, S.M. Shahriar, M.S. Anower, M. Ahsan, J. Haider, M. Kowalski, Parallel CNN-ELM: A multiclass classification of chest X-ray images to identify seventeen lung diseases including COVID-19., Expert Syst. Appl. 229 (2023) 120528. https://doi.org/10.1016/j.eswa.2023.120528.
[41] S.M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Adv. Neural Inf. Process. Syst., Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.
[42] M. Bhandari, T.B. Shahi, B. Siku, A. Neupane, Explanatory Classification of CXR Images into COVID-19, Pneumonia and Tuberculosis Using Deep Learning and XAI, Comput. Biol. Med. 150 (2022). https://doi.org/10.1016/j.compbiomed.2022.106156.
[43] D.M.W. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, ArXiv Prepr. ArXiv2010.16061. (2020).
[44] B.N. Chaithanya, T.J. Swasthika Jain, A. Usha Ruby, A. Parveen, An approach to categorize chest X-ray images using sparse categorical cross entropy, Indones. J. Electr. Eng. Comput. Sci. (2021) 1700–1710.
[45] H. Chen, X. Chen, H. Tao, Z. Li, X. Wang, Low-Rank Representation with Adaptive Dimensionality Reduction via Manifold Optimization for Clustering, ACM Trans. Knowl. Discov. Data. 17 (2023). https://doi.org/10.1145/3589767.