**Detection of various Gastrointestinal Tract Diseases through Deep Learning method with Ensemble ELM and Explainable AI**

**Md. Faysal Ahamed1, Md. Nahiduzzaman2, Md. Rabiul Islam1, Mansura Naznine1, Mohamed Arselene Ayari3, Amith Khandakar3, and Julfikar Haider4**

*1Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh.*

*2Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh.*

*3Department of Electrical Engineering, Qatar University, Doha 2713, Qatar.*

*4Department of Engineering, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK*

**Corresponding Author:**

**Abstract**

The worldwide rising prevalence of gastrointestinal (GI) tract disorders highlights the urgent need for precise diagnosis, as these diseases greatly affect human life and contribute to high mortality rates. Fast identification, accurate classification, and efficient treatment approaches are essential to deal with this critical health issue. The current study intends to develop a deep learning (DL)-based approach that automatically classifies the GI tract diseases. For the first time, GastroVision dataset with 8000 images of 27 different GI diseases were utilized in this work to design a computer-aided diagnosis (CAD) system. This study presents a novel lightweight feature extractor with compact size and minimum layers named Parallel Depthwise Separable Convolutional Neural Network (PD-CNN) and a Pearson Correlation Coefficient (PCC) as the feature selector. Furthermore, a robust classifier named Ensemble Extreme Learning Machine (EELM), combined with pseudo inverse ELM (ELM) and L1 Regularized ELM (RELM) has been proposed to identify the diseases more precisely. A hybrid preprocessing technique, including scaling, normalization, and image enhancement techniques such as erosion, CLAHE, sharpening, and gaussian filters, are employed to enhance the image representation and improve classification performance. The proposed approach consists of twenty-four layers and only 0.815 million parameters with 9.79 MB model size. The proposed PD-CNN-PCC-EELM significantly extracts features, reduces computational overhead, and achieves excellent classification performance on the multiclass GI images. The PD-CNN-PCC-EELM achieves highest accuracy, precision, recall, f1, ROC-AUC, and AUC-PR values are 88.12±0.332%, 87.75±0.348%, 87.12±0.324%, 87.75%, 98.89%, and 92%, respectively with maintaining minimum testing time of 0.000001 seconds. A comparative study utilizes various State-of-the-art (SOTA) transfer learning (TL) models as feature extractors. Then, PCC and EELM are integrated with TL to generate the predictions, notably in performance and real-time processing capability; the proposed model outperforms significantly. Moreover, various explainable AI (XAI) methods such as SHAP (Shapley Additive Explanations), heatmap, guided heatmap, Grad-Cam (Gradient-weighted Class Activation Mapping), guided Grad-CAM, and guided Saliency mapping techniques has been employed to explore interpretability and decision-making capability of the proposed model. Therefore, the model provides practical intelligence for increasing confidence in diagnosing GI diseases in real-world scenarios.

**Keywords**

Gastrointestinal tract, GastroVision, Depthwise Separable Convolutional Neural Network, Pearson Correlation Coefficient (PCC), Ensemble Extreme Learning Machine (EELM), Contrast-Limited Adaptive Histogram Equalization (CLAHE), Guided GradCam, Saliency map, and Shapley Additive Explanations (SHAP), Explainable AI (XAI).

# Introduction

The gastrointestinal (GI) system, which includes the organs associated with digestion and food absorption, is essential for sustaining good health. This intricate system is susceptible to multiple disorders that can significantly impact its daily functioning. GI diseases such as polyps, esophageal disorders, colon cancer, and ulcerative colitis affect organs including the stomach, intestines, liver, and pancreas. Medical imaging technology has made significant strides towards automatic diagnosis of these diseases in the last 20 years. Early identification and accurate diagnosis are essential for successful treatment of many diseases, but it requires a large number of healthcare experts which is costly, error-prone, and time-consuming. Moreover, rural areas often struggle to fulfill the need for more skilled medical professionals [1]. Addressing these issues necessitates technological solutions that can automatically and accurately detect and assess GI diseases.

Digestive diseases significantly accelerate mortality rates, indicating an alarming phenomenon in public health recently. Colorectal cancer is one of the most common GI illnesses. Since 2015, almost 132,000 new cases of colorectal cancer have been reported in the USA, affecting 1.6 million individuals with bowel infections. Approximately 200,000 new cases occur each year [2]. The USA documented 135,430 cases of various GI diseases in 2017 [3]. Additionally, 18% of adults in Brazil, 11% in China, 20% in EU-5, 12% in Russia, and 21% in the US were diagnosed with GI diseases [4]. In 2017, a worldwide survey reported 765,000 fatalities from stomach diseases, with colon cancer being responsible for 525,000 deaths [5]. Worldwide, there were approximately 4.8 million new instances of GI malignancies and 3.4 million deaths related to these illnesses in 2018 [6]. About 3.6 million children are affected by stomach infections each year [2]. Esophageal cancer is the seventh most prevalent cancer worldwide, whereas stomach cancer is the third leading cause of cancer-related fatalities globally.

Accurate and early diagnostic of GI diseases play a significant role in reducing the mortality rate. Endoscopy, which includes esophagogastroduodenoscopy (EGD) and colonoscopy, is one of the best and most effective ways to examine the upper and lower intestines for potential health problems. Furthermore, capsule endoscopy, endoscopic ultrasonography (EUS), CT scan, magnetic resonance imaging (MRI), and positron emission tomography (PET) scan are other important techniques for the thorough diagnosis of GI disorders. These techniques support medical personnel to observe the GI tract, evaluate organ health, and detect abnormality quickly, which helps in timely intervention and enhances the possibility of successful treatment outcomes [6].

The complicated structure of small bowls makes push gastroscopy instruments unsuitable for the identification and analysis of GI infections such as polyps, ulcers, and bleeding. A standard endoscopy may fail to detect numerous lesions because of the presence of secretions. During colon cleansing operations intended for cancer or precursor lesion diagnosis, a significant number of polyps remain undiscovered, with rates ranging from 21.4% to 26.8% which pose significant challenges [7]. Furthermore, as polyp growths may exhibit similarities in multiple categories, accurate diagnosis can be challenging. In 2000, a new technology called Wireless Capsule Endoscopy (WCE) was introduced to resolve these issues to a certain extent [8]. During WCE, a medical professional visually inspects the interior of the GI tract to identify any diseases. The patient swallows a capsule with a wireless camera, light-emitting diodes, radio frequency emitter, and a battery throughout this procedure. The system autonomously navigates through the GI tract and the camera captures thousands of images. The images are stored on recorders and then transmitted to a computer with specialized software that compiles them to form a video. The gastroenterologists evaluate those images and track the lesion. However, the primary issue of this process is the longer time in classifying many types of GI diseases. Over 50,000 pictures are generated during a WCE scan. The physician needed an average of two hours to analyze the images, and the risk of incorrect detection is very high [9].

Previous research shows substantial progress in developing various artificial intelligence (AI) models for classifying the GI tract. These models utilize a variety of methodologies, such as rule-based reasoning and neural networks (NN) [10–12]. Although significant progress has been made, certain challenges need to be addressed. Previous research mostly concentrated on developing image diagnosis methods to classify precursor lesions associated with GI disorders precisely [13–18]. Traditional approaches in these studies included improving contrast, removing noise, and segmenting regions. Several studies have also delved into classifying diseases within the GI tract [19–21]. Yet, a significant drawback of these approaches is their focus on a limited number of diseases and a limited number of samples around 1650 to 4854, and lack of models’ interpretability. Additionally, certain researchers utilized transfer learning (TL) based methods containing a significant number of parameters and layers. Although, these methods required substantial processing time because of extracting irrelevant features, which creates significant challenges for real-time applications.

This study aims to mitigate those problems to identify them in prior studies by introducing a comprehensive disease classification framework. The primary contribution of this research is outlined as follows:

* For the first time, a DL model is employed to classify large number of GI tract diseases (27 classes) which contains large number of upper, lower, and combined GI samples.
* A hybrid preprocessing method (CLAHE, erosion, sharpening, and gaussian filters) has been introduced to enhance image quality on the multi-class GastroVision dataset.
* A lightweight novel DL model proposed called Parallel Depth-wise separable CNN (PD-CNN) to perform feature extraction with distinct characteristics and minimal parameters, resulting in a significant decrease in model size, parameters, layers, and testing time.
* The Pearson correlation coefficient (PCC) has been utilized to reduce irrelevant features assessing the linear connection between the features and target classes, which improves the effectiveness of the proposed PD-CNN model.
* A novel Ensemble Extreme Learning Machine (EELM) classifier has been designed, which is a combination of pseudo inverse-ELM (ELM) and L1-regularized ELM (RELM) to accelerate performance in classifying GI tract diseases.
* The study has evaluated the classification performance, parameters, layers, and sizes of the proposed PD-CNN model with different established transfer learning (TL) models to demonstrate the proposed model's superiority.
* The framework's interpretability is highlighted by using various explainable AI (XAI) techniques including SHapley Additive exPlanations (SHAP), heatmap, guided heatmap, Grad-CAM, guided Grad-CAM and guided Saliency mapping which demonstrate significant insights into the model's decision-making process.

Section 2 provides a detailed summary of the previous relevant research. Section 3 outlines the suggested methodology, consisting of a proposed framework, dataset and section 4 demonstrates detailed model architecture. Section 5 provides an elaborate overview of complete classification outcomes, accompanied by interpretability of the proposed model using XAI. Section 6 presents the key conclusions.

# Related Works

Multiple approaches have been explored in the field of medical diagnosis and decision support for GI diseases [22–29]. Researchers have utilized several ML and DL approaches to detect and analyze various GI diseases, including cancer and ulcer from endoscopic images [13–15]. Also, ongoing research is being conducted in the field of multi-class classification, but mostly encompassing a smaller number of classes [19–21,35,36].

A medical diagnosis decision support model for gastrointestinal cancer was presented by Saraiva et al. [11] using a mix of rule-based and case-based reasoning. Subsequently, Aruna et al. [10] presented a NN model for GI diagnosis that used radial basis functions and backpropagation. The fuzzy inputs used in this model were derived from patient interviews. Furthermore, Awais et al. [12] presented a unique model for myocardial infarction detection that was based on the defense mechanism of the human digestive system. Also, a polyp detection method for WCE images was presented by Li et al. [37]. Using support vector machine (SVM) classifier features were extracted through an integration of wavelet transform and uniform local binary pattern. All these studies have been carried out on comparable tract areas, yet very few of these specifically addressed multiclass classification.

Researchers have also used different aspects of the human body like cancer, blood, polyps, ulcer lesions, dyed-lifted-polyps, and ileocecal to detect problems in endoscopic images [13–18]. Musha et al. [30] suggested a method utilizing chromaticity moment color feature and uniform local binary pattern for bleeding region detection in endoscopy images. Similarly, Pan et al. [15] used probabilistic NN for bleeding detection. On the other hand, Noya et al. [13] applied a boosted decision tree (DT) classifier using a combination of color-based, texture, statistical and morphological features for detecting angiodysplasia lesions. Li et al. [14] introduced a texture extraction process curvelet-based local binary pattern for the detection of ulcer regions in capsule endoscopy images. Utilizing multilayer perceptron NN and SVM they classify ulcer regions. Most of these studies have emphasized extracting many features overachieving a balanced real-time benchmark. Morphological operations and statistical analysis were conducted to produce the data.

Yeh et al. [31] proposed a method for detecting ulcers and bleeding in images acquired using WCE. Color features have been used to evaluate the condition of the small intestine, and various feature selection approaches, and classifiers were utilized, with the DT demonstrating the highest accuracy in bleeding detection. In a separate study, Lee et al. [16] evaluated TL models such as ResNet50, Inceptionv3, and VGG16 to classify stomach endoscopic images and distinguish between normal and benign ulcers. The results showed high accuracy, with ResNet50 consistently performing better than other methods. Yuan et al. [32] generated a computer-aided technique for detecting ulcers. The approach effectively identified ulcer regions through a multi-level super-pixel representation. The method utilized Locality-Constrained Linear Coding (LLC) and saliency max-pooling to save visual features. Furthermore, Pogorelov et al. [33] created Kvasir, a multi-class image dataset designed for computer-aided identification of GI diseases, with annotated images from the GI tract. The collection contains anatomical landmarks, pathological observations (such as esophagitis, polyps, and ulcerative colitis), and images associated with endoscopic polyp removal. Jain et al. [34] developed WCENet, a deep CNN model for detecting and locating anomalies in WCE images. Their proposed model worked in two stages: an initial stage utilizing an attention-based CNN to categorize images into certain groups (polyp, vascular, inflammatory, or normal), followed by a phase that combines Grad-CAM++ with a customized SegNet for locating anomalies in images.

As mentioned before, very little work has been done recently on the multi-class classification of GI tract diseases. Nouman et al. [21] proposed method for classifying GI tract diseases using Kvasir [33] and Hyper-Kvasir [38] dataset which contains 4854 images of five different classes. The process has included contrast optimizing using a genetic algorithm (GA), utilizing MobileNetV2 for feature extraction, and applying machine-learning classifiers SoftMax. Similarly, Noor et al. [20] suggested a computer-aided diagnosis system for GI diseases with lightweight MobileNetV2 feature extractor, and SoftMax classifier. They also have integrated attention mechanisms and a cosine similarity-based feature selection technique to reduce the number of features to improve the effectiveness of the classification. Using 810 key features, the framework has obtained high accuracy (97.68%) to classify GI tract images into 5 different classes of the Kvasir dataset [33]. Gunasekaran et al. [19] presented GIT-NET, an ensemble model that has used pre-trained models DenseNet201, InceptionV3, and ResNet50 to classify GI diseases accurately. The Kvasir v2 [33] dataset, which has 8000 photos from 8 classes, was utilized in this study. With an accuracy of 95%, the proposed weighted average ensemble method outperforms individual models. Sivari et al. [35] also utilized the Kvasir v2 and Hyper-Kvasir datasets to develop a DL-based hybrid stacking ensemble model for detection and classification of GI tract from endoscopic images. The models were trained using a two-level stacking architecture. The second level includes logistic regression, linear SVM, multi-layer perceptron, and k-nearest neighbor algorithms. Rustam et al. [36] introduced bleedy image recognizer (BIR) combining MobileNet and a custom-built CNN model for automatic analysis of WCE images. With a dataset of 1650 images, BIR achieved impressive performance, demonstrating high accuracy of 99.3%. Lan et al. [39] carried out a study that introduced a hybrid unsupervised DL technique to summarize videos within a weakly supervised cross-modal embedding framework. They used networks such as long short-term memory (LSTM) and autoencoder to help healthcare professionals analyze WCE videos in detail. In [40], a method that integrates Mask R-CNN, fine-tuned ResNet models, and an Enhanced Ant Colony Optimization algorithm was proposed. ResNet-50 and ResNet-152 models achieved an impressive classification accuracy of 96.43%. Another study [41] utilized discrete wavelength transformation (WT) and CNN approaches to categorize polyp and esophagitis classes. The method achieved an impressive accuracy rate of 96.65%.

Several researchers achieved a higher level of accuracies ranging from 93% to 98% in their research. However, it is important to note that these researchers utilized datasets with a limited number of classes (ranging from 5 to 8) and small number of image samples (ranging from 4000 to 8000) to demonstrate performance of their proposed models. Furthermore, most researchers employed TL and DL-based models, including GIT-NET, MobileNetV2, ResNet-50, ResNet-152, and a hybrid stacking ensemble model. Nevertheless, a major problem arises regarding the computational requirements (parameter, size, layer) led to longer processing times, which made it difficult to use the classification model effectively. Furthermore, certain advanced techniques require high-resolution images to achieve precise classification, which is crucial for real-world applications, especially in embedded systems [2,10,11,13–17,20,21,40,42]. To accelerate the widespread use of the GI disease classification model, it is crucial to improve existing models by reducing parameter counts, size, layer, minimizing processing times, and boosting classification accuracy. In addition, certain research demonstrated the application of XAI techniques such as heatmaps and Grad-CAM. However, most of the SOTA research has not focused much on assessing the impact of individual features. This study acknowledges these challenges and suggests a novel and efficient solution.

# Methodology

## Proposed Framework

A novel methodology utilizing DL approach has been developed to address the complicated issues linked to identifying large number of GI disorders. The procedural phases involved in this research are illustrated in **Figure 1**. Initially, the annotated dataset was divided into twenty-seven unique disease categories by 90:10 ratio for the training and testing sets. Several data preprocessing steps were executed to enhance the model's learning ability. These procedures comprised implementing data normalization, CLAHE, erosion, sharpening, gaussian filters, and resizing images from the training set to 124×124 pixels. Following that, a lightweight novel architecture Parallel Depthwise separable CNN (PD-CNN) was constructed in conjunction with parallel CNN (Pr-CNN) where traditional convolutional layer used. This architecture was utilized to test multiple TL models simultaneously and extract 200 features. 39 significant features were ascertained by eradicating 161 irrelevant ones using the PCC algorithm and t-SNE visualization on the feature distribution. Subsequently, Z score normalization was applied to achieve standardization. An ensemble extreme learning machine (ELM) classifier was designed to improve the model's classification performance from the PCC features. This classifier incorporated the ELM and RELM approaches. The model weights that yielded the most accurate predictions were maintained after a comparative analysis. Moreover, employing various XAI techniques, the decision-making capabilities of the proposed models were graphically presented.

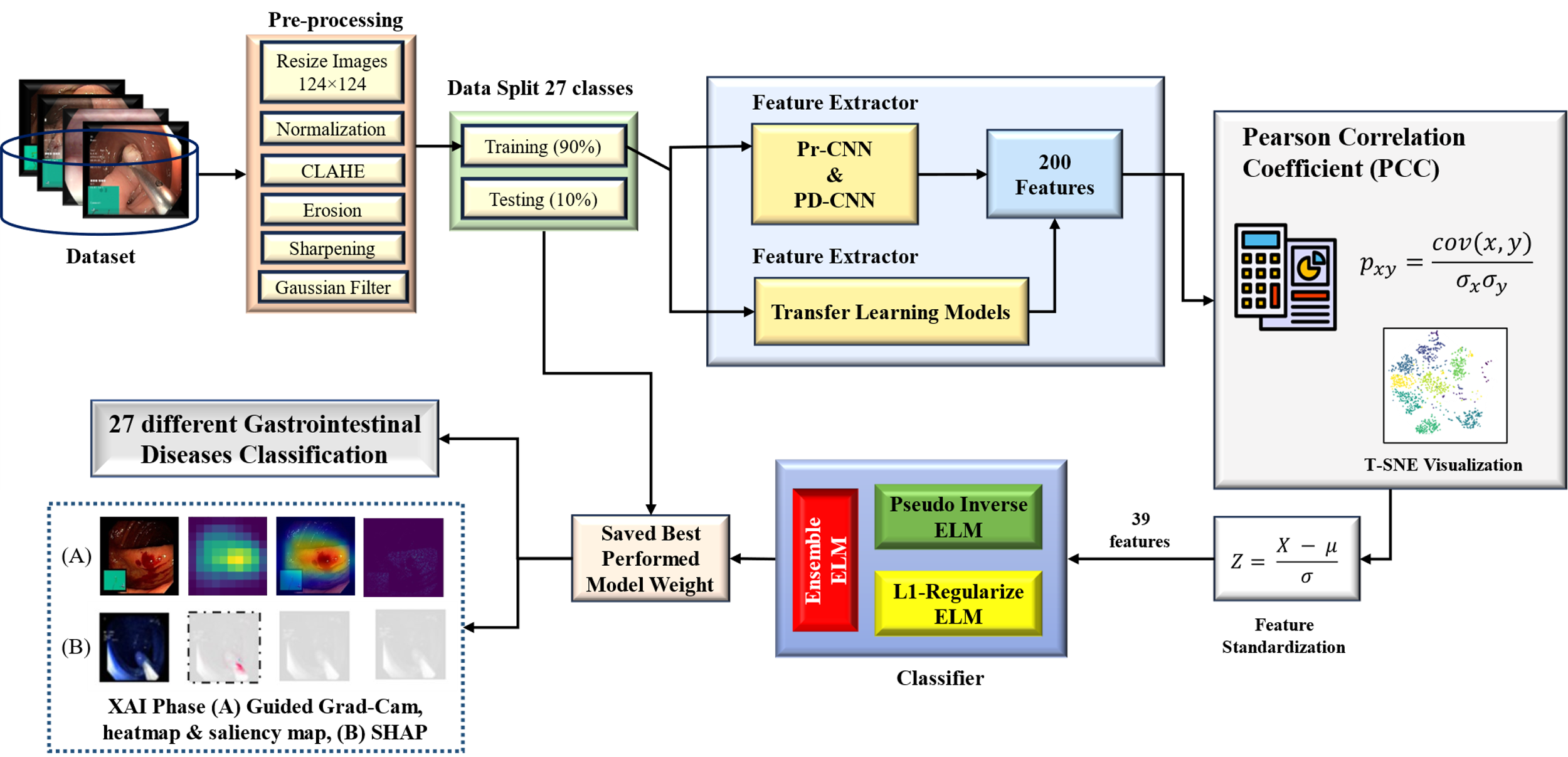


Figure 1: Proposed working framework for GI diseases multi-class classification.

## Dataset Description

A comprehensive multi-center dataset developed for GI endoscopy applications denoted as GastroVision, was used in this study [43]. The objective of GastroVision is to aid in the advancement and assessment of AI-driven algorithms utilized for the identification and categorization of gastrointestinal disorders. With a total of 27 unique classes representing different GI tract diseases, the dataset contains a broad variety of anatomical landmarks, clinical abnormalities, normal results, and instances of polyp removal. A combined group of images representing both normal variations and GI pathology is available in addition to the upper GI and lower GI categories, which are based on the digestive tract. The data collection process for GastroVision was a collaborative effort between Bærum Hospital in Norway and Karolinska University Hospital in Sweden. Skilled GI endoscopists meticulously conducted endoscopic procedures, capturing high-resolution images of diverse GI tract regions, including the esophagus, stomach, small intestine, colon, rectum, and terminal ileum. Following data acquisition, expert GI endoscopists meticulously annotated and verified the images. The dataset comprised a comprehensive collection of 8,000 high-resolution endoscopic images from various GI regions on the human body. Moreover, the dataset is thoughtfully distributed across different areas of the GI tract, ensuring comprehensive coverage of GI pathology and normal variations. Comprehensive details of the dataset, along with sample images of model training and testing set, are provided in **Table 1 and Figure 2**.

Table 1. Dataset distribution in twenty-seven classes and data split of training and testing set.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Testing phase** | **GI**  **Position** | **labeled areas** | **Disease Types** | **Class No.** | **Training** | **Testing** |
| GI tract Diseases  (27-classes) | Upper GI | Normal findings | Normal stomach | 20 | 872 | 97 |
| Normal esophagus | 18 | 126 | 14 |
| Anatomical Landmarks | Gastroesophageal\_junction\_normal z-line | 15 | 297 | 33 |
| Duodenal bulb | 8 | 185 | 20 |
| Pylorus | 21 | 354 | 39 |
| Pathological Findings | Barrett's esophagus | 2 | 86 | 9 |
| Esophagitis | 13 | 96 | 11 |
| Gastric polyps | 14 | 59 | 6 |
| Ulcer | 26 | 5 | 1 |
| Esophageal varices | 12 | 6 | 1 |
| Lower GI | Normal Findings | Normal mucosa and vascular pattern in the large bowel | 19 | 1320 | 147 |
| Anatomical Landmarks | Cecum | 4 | 102 | 11 |
| Colon diverticula | 5 | 26 | 3 |
| Ileocecal valve | 16 | 180 | 20 |
| Retroflex rectum | 24 | 60 | 7 |
| Small bowel\_terminal ileum | 25 | 761 | 85 |
| Pathological Findings | Angioectasia | 1 | 15 | 2 |
| Mucosal inflammation large bowel | 17 | 26 | 3 |
| Colon polyps | 6 | 738 | 82 |
| Colorectal cancer | 7 | 125 | 14 |
| Therapeutic interventions | Dyed-lifted-polyps | 9 | 127 | 14 |
| Dyed-resection-margins | 10 | 221 | 25 |
| Resected polyps | 22 | 83 | 9 |
| Resected margins | 23 | 23 | 2 |
| Upper & Lower GI | Pathological findings | Blood in lumen | 3 | 154 | 17 |
| Erythema | 11 | 14 | 1 |
| Therapeutic interventions | Accessory tools | 0 | 1139 | 127 |
| **Total** | | | | **7200** | **800** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| (A) | (B) | (C) | (D) | (E) | (F) |
|  |  |  |  |  |  |
| (G) | (H) | (I) | (J) | (K) | (L) |
|  |  |  |  |  |  |
| (M) | (N) | (O) | (P) | (Q) | (R) |
|  |  |  |  |  |  |
| (S) | (T) | (U) | (V) | (W) | (X) |
|  | |  | |  | |
| (Y) | | (Z) | | (AA) | |

Figure 2: GastroVision dataset include (A) Accessory tools, (B) Angiectasia, (C) Barrett's esophagus, (D) Blood in lumen, (E) Cecum, (F) Colon diverticula, (G) Colon polyps, (H) Colorectal cancer, (I) Erythema, (J) Dyed-lifted-polyps, (K) Dyed-resection-margins, (L) Erythema, (M) Esophageal varices, (N) Esophagitis, (O) Gastric polyps, (P) Gastroesophageal\_junction\_normal z-line, (Q) Ileocecal valve, (R) Mucosal inflammation large bowel, (S) Normal esophagus, (T) Normal mucosa and vascular pattern in the large bowel, (U) Normal stomach, (V) Pylorus, (W) Resected polyps, (X) Resection margins, (Y) Retroflex rectum, (Z) Small bowel\_terminal ileum, and (AA) Ulcer multi-classes.

## Preprocessing

The image processing step is vital for optimizing the performance of deep learning models. Each image was scaled to a consistent resolution of 124×124 pixels and z-score normalization procedures were performed to ensure consistency and optimal data representation for model training. The dataset is partitioned into two distinct subsets: 90% is allocated for training, 10% for testing phase. This preprocessing step significantly improves the resilience and ability of the model to perform well on various datasets. In addition, a range of image enhancing methods, including erosion, CLAHE, sharpening, and gaussian filters was used as part of the preprocessing workflow. Erosion aids in reducing noise and refining picture boundaries, but CLAHE increases contrast, especially in areas with shifting illumination. Sharpening methods boost the identification of edges and improve the overall clarity of an image, while gaussian filters efficiently decrease noise and enhance the characteristics of images. The integration of various preprocessing approaches guarantees that the model is provided with refined and standardized input data, resulting in enhanced performance and accuracy in classification tasks. Examples of original selected images with the corresponding preprocessed images are presented in Figure 3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| (A) | (B) | (C) | (D) | (E) |

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

# Architecture

The dataset is ready for training after successfully performing the preprocessing stage. In the current research era, the most challenging term is to build a robust, lightweight model that can best serve while maintaining minimum testing time, parameters, layers, and sizes. Following that, a lightweight novel PD-CNN feature extractor has been proposed and for fundamental feature selection, PCC was utilized concurrently. Furthermore, a novel classifier, EELM has been proposed to boost up the model's decision-making capability.

## Feature Extraction

The primary objective of this research was to construct a custom CNN design that can extract critical features while simultaneously reducing both parameters and network depth. The configuration of layers in a CNN is of the utmost importance. An overabundance of parameters and layers may impede the model's ability to differentiate distinctive attributes, thus imposing a performance constraint. Conversely, an overabundance of parameters and layers increases the demand for computational resources and lengthens processing times due to the chance of overfitting. Therefore, it is crucial to achieve an ideal equilibrium to guarantee precise feature extraction and feasible implementation.

Figure 4 presents the proposed PD-CNN model architecture adept at managing layer complexities and adhering to parameter constraints to fulfill its objective effectively. The model incorporates convolution layers (CL) and fully connected layers (FC) to find an optimal balance. Instead of relying on a single CL model, five parallel CL models were employed to identify the essential features. This challenge was addressed by concurrently running the initial parallelly connected five depth-wise separable CLs instead of employing five consecutive CLs, which would amplify network depth and complexity [44]. Their selection was determined through a systematic trial-and-error process.

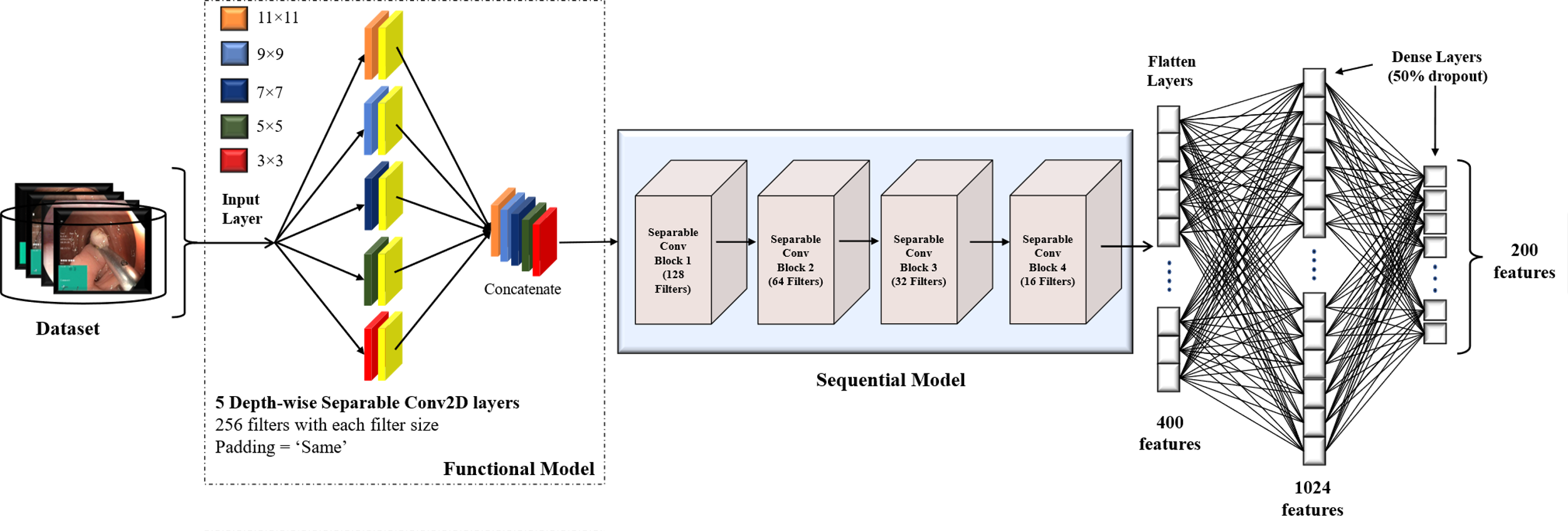


Figure 4: Proposed PD-CNN-EELM architecture.

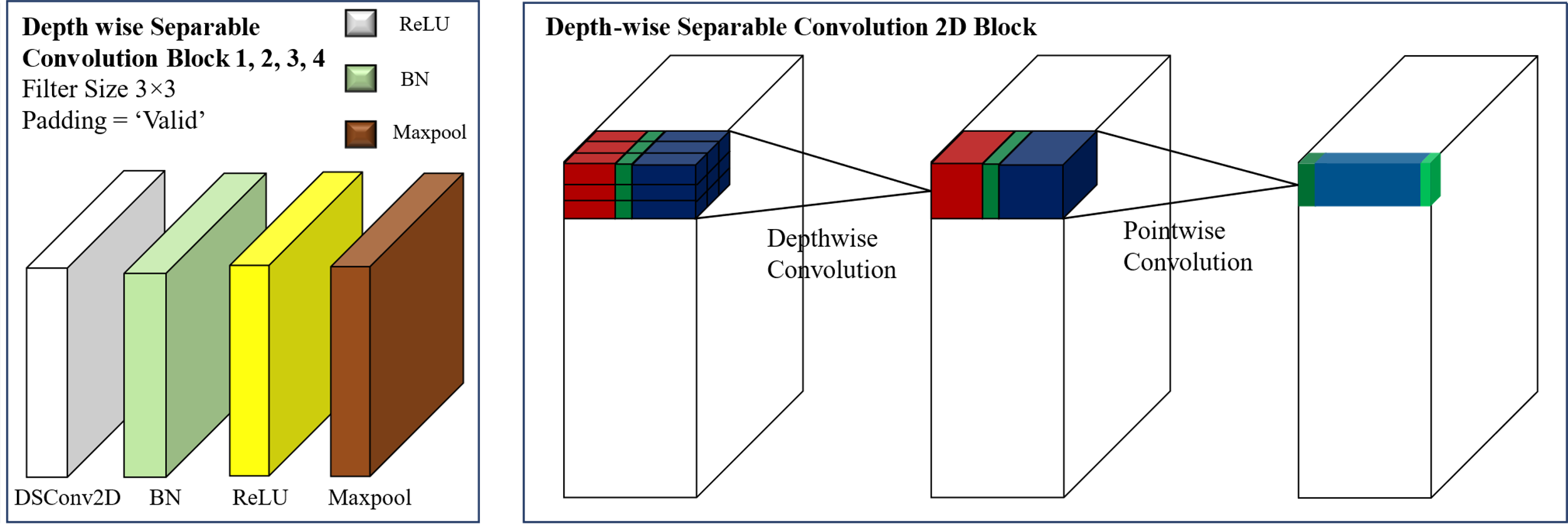


Figure 5: Detail overview of the convolution block.

The model begins with an input layer that accommodates images of variable dimensions. Subsequently, a series of parallel CL with varying kernel sizes (11×11, 9×9, 7×7, 5×5, and 3×3) are applied to capture spatial hierarchies and diverse patterns within the input data. For this research, the kernel size selection method suggested by Krizhevsky et al. was implemented [45]. This method entails the utilization of 11×11 kernel sizes, which have been found to yield satisfactory classification performance. Acknowledging the importance of the diverse proportions of the kernels, we undertook a comprehensive examination and synthesis of various kernels to identify critical characteristics and enhance the efficacy of classification. This methodology recognizes the distinct feature maps produced by various kernels. To optimize the results obtained from extracting critical data from the frame features of GI images, it is crucial to ensure that the initial five CLs have a consistent buffer size. The feature maps acquired from these concurrent CLs must be error-free and seamlessly integrated into a sequential CL to preserve the classification process's integrity.

The concatenated feature maps are then fed into the subsequent layers comprising standard separable convolutions, each followed by batch normalization and rectified linear unit (ReLU) activation functions. This hierarchical architecture enables the model to progressively distil and abstract high-level features from the input data while mitigating the risk of overfitting through regularization techniques. An updated feature map with fewer channels is produced by applying a 1×1 convolutional kernel separately to each channel during the pointwise convolution process. This emphasizes the pivotal significance of DSC (Depthwise Separable Convolution), as it induces a substantial reduction in computation complexity. During the concluding stage, three CLs were incorporated in addition to implementing BN and MP using a 2×2 kernel. The respective filter values for these CLs were 128, 64, 32, and 16; each filter utilized 3×3 kernels and VALID padding. Integrating BN enhances the model's efficiency by recalibrating each layer's input mean and standard deviation, improving execution speed and stability. Further details of the convolution block are presented in Figure 5.

The ReLU activation function was used in each CL. In addition to two FC layers, dropout was used to mitigate overfitting and enhance the efficiency of the training process. In every training cycle, random deactivation was applied to 50% of all nodes, which helped enhance generalization and accelerate convergence. The final FC layer retrieved 200 features. After that, PCC algorithm is employed to get significant features (39 features) from the final layer of 200 features. Furthermore, EELM classifier is integrated to generate final classification performance among 27 classes (Figure 6). Table 2 provides a comprehensive overview of the model summary.



Figure 6: Integration of EELM architecture with PCC from PD-CNN model’s last layer.

Table 2: Summary of proposed PD-CNN model.

|  |  |  |
| --- | --- | --- |
| **Layer Type** | **Output Shape** | **Params** |
| Model input layer | (None, 124, 124, 3) | 0 |
| Functional model | (None, 124, 124, 1280) | 5975 |
| Separable conv2d layer | (None, 122, 122, 128) | 175488 |
| Batch normalization | (None, 122, 122, 128) | 512 |
| Activation | (None, 122, 122, 128) | 0 |
| Max pooling | (None, 61, 61, 128) | 0 |
| Separable conv2d layer | (None, 59, 59, 64) | 9408 |
| Batch normalization | (None, 59, 59, 64) | 256 |
| Activation | (None, 59, 59, 64) | 0 |
| Max pooling | (None, 29, 29, 64) | 0 |
| Separable conv2d layer | (None, 27, 27, 32) | 2656 |
| Batch normalization | (None, 27, 27, 32) | 128 |
| Activation | (None, 27, 27, 32) | 0 |
| Max pooling | (None, 13, 13, 32) | 0 |
| Last convolution layer | (None, 11, 11, 16) | 816 |
| Batch normalization | (None, 11, 11, 16) | 64 |
| Activation | (None, 11, 11, 16) | 0 |
| Max pooling | (None, 5, 5, 16) | 0 |
| Dropout | (None, 5, 5, 16) | 0 |
| Flatten | (None, 400) | 0 |
| Dense | (None, 1024) | 410624 |
| Batch normalization | (None, 1024) | 4096 |
| Dropout | (None, 1024) | 0 |
| Dense Last | (None, 200) | 205000 |
| **Total Parameters:** | 815,023 | |
| **Trainable Parameters:** | 812,495 | |
| **Non-Trainable Parameters:** | 2,528 | |

## Ensemble Extreme Learning Machine (EELM)

The pseudo-inverse Extreme Learning Machine (ELM) is a well-recognized approach of single-hidden layer feedforward neural networks (SLFNs) [46]. In contrast to traditional NN training methods, ELM utilizes a unique strategy of randomizing and fixing the parameters that connect the input layer to the hidden layer by using the pseudo-inverse technique. This allows for the exclusive training of the parameters that link the hidden layer to the output layer. Randomized initialization accelerates training processes and enhances generalization capacities. A useful feature selection and regularization procedure is introduced into the ELM framework by including L1 regularization, often known as Lasso regularization [47]. L1 regularization encourages sparsity in feature representations by adding a penalty term to the loss function, which effectively pushes many feature weights towards zero. By incorporating L1 regularization into ELM (RELM), the ability to distinguish features is improved and the chance of overfitting is reduced, resulting in a stronger performance of the model in generalizing.

A novel ensemble method that combines ELM and RELM classifiers has been proposed. Ensemble learning is a method that integrates many classifiers to enhance accuracy and resilience. However, the proposed methodology distinguishes itself by including ELM and L1 Regularized ELM models in this strategy. The ensemble operation combines predictions from individual ELM and RELM classifiers, each trained on separate subsets or with various initializations. This integration combines ELM's efficiency with RELM's feature selection, improving classification performance. The ELM has shown remarkable competence in handling large-scale multi-class classification tasks, outperforming current machine learning models. However, this work enhances the complexity by substituting the pseudoinverse method and L1 Regularized methodology with an ensemble approach. This augmentation greatly enhances the model's ability to learn and control features, improving its potential for generalization and obtaining unmatched accuracy compared to each approach. In the classifier's design, 39 nodes are in the input layer after employing PCC, and a staggering 900 nodes are in the hidden layer. In addition, the EELM algorithm produces twenty-seven nodes crucial in categorizing different samples from GI tract images. The ELM, RELM and proposed EELM are described in Algorithm 1.

The explanation of EELM algorithm is given below:

|  |
| --- |
| **Algorithm 1:** Proposed EELM Classifier algorithm. |
| 1. Feature sample is , and output is .  |  |  | | --- | --- | |  |  | |
| 1. Input weight and bias metrics is presented as , and .  |  |  | | --- | --- | |  |  | |
| 1. Hidden layer is used to generate the output.  |  |  | | --- | --- | |  |  |   Where, denotes activation function |
| 1. In ELM, output weight metric is presented as , |
| 1. In RELM, output weight metric is presented as , where the ELM equations are replaced as follows:   Here, is called regularizationparameter. |
| 1. The following formula denotes the proposed Ensemble operation: |
| 1. The generated prediction, . |

## Transfer Learning (TL)

The diagnosis of GI diseases across many classes can be greatly improved by using transfer learning models such as DenseNet201 [48], EffiecientNetB6 [49], InceptionResNetV2 [50], MobileNetV2 [51], ResNet152V2 [52], VGG16 [53], and Xception [54]. These models extracting large number of features from images due to their extensive pre-training on large datasets. Fine-tuning them on limited data for the specific task enables effective capture of intricate patterns and subtle details associated with GI diseases. The pre-trained models were trained using more than 14 million classifications from 1,000 categories (ImageNet). We integrated the training of TL models with the PCC for reducing unnecessary features, and EELM classifier to attain accurate classification outcomes and compared the PD-CNN model to TL approaches in terms of classification results and computational resources as there are no previous research on this dataset. This comparison encompasses performance metrics, model parameters, layer, sizes, and the duration of testing time. After initializing the TL models, two FC layers were added with 1024 and 200 nodes each to improve the detection of GI diseases. PCC was employed simultaneously to eliminate 200 to 39 necessary features. **Figure 7** shows the comprehensive illustration of TL models with PCC and EELM classifier.



Figure 7: The modified TL architecture with PCC and EELM to classify GI diseases.

DenseNet is a CNN architecture that utilizes dense connectivity, allowing each layer to receive input from all preceding layers [48]. This promotes effective information transmission and improves overall performance. The variants of DenseNet, including DenseNet-121, DenseNet-169, DenseNet-201, and DenseNet-264, differ in the number of layers they contain, with DenseNet-121 having 121 layers and DenseNet-201 having 201 layers. EfficientNetB6 architecture employs compound scaling, which involves scaling the network's depth, width, and resolution correspondingly [49]. This model is trained using the ImageNet dataset and comprises 87 million parameters. It has been used for several kinds of TL applications, such as semantic segmentation, object detection, and image categorization. The InceptionResNetV2 [50] model combines the Inception and ResNet architectures, utilizing inception modules and residual connections to extract features effectively. MobileNetV2 [51] is another CNN architecture introduced by Sandler et al., which is based on an inverted residual structure. It employs lightweight depth-wise convolutions and bottleneck layers to achieve better performance while being computationally efficient, so this architecture is particularly efficient for mobile devices. In 2017, He et al. [52] presented ResNet152V2, a NN model that utilizes residual learning by incorporating shortcut connections across layers to enhance learning efficiency. ResNet152V2 contains 60 million parameters. The Visual Geometry Group (VGG) is characterized by multiple CLs and filters [53]. After each CL, feature extraction is enhanced with a Max Pooling (MP) layer and a Rectified Linear Unit (ReLU) function. Google developed Xception, an architecture based on the Inception framework, in 2016 [54]. The system employs pointwise convolutions and DSCs to filter each channel of the input feature map separately. This approach preserves precision while significantly decreasing memory consumption and processing requirements. Xception is commonly utilized for various computer vision tasks due to its high efficiency, especially in situations with limited computational resources.

## Feature Selection with PCC

In the current era of ML, where data is paramount, it is critical to emphasize the importance of identifying pertinent features. Pattern recognition systems are based on features, measured parts of things that help identify patterns. However, of the many characteristics that may be provided, only a specific subset is significantly relevant to the final output. The extensive feature space with ML methods causes many problems, like slow learning and more complicated computations. So, finding the best group of features, which can be done by carefully choosing which features to use, becomes significant for getting through these problems. The Pearson Correlation Coefficient (PCC)-based method stands out among the many feature selection procedures as a potential way to isolate essential characteristics from various possibilities [55]. Using PCC, this method aims to reduce complexity and improve efficiency and processing speed by selecting the most relevant feature subset from those recovered by CNNs. By computing correlation values across all features, finding pairs with correlations more robust than certain limits is easier. This reduces the features that are not needed and improves the feature space. Additionally, the correlation coefficient is calculated by dividing the product of the standard deviations of two variables by the covariance between them. It ensures that its output remains within the interval of -1 to 1. Standardization approaches, such as the ordinary score equation for a sample, highlight the careful focus on data pretreatment essential for achieving effective machine learning outcomes. This thorough procedure of selecting features and standardizing them emphasizes the crucial significance of precise data pretreatment techniques in fully harnessing the capabilities of machine learning algorithms. The algorithm 2 presents the working steps of PCC.

|  |
| --- |
| **Algorithm 2:** Feature selection utilizing PCC |
| 1. **BEGIN** |
| 1. Define data, |
| 1. CorrMat = features.corr() |
| 1. Calculate the mean of each data set:   , and |
| 1. Calculate the standardized values for each data point: |
| 1. Calculate the covariance: |
| 1. Calculate the standard deviation:   , and |
| 1. Calculate PCC, corr(): |
| 1. for i ≤ CorrMat.col: |
| 1. for j ≤ i: |
| 1. If CorrMat.iloc[i, j] > threshold: |
| 1. ColName = CorrMat.col[i] |
| 1. CorrCol.add(ColName) |
| 1. Dropped.features(CorrCol) |
| 1. **END** |

## Explainable Artificial Intelligence (XAI)

In the realm of DL, XAI refers to understanding and clarifying the decision-making process of a deep neural network [56]. This is essential due to the model's complexity and difficulty to understand. XAI is used in this study to diagnose GI illnesses across 27 different categories to verify accurateness. SHAP, heatmap, guided heatmap, Grad-CAM, guided Grad-CAM, and saliency mapping were employed to address the "black box" character of the DL models, which can hinder their usefulness. The goal was to enhance the transparency and interpretability of the proposed PD-CNN model by utilizing XAI. By combining the PD-CNN model with XAI for disease classification, endoscopists can make more accurate and confident decisions when diagnosing diseases more efficiently. It will help healthcare professionals confirm the model's predictions, identify errors, reduce biases and missing rates by providing accurate diagnoses. This advancement creates new opportunities for better disease management techniques and more efficient therapies for GI issues.

### 4.5.1 Shapley Additive Explanations (SHAP)

This study utilized Shapley values for determining the importance of individual pixels, which exhibited a clear pattern. Red pixels enhance accurate class recognition, but blue pixels hinder it by reducing the likelihood of successful categorization [57]. The Shapley values were computed using Equation (1).

|  |  |
| --- | --- |
|  | (1) |

indicates the impact on output resulting from the Shapely values of a specific feature, . The subset includes all features coming from feature , excluding feature . represents the weighted factor of the subset permutations. Equation (2) gives the predicted result, denoted by the sign .

|  |  |
| --- | --- |
|  | (2) |

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

|  |  |
| --- | --- |
|  | (3) |

The contribution of the feature is represented by in the proposed framework , where is the substitute model for the framework. The bias is indicated by . A crucial component that helps comprehend the fundamental workings of the model is the contribution of feature to the result and the function of .

### 4.5.2 Heatmap Visualization

A heatmap is generated to display the regions of the original image that had the most impact on the ultimate classification outcome. This heatmap is generated by computing the gradient of the last convolutional layer's output class score with respect to the feature maps [58].

|  |  |
| --- | --- |
|  | (4) |

Here, represents heatmap for a given input image , represents the class score, and corresponds to the ith feature map.

### 4.5.3 Guided Heatmap Visualization

The heatmap from the previous phase is refined using guided visualization techniques. Gradients are calculated via guided backpropagation and then scaled by the ReLU activation of the corresponding feature map [58].

### 4.5.4. Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM merges class discrimination with location. It creates a heatmap after computing weights for each feature map based on the gradient of the class score compared to the feature maps [59].

|  |  |
| --- | --- |
|  | (5) |
|  | (6) |

Here, indicate score for class c with respect to feature map , indicates calculated weight for every neuron, defines a global average pooling over the width (i) and height (j).

### 4.5.5 Guided Grad-CAM

Guided Grad-CAM visualization is an interpretability technique that merges guided backpropagation and Grad-CAM principles. This technique enhances the heatmap produced by Grad-CAM by including guided backpropagation gradients and highlighting the most significant features in the ultimate classification determination [60].

|  |  |
| --- | --- |
|  | (7) |

Here, indicates the class of interest, is co-ordinates of a pixel in the input image, indicates important weights for each feature map, is activation of feature map at pixel , indicates refined gradient map obtained through guided backpropagation for class .

### 4.5.6. Guided Saliency Mapping

Saliency mapping assesses the spatial support of a class. It facilitates interpretability in neural networks by presenting an image that emphasizes the region of interest. The saliency map is generated using backpropagation. This method enhances comprehension of the model's judgments by locating pixels having minimal influence on the score and calculating the derivative of the class score regarding the image [61].

First, the distance of each pixel is calculated to the rest of pixels in the same frame:

|  |  |
| --- | --- |
|  | (8) |

is the value of the pixel . The following equation is expanded form of equation 9.

|  |  |
| --- | --- |
|  | (9) |

Where, represents all the pixels in the present frame. After that, formula 8 is refined. The values are combined that have the same .

|  |  |
| --- | --- |
|  | (10) |

Where, is the frequency of . And the value of belongs to [0, 255].

## Classification Matrices and Experimental Set-up

The model was trained using sparse categorical loss function, ADAM optimizer and a batch size of 32. The training process consisted of 200 epochs, during which a learning rate of 0.001 was employed for iterative experimentation. To assess the classification effectiveness of the proposed model, various metrics such as accuracy, precision, recall, f1-score, and area under the curve (AUC) were computed using the respective equations [62].

(11)

(12)

(13)

(14)

(15)

Where, = True positive, = True negative, = False positive, and = False negative.

The cross-entropy formula evaluates the correspondence between the integer-based actual class label and the probability distribution generated by the model. It calculates the difference between the real and predicted labels to reduce cross-entropy loss to the greatest extent possible. The sparse categorical cross-entropy loss is commonly utilized in deep learning scenarios like image classification, particularly when dealing with numerous classes [63]. The formula is as follows:

 (16)

Where, denotes the class number, truth label is defined as, and as the probability.

# Results and Discussions

This section provides a concise overview of the performance of the proposed framework that combines depth-wise separable CNN with and without (w/o) PCC and EELM classifiers. In addition, SOTA TL models are evaluated against this framework in terms of their classification performance, including parameters, layers, size, and computational cost.

## Depth-wise Separable CNN without PCC

**Table 3** presents the performance metrics including precision, recall, F1-score, accuracy, and AUC for various GI disease classes using the PD-CNN model without PCC approach. Across the board, the EELM classifier consistently outperforms both ELM and RELM classifiers within the PD-CNN framework. EELM demonstrates superior performances across all disease classes. The confusion matrices among the ELM, RELM, and EELM classifiers are displayed in **Figures 8**, providing essential insights into the classification outcomes. According to the analysis, EELM was identified as the most precise classifier.For instance, in disease class 7, EELM achieved a precision of 0.91, recall of 0.71, and an F1-score of 0.80, surpassing the performance of ELM and RELM. Notably, in disease class 24, EELM exhibited competitive improvements in precision, recall, and F1-scores, with values of 0.88, 1.00, and 0.93 respectively, outscoring both ELM and RELM. In disease class 25, ELM and RELM exhibited competitive precision, recall, and F1-scores of 0.96, 0.84, and 0.89 respectively. Additionally, the model couldn't identify classes 11, 12, and 26 due to a lack of training data. Only one sample was tested (see **Table 1**).

In comparison to ELM (average precision of 87.59%, recall of 87.62%, and F1-score of 87%), RELM (average precision of 87.31%, recall of 87.25%, and F1-score of 87.12%), and EELM demonstrated a considerable improvement with highest average precision of 88.11%, recall of 87.75%, and F1-score of 87.12%. Among the examined GI diseases EELM stands out with an estimated 0.594% improvement in precision, 0.15% in recall, and 0.14% in F1-score compared to ELM. This significant improvement highlights how EELM effectively tackles the categorization issues presented by certain classes.

With respect to accuracy, EELM achieved a competitive score of 87.75%, where 0.15% improved than ELM (87.62%), and 0.57% improved than RELM (87.25%). While the reason for some classes' lower EELM scores is that combining multiple base learners (ELM and RELM) improved overall predictive performance (precision, recall, F1 score), in specific scenarios, insufficient data diversity among the base learners prevented the ensemble from harnessing distinct perspectives and patterns to enhance predictive accuracy immensely. Another performance measure with AUC showed that EELM is better area coverage of 98.92% (see **Figure 10A**). This represents an increase of approximate 0.2% in comparison to ELM (98.88%) and RELM (98.50%). **Figure 11A** presents the AUC-PR curves where EELM achieved maximum area coverage of 92.10%. It is evident from the performance study that using the EELM classifier improves average precision, recall, and F1-scores, underscoring the classifier's potential application in a variety of contexts, particularly in the fields of predictive modeling and medical diagnostics.

|  |  |
| --- | --- |
|  |  |
| (A) | (B) |
|  | |
| (C) | |

Figure 8: Confusion metrics of (A) PD-CNN-ELM and (B) PD-CNN-RELM, and (C) PD-CNN-EELM models on GI tract diseases classification.

**Table 3. Class-wise performance using PD-CNN without PCC on test set.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GI Disease classes** | **Precision** | | | **Recall** | | | **F1-score** | | | **Accuracy (%)** | | | **ROC-AUC (%)** | | |
| **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** |
| 0 | 0.95 | 0.94 | **0.95** | 0.99 | 1 | **1** | 0.97 | 0.97 | **0.97** | 87.62 | 87.25 | **87.75** | 98.88 | 98.50 | **98.92** |
| 1 | 1 | 0 | **1** | 0.5 | 0 | **0.5** | 0.67 | 0 | **0.67** |
| 2 | 0.73 | 0.7 | **0.73** | 0.89 | 0.78 | **0.89** | 0.8 | 0.74 | **0.8** |
| 3 | 0.94 | 1 | **1** | 0.88 | 0.88 | **0.88** | 0.91 | 0.94 | **0.94** |
| 4 | **1** | 0.9 | 0.91 | 0.91 | 0.82 | **0.91** | **0.95** | 0.86 | 0.91 |
| 5 | 1 | 1 | **1** | **1** | 0.67 | 0.67 | **1** | 0.8 | 0.8 |
| 6 | 0.74 | 0.74 | **0.76** | 0.88 | 0.9 | **0.91** | 0.8 | 0.81 | **0.83** |
| 7 | 0.83 | 0.83 | **0.91** | 0.71 | 0.71 | **0.71** | 0.77 | 0.77 | **0.8** |
| 8 | **0.77** | 0.74 | 0.74 | 0.85 | 0.85 | **0.85** | **0.81** | 0.79 | 0.79 |
| 9 | 0.92 | **0.92** | 0.91 | 0.79 | **0.79** | 0.71 | 0.85 | **0.85** | 0.8 |
| 10 | **0.92** | 0.88 | 0.85 | 0.92 | 0.92 | **0.92** | **0.92** | 0.9 | 0.88 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 1 | 0.83 | **1** | 0.45 | 0.45 | **0.45** | 0.62 | 0.59 | **0.62** |
| 14 | 0.67 | 0.67 | **0.67** | 0.33 | 0.33 | **0.33** | 0.44 | 0.44 | **0.44** |
| 15 | 0.82 | 0.79 | **0.82** | 0.94 | 0.94 | **0.94** | 0.87 | 0.86 | **0.87** |
| 16 | 0.85 | **0.92** | 0.91 | 0.55 | **0.55** | 0.5 | 0.67 | **0.69** | 0.65 |
| 17 | 1 | 1 | **1** | 0.33 | 0.33 | **0.33** | 0.5 | 0.5 | **0.5** |
| 18 | 0.86 | 0.86 | **0.86** | 0.86 | 0.86 | **0.86** | 0.86 | 0.86 | **0.86** |
| 19 | 0.88 | 0.88 | **0.88** | 0.91 | 0.91 | **0.91** | 0.89 | 0.9 | **0.9** |
| 20 | 0.9 | **0.9** | 0.89 | 0.94 | 0.94 | **0.95** | 0.92 | 0.92 | **0.92** |
| 21 | 0.83 | 0.83 | **0.83** | 0.87 | 0.87 | **0.87** | 0.85 | 0.85 | **0.85** |
| 22 | 1 | 1 | **1** | **0.67** | 0.33 | 0.44 | **0.8** | 0.5 | 0.62 |
| 23 | 0 | 1 | **1** | 0 | 0.5 | **0.5** | 0 | 0.67 | **0.67** |
| 24 | 0.78 | 0.88 | **0.88** | 1 | 1 | **1** | 0.88 | 0.93 | **0.93** |
| 25 | 0.95 | 0.96 | **0.96** | 0.84 | 0.84 | **0.84** | 0.89 | 0.89 | **0.89** |
| 26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Average (µ) ± SD (𝜎) (%)** | 87.59±0.333 | 87.31±0.33 | **88.11±0.3** | 87.62±0.343 | 87.25±0.338 | **87.75±0.316** | 87±0.321 | 86.54±0.316 | **87.12±0.285** |

\*Bold values indicate the best results.

## Depth-wise Separable CNN with PCC

The performance of the proposed feature extractor, PD-CNN, after incorporating PCC with three different classifiers: ELM, RELM, and EELM is shown in **Table 4**. **Figures 9A, 9B and 9C** illustrate the confusion matrices among the EELM, RELM, and ELM classifiers. These matrices offer crucial insights into the classification process's outcomes. The novel PD-CNN model extracted an excessive 200 features. Unnecessary and insignificant features were subsequently eliminated to reduce the predictive complexity of the classification. Upon completion of the feature extraction process, the PCC algorithm was implemented to eliminate 161 redundant features, resulting in the retention of only 39 of the most salient features. The classification was then performed using PD-CNN-PCC-EELM framework and displayed by employing test set.

**Table 4** compares precision, recall, F1-score, accuracy, and ROC-AUC of several GI disorders. The EELM classifier frequently outperforms ELM and RELM in maximum evaluated metrics. EELM consistently shows higher performances compared to others. In class 5, EELM demonstrated perfect precision, recall, and F1 scores (1.00), but without PCC, it performed inadequately (**see Table 3,** and **Table 4**).

The performance enhancement of EELM compared to ELM and RELM is substantial. EELM obtained the highest average accuracy of 87.75%, 0.15% improved than ELM's accuracy of 87.62%, and 0.57% higher than RELM's 87.25%.

The EELM classifier accurately identifies and classifies the GI disorders within the PD-CNN-PCC model, showcasing its outstanding capabilities. Reducing superfluous features is a crucial benefit of PCC, substantially improving the PD-CNN model's performance. Compared to the PD-CNN-EELM model, PD-CNN-PCC-EELM shows competitive scores in both ROC-AUC (98.89%) and AUC-PR (98.89%) (see **Figure** **10B** and **11B**). **Figure 10** and **Figure** **11** illustrate the ROC-AUC and AUC-PR curves comparing ELM, RELM, and EELM within the proposed PD-CNN with and without PCC integration.

In summary, the proposed configuration emerges as the optimal selection, amalgamating the robust feature extraction capabilities inherent to the PCC. This synergy yields exceptional performance, rendering PD-CNN-PCC-EELM the most efficacious solution for precise and dependable GI disease classification tasks.

|  |  |
| --- | --- |
|  |  |
| (A) | (B) |
| A close-up of a crossword puzzle  Description automatically generated | |
| (C) | |

Figure 9: Confusion metrics of (A) PD-CNN-PCC-ELM and (B) PD-CNN-PCC-RELM, and (C) PD-CNN-PCC-EELM models on GI tract diseases classification.

**Table 4. Class-wise performance using PD-CNN-PCC on test set.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GI Disease classes** | **Precision** | | | **Recall** | | | **F1-score** | | | **Accuracy (%)** | | | **ROC-AUC (%)** | | |
| **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** |
| 0 | **0.95** | 0.94 | 0.94 | 0.99 | 0.99 | **0.99** | 0.97 | 0.97 | **0.97** | 87.62 | 87.25 | **87.75** | **98.91** | 98.43 | 98.89 |
| 1 | 0 | **1** | 0 | 0 | **0.5** | 0 | 0 | **0.67** | 0 |
| 2 | 0.73 | 0.67 | **0.73** | 0.89 | 0.89 | **0.89** | 0.8 | 0.76 | **0.8** |
| 3 | 0.88 | 0.83 | **0.94** | 0.88 | 0.88 | **0.88** | 0.88 | 0.86 | **0.91** |
| 4 | 1 | 0.91 | **1** | 0.91 | 0.91 | **0.91** | 0.95 | 0.91 | **0.95** |
| 5 | 1 | 0.67 | **1** | 0.67 | 0.67 | **1** | 0.8 | 0.67 | **1** |
| 6 | 0.76 | **0.76** | 0.75 | 0.9 | 0.87 | **0.9** | **0.83** | 0.81 | 0.82 |
| 7 | **0.91** | 0.83 | 0.83 | 0.71 | 0.71 | **0.71** | **0.8** | 0.77 | 0.77 |
| 8 | 0.74 | 0.74 | **0.74** | 0.85 | 0.85 | **0.85** | 0.79 | 0.79 | **0.79** |
| 9 | 1 | 1 | **1** | 0.71 | 0.71 | **0.71** | 0.83 | 0.83 | **0.83** |
| 10 | 0.89 | 0.89 | **0.89** | 0.96 | 0.96 | **0.96** | 0.92 | 0.92 | **0.92** |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0.83 | **1** | 0.83 | 0.45 | 0.45 | **0.45** | 0.59 | **0.62** | 0.59 |
| 14 | 0.67 | 0.67 | **0.67** | 0.33 | 0.33 | **0.33** | 0.44 | 0.44 | **0.44** |
| 15 | 0.84 | 0.84 | **0.84** | 0.94 | 0.94 | **0.94** | 0.89 | 0.89 | **0.89** |
| 16 | **0.92** | 0.85 | 0.85 | 0.55 | 0.55 | **0.55** | **0.69** | 0.67 | 0.67 |
| 17 | 1 | 1 | **1** | 0.33 | 0.33 | **0.33** | 0.5 | 0.5 | **0.5** |
| 18 | 0.87 | 0.87 | **0.87** | 0.93 | 0.93 | **0.93** | 0.9 | 0.9 | **0.9** |
| 19 | 0.87 | 0.89 | **0.9** | 0.93 | 0.93 | **0.93** | 0.9 | 0.91 | **0.91** |
| 20 | 0.91 | 0.91 | **0.91** | 0.93 | 0.94 | **0.94** | 0.92 | 0.92 | **0.92** |
| 21 | 0.83 | 0.81 | **0.83** | 0.87 | 0.87 | **0.87** | 0.85 | 0.84 | **0.85** |
| 22 | 1 | 1 | **1** | 0.56 | **0.56** | 0.44 | 0.71 | **0.71** | 0.62 |
| 23 | 1 | 0 | **1** | 0.5 | 0 | **0.5** | 0.67 | 0 | **0.67** |
| 24 | 0.88 | 0.88 | **0.88** | 1 | 1 | 1 | 0.93 | 0.93 | **0.93** |
| 25 | 0.95 | **0.96** | 0.95 | **0.82** | 0.81 | 0.81 | 0.88 | **0.88** | 0.87 |
| 26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Average (µ) ± SD (𝜎) (%)** | 87.59±0.334 | 87.31±0.329 | **88.12±0.332** | 87.62±0.339 | 87.25±0.3387 | **87.75±0.348** | 87.04±0.318 | 86.54±0.314 | **87.12±0.324** |

\*Bold values indicate the best results

|  |  |
| --- | --- |
|  |  |
| (A) | (B) |

Figure 10: ROC-AUC performances on (A) PD-CNN and (B) PD-CNN-PCC models on GI tract diseases classification.

|  |  |
| --- | --- |
|  |  |
| (A) | (B) |

Figure 11: AUC-PR performances on (A) PD-CNN and (B) PD-CNN-PCC models on GI tract diseases classification.

## Performance comparison with TL models

PCC was incorporated with TL models for comparative study, which is showed in **Table 5**. Compared to regular CL, DSC layers performed better in the suggested architecture.

Among the models, Pr-CNN, Pr-CNN-PCC, PD-CNN, and proposed PD-CNN-PCC models demonstrate substantial improvements. In every metric, the PD-CNN-PCC model outperformed the Pr-CNN-PCC model. Similarly, PD-CNN-PCC achieved 0.863 for recall, 4.87% higher than Pr-CNN-PCC's score of 0.821. An analysis of the F1-score revealed that it increased by 5.26% from 0.811 for Pr-CNN-PCC to 0.856 for PD-CNN-PCC. Regarding computational time, it was seven times faster than the Pr-CNN-PCC technique.

PD-CNN-PCC-EELM outperformed EfficientNetB6-PCC in accuracy, scoring an impressive 0.866, well above 0.11 with an enormous improvement of 87.29%. The recall metric also showed significant improvement, reaching 0.863, better than ResNet152V2, the second-best transfer model, which scored 0.838. The recall has been improved by an impressive 2.9%. The F1-score, which stands at 0.856 and surpasses the transfer models' range of 0.672 to 0.828, further demonstrates the PD-CNN-PCC model's superiority. Regarding model performance, this corresponds to gains of 21.5% with Xception and 85.3% with EfficientNetB6. Finally, in accuracy, PD-CNN-PCC exceeded Xception and VGG16, reaching an impressive score of 86.13%. There is a noticeable improvement of 1.66% compared to the best performing VGG16. Compared to traditional TL approaches, the PD-CNN approach—which involves integrated PCC thresholding—has produced significant detection capability ranging from 2.9 to 87.29% across major standards. This highlights the critical need for domain-specific model design and training.

The PD-CNN-PCC-EELM model was 1,410 times faster than the VGG16-PCC model, which took 0.0141 seconds to test. It is 120 times quicker than Pr-CNN, 14 times faster than PD-CNN, and 7-8 times faster than other TL models, with processing durations ranging from 0.00007 to 0.00009 seconds. The substantial decrease in testing time demonstrates the efficiency advantages of the PD-CNN-PCC method. The computational load is reduced while testing by condensing the CNN model into a compact feature vector before implementing PCC thresholding. When real-time or low-latency forecasts are crucial, producing findings in 10 microseconds instead of multiple milliseconds can be very beneficial. The PD-CNN-PCC model demonstrates superior testing efficiency compared to the conventional deep TL methods.

The ROC-AUCs of DensNet201, EfficientNetB6, InceptionResNetV2, MobileNetV2, ResNet152V2, VGG16, and Xception with PCC-EELM were 98.51%, 82.39%, 98.45%, 98.84%, 97.68%, 98.83%, and 97.69%, respectively (see **Figure 12**). Furthermore, AUC-PRs of those models were achieved 93.68%, 15.56%, 88.25%, 91.46%, 90.07%, 94.20%, and 84.29%, respectively (see **Figure 13**).

**Table 5. Twenty-seven class classification performance of TL models and proposed PD-CNN-PCC model.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Models** | **Precision** | | | **Recall** | | | **F1-score** | | | **Accuracy (%)** | | | **Testing Time (Seconds)** | | | |
| **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** |
| DenseNet201-PCC | 0.831 | 0.808 | 0.829 | 0.825 | 0.815 | 0.825 | 0.805 | 0.792 | 0.804 | 82.5 | 81.5 | 82.5 | 0.00008 | 0.00007 | 0.00008 |
| EfficientNetB6-PCC | 0.121 | 0.0896 | 0.11 | 0.208 | 0.207 | 0.207 | 0.128 | 0.123 | 0.1259 | 20.87 | 20.75 | 20.75 | 0.00007 | 0.00006 | 0.00007 |
| InceptionResNetV2-PCC | 0.71 | 0.76 | 0.753 | 0.738 | 0.751 | 0.761 | 0.709 | 0.725 | 0.734 | 73.8 | 75.1 | 76.1 | 0.00009 | 0.00009 | 0.00009 |
| MobileNetV2-PCC | 0.727 | 0.758 | 0.751 | 0.752 | 0.772 | 0.778 | 0.717 | 0.738 | 0.745 | 75.25 | 77.25 | 77.87 | 0.00009 | 0.00009 | 0.00009 |
| ResNet152V2-PCC | 0.834 | 0.827 | 0.832 | 0.837 | 0.831 | 0.838 | 0.823 | 0.823 | 0.826 | 83.75 | 83.12 | 83.87 | 0.00007 | 0.00007 | 0.00007 |
| VGG16-PCC | 0.81 | 0.812 | 0.838 | 0.832 | 0.827 | 0.848 | 0.81 | 0.81 | 0.828 | 83.25 | 82.75 | 84.87 | 0.00005 | 0.00032 | 0.0141 |
| Xception-PCC | 0.669 | 0.674 | 0.654 | 0.686 | 0.72 | 0.715 | 0.648 | 0.677 | 0.672 | 68.62 | 72 | 71.5 | 0.00009 | 0.00008 | 0.00023 |
| PR-CNN | 0.828 | 0.831 | 0.831 | 0.818 | 0.826 | 0.821 | 0.807 | 0.813 | 0.809 | 0.818 | 0.826 | 82.1 | 0.00078 | 0.00094 | 0.00143 |
| PR-CNN- PCC | 0.826 | 0.812 | 0.817 | 0.83 | 0.81 | 0.821 | 0.819 | 0.797 | 0.811 | 83 | 81 | 82.12 | 0.00006 | 0.00008 | 0.00007 |
| PD-CNN | 0.8759 | 0.8731 | 0.8811 | 0.8762 | 0.8725 | 0.8775 | 0.87 | 0.8654 | 0.8712 | 87.62 | 87.25 | 87.75 | 0.01562 | 0.0156 | 0.00001 |
| PD-CNN-PCC | 0.8759 | 0.8731 | **0.8812** | 0.8762 | 0.8725 | **0.8775** | 0.8704 | 0.8654 | **0.8712** | 87.62 | 87.25 | 87.75 | 0.00006 | 0.00004 | 0.000001 |

\*Bold values indicate best results.

|  |  |
| --- | --- |
|  |  |
| **(A)** | **(B)** |
|  |  |
| **(C)** | **(D)** |
|  |  |
| **(E)** | **(F)** |
|  | |
| **(G)** | |

**Figure 12. ROC-AUC curves on (A) DensNet201, (B)** **EfficientNetB6, (C) InceptionResNetV2, (D) MobileNetV2, (E)** **ResNet152V2, (F) VGG16, and (G) Xception with PCC and ELM, RELM, EELM classifier on test-set.**

|  |  |
| --- | --- |
|  |  |
| **(A)** | **(B)** |
|  |  |
| **(C)** | **(D)** |
|  |  |
| **(E)** | **(F)** |
|  | |
| **(G)** | |

**Figure 13. AUC-PR curves on (A) DensNet201, (B)** **EfficientNetB6, (C) InceptionResNetV2, (D) MobileNetV2, (E)** **ResNet152V2, (F) VGG16, and (G) Xception with PCC and ELM, RELM, EELM classifier on test-set.**

## Optimization of PCC threshold

Once the feature extraction was completed, a total of 200 features were found, subsequently the PCC was used to get rid of unnecessary features and select the most important ones and the suggested EELM classifier was applied for classification. This led to creating the PD-CNN-PCC-EELM model**. Table 6** displays the PCC values and their performance metrics for different schemes. The PCC threshold value was determined through trial and error.

Significant features are identified in step 6 of algorithm 2 by examining threshold values with correlational metrics. A variation in the quantity of features is observed when the threshold levels are adjusted, as illustrated in **Table 6.** It is evident that an increase in the number of features results from the lower threshold values, whereas a decrease in the number of features occurs from higher thresholds. However, the performance assessments are heavily impacted by these differences in feature values. Most importantly, when combined with the suggested EELM classifier, a PCC value of 0.78 produced the best results out of all the tested PCC values, outperforming both the lower and higher thresholds. Moreover, a threshold value of 0.78 for incrementing or decrementing diminishes the detection capabilities. A greater value (≥0.79) will decrease the number of relevant aspects, resulting in less accurate findings with reductions of 1.017% in recall and accuracy, 0.863% in precision, and 1.104% in the F1-score. Correspondingly, if the PCC value falls below the threshold (≤0.77), it results in a reduction of discriminant features, thereby yielding unsatisfactory outcomes. Upon conducting time-cost analyses, optimal processing efficiency was observed at a threshold value of 0.78. With the PD-CNN-PCC model integrated into the EELM framework, testing scores registered at 0.00001, presenting a performance that is five times faster than that achieved with the thresholds greater than 0.79 and six times faster than that achieved with the thresholds less than or equal to 0.77. Notably, both the ELM and RELM classifier models exhibited superior performance compared to the other threshold levels. Consequently, based on this comprehensive examination, the threshold value of 0.78 emerged as the optimal choice for PCC in this research endeavor. **Figures 14** and **Figure 15** present the visualization of the features number and performances on various threshold levels.

Table 6. Results on proposed PD-CNN-PCC model with different PCC values.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PCC Threshold** | **Features on PD-CNN** | | **Accuracy** | | | **Precision** | | | **Recall** | | | **F1-score** | | | **Testing Time**  **(Seconds)** | | |
| **W/O PCC** | **With PCC** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** | **ELM** | **RELM** | **EELM** |
| 0.79 | 200 | 36 | 0.85 | 0.8487 | 0.8569 | 0.8558 | 0.8552 | 0.8614 | 0.85 | 0.8487 | 0.8562 | 0.8459 | 0.8436 | 0.8509 | 0.0157 | 0.0157 | 0.00005 |
| 0.78 | 200 | **39** | 0.8762 | 0.8725 | **0.8775** | 0.8759 | 0.8731 | **0.8812** | 0.8762 | 0.8725 | **0.8775** | 0.8704 | 0.8654 | **0.8712** | 0.00006 | 0.00004 | **0.000001** |
| 0.77 | 200 | 46 | 0.8537 | 0.8537 | 0.8531 | 0.856 | 0.8578 | 0.8568 | 0.8537 | 0.8537 | 0.8537 | 0.8593 | 0.8489 | 0.8479 | 0.00007 | 0.00005 | 0.00006 |
| 0.76 | 200 | 53 | 0.8412 | 0.855 | 0.8609 | 0.8477 | 0.8585 | 0.8654 | 0.8512 | 0.855 | 0.8612 | 0.8356 | 0.8505 | 0.8569 | 0.00009 | 0.00009 | 0.00009 |
| 0.75 | 200 | 60 | 0.85 | 0.8562 | 0.8538 | 0.8572 | 0.8601 | 0.8632 | 0.85 | 0.8562 | 0.8575 | 0.8453 | 0.8518 | 0.8536 | 0.0156 | 0.00005 | 0.007825 |
| 0.74 | 200 | 72 | 0.845 | 0.8475 | 0.8499 | 0.8497 | 0.8539 | 0.842 | 0.845 | 0.8475 | 0.8475 | 0.839 | 0.8423 | 0.8417 | 0.0008 | 0.0006 | 0.0007 |

\*Bold values indicate the best results.

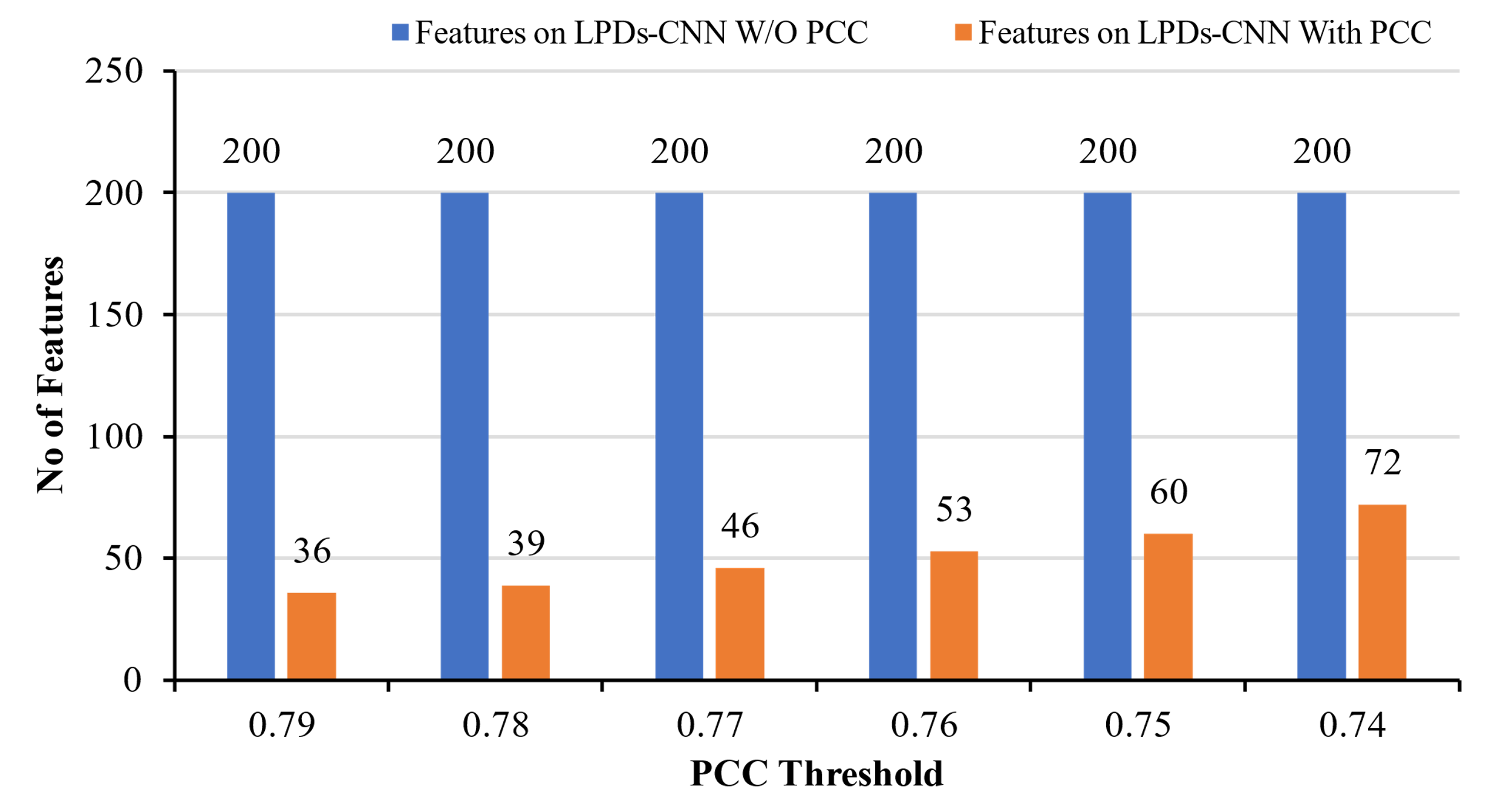


Figure 14: Number of features with various threshold values.

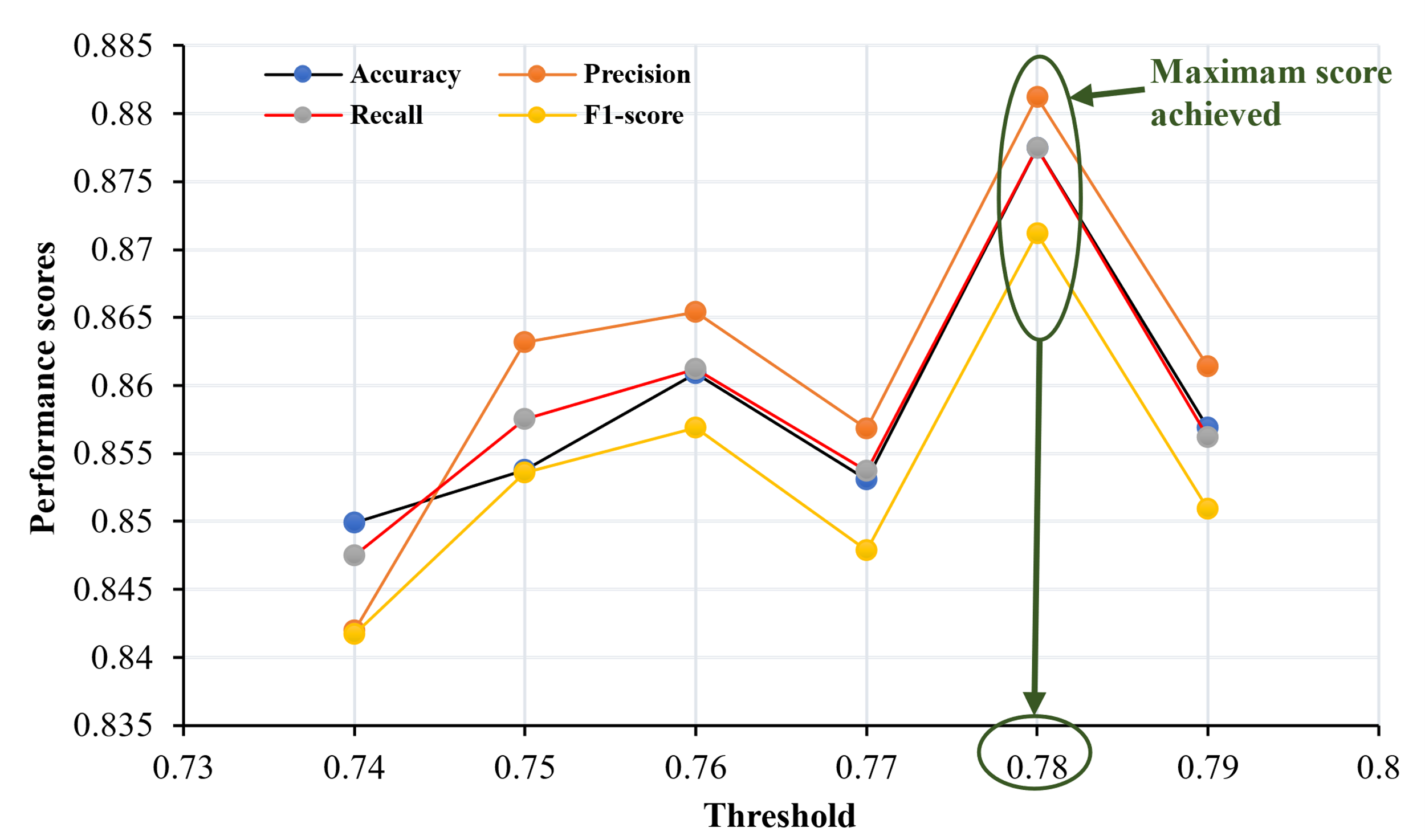


Figure 15: Performance scores on various threshold values at PD-CNN-PCC-EELM model.

## Data distribution with PCC

T-SNE is a method used to decrease nonlinear dimensions and display complicated information [64]. Data visualization is achieved by transforming data from higher dimensions into two or three dimensions, highlighting the proximity of close spots and the uniqueness of distant ones. The t-SNE method works in two steps. It first creates a probability distribution for pairs of high-dimensional objects, assigning greater probabilities to comparable pairings and lower probabilities to dissimilar pairs. Subsequently, it establishes a corresponding probability distribution in lower dimensions, minimizing the Kullback–Leibler divergence (KL divergence) between the two distributions concerning the positions of the points on the map. Evaluating the model's effectiveness involves visualizing its learned insights. The widely used t-SNE method aids in representing learning within the embedded space of a trained model. This analysis indicates that 27 class classifications exhibit reduced 2D representations for testing datasets. In **Figure 16**, the embedded space segregates sample points at the testing phase among the multiclass classification datasets.

For a comprehensive analysis, a two-dimensional t-SNE embedding were utilized on PD-CNN-PCC in **Figure 16A**. Significantly distinct categories such as "Normal esophagus (class 18)" and "Small bowel\_terminal ileum (class 25)" demonstrate a reduced occurrence of misclassification in the t-SNE embedding. The high F1 scores of 0.90 and 0.87 for the "Normal esophagus (class 18)" and "Small bowel\_terminal ileum (class 25)" classes are likely attributable to these distinct separations. Conversely, certain classes display overlap, such as "Ileocecal valve (class 16)", "Erythema (class 11)", "Esophageal varices (class 12)" and "Angioectasia (class 1)", making them susceptible to misclassification due to the absence of well-defined boundaries and limited data diversity with other classes. Additionally, t-SNE visualizations with PCC were generated for the remaining TL models (see **Figure 16**).

|  |  |
| --- | --- |
|  |  |
| (A) | (B) |
|  |  |
| (C) | (D) |
|  |  |
| (E) | (F) |
|  |  |
| (G) | (H) |
|  |  |
| (I) | (J) |

**Figure 16: T-SNE visualization in a 2D space of testing samples after training the models (A) PD-CNN, (B) PD-CNN-PCC, (C) Pr-CNN-PCC, (D) DensNet201-PCC, (E)** **EfficientNetB6-PCC, (F) InceptionResNetV2-PCC, (G) MobileNetV2-PCC, (H)** **ResNet152V2-PCC, (I) VGG16-PCC, and (J) Xception-PCC with EELM for twenty-seven test-set classification.**

## Computational time and resource comparison

**Figure 17** illustrates the comparison of resources among the models. The PD-CNN-PCC model stands out among the evaluated models across various performance measures. PD-CNN-PCC is notable for its reduction in architectural complexity compared to the other models, consisting of a mere 0.815 million total parameters. On the other hand, Pr-CNN-PCC required 1.57 times higher number of parameters. With only 0.812 million trainable parameters, it is the most parameter-efficient model compared to Pr-CNN by 1.58% and other TL models by 5.06 to 45.74% (see **Table 7**). Furthermore, compared to DensNet201-PCC models, which have 710 layers and are more difficult to analyze, the proposed model’s 24-layer design denotes a streamlined architecture and a 28.58% decrease in the number of layers. The compactness of PD-CNN-PCC is evident in its size, which is just 9.79 megabytes, positioning it as one of the smallest models in comparison. This model's compact size and fewer layers enhance its efficiency for deployment on edge devices, enabling real-time processing. PD-CNN-PCC is around 5 times faster than Pr-CNN-PCC, 10 times faster than EfficientNetB6-PCC, and nearly 140 times faster than VGG16-PCC when considering testing time with EELM. This computational time analysis demonstrates PD-CNN-PCC's remarkable speed and efficiency compared to others. In conclusion, the PD-CNN-PCC model has achieved the best results while maintaining optimal resource requirements, showcasing remarkable efficiency and performance.

Table 7. Comparative resource analysis among the trained models on test set.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Performance Criteria** | **Pr-CNN-PCC** | **PD-CNN-PCC** | **DenseNet201-PCC** | **EfficientNetB6-PCC** | **Inception**  **ResNetV2-PCC** | **MobileNet**  **V2-PCC** | **ResNet**  **152V2-PCC** | **VGG16-PCC** | **Xception-PCC** |
| Total Parameters (Million) | 2.095 | **0.815** | 36.22 | 78.91 | 60.83 | 23.44 | 92.09 | 19.64 | 54.62 |
| Trainable Parameters (Million) | 2.093 | **0.812** | 17.9 | 37.95 | 6.5 | 21.18 | 33.76 | 4.92 | 33.76 |
| Number of Layers | 41 | **24** | 710 | 669 | 783 | 157 | 567 | 22 | 135 |
| Size (Megabytes) | 28.49 | **9.79** | 111 | 606.6 | 283.32 | 257.54 | 625.07 | 115.35 | 477.6 |
| Testing Time (Seconds)- EELM | 0.00005 | **0.000001** | 0.000098 | 0.000099 | 0.000085 | 0.000076 | 0.000086 | 0.0141 | 0.000069 |

\*Bold values indicate the best score.

|  |
| --- |
|  |
| **(A)** |
|  |
| **(B)** |
|  |
| **(C)** |

Figure 17: Computational resource (A) parameters, (B) number of layers and (C) size comparisons among the proposed PD-CNN-EELM with TL models.

## Interpretability with XAI

The proposed framework enhanced transparency and interpretability in its decision-making processes by utilizing multiple XAI approaches. Various XAI techniques were generated to provide a visual representation that highlights the affected positive regions of GI diseases. Several samples of images were selected randomly from the original dataset to generate the XAI prediction (**Figure 18**). In the first row, the sample image is of the resection margins class. The heatmap visualizes the affected area of the sample image that is important for identification of the class. But the guided heatmap visualizes the particular affected region of the image more precisely than the normal Grad-Cam. Remarkably, guided grad-CAM reveals the affected area integrating the original image as background to understand the specific features of that class prediction. The sixth column indicates that the model mostly focuses on the center of the resection margin regions. The last column represents guided saliency map visualization. The brighter pixels of guided saliency map precisely indicate the affected pixels rather than the irrelevant background. Similarly, in the third row, the sample image shows colon polyp, and from the visualization, it is clear that the guided heatmap is indicating polyp area more accurately than the heatmap. Similarly, guided grad-CAM focuses on the polyp more precisely. Also, the brighter pixels of the guided saliency map accurately show affected pixels. The visualization indicates that the PD-CNN-PCC-EELM model accurately detects the features of different GI diseases.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Selected Diseases Class Name** | **Original Image** | **Heatmap**  **visualization** | **Guided Heatmap**  **visualization** | **Grad-Cam**  **visualization** | **Guided Grad-Cam**  **visualization** | **Guided Saliency mapping** |
| **Resection margins** |  |  |  |  |  |  |
| **Blood in lumen** |  |  |  |  |  |  |
| **Colon polyps** |  |  |  |  |  |  |
| **Colorectal Cancer** |  |  |  |  |  |  |
| **Dyed-resection margins** |  |  |  |  |  |  |
| **Ileocecal valve** |  |  |  |  |  |  |

Figure 18: Gradient-weighted class activation mapping of some diseases using GastroVision dataset on PD-CNN feature extraction technique.

An exhaustive analysis of several GI features led to the generation of Shapley values, which produced pixelated visualizations. The analysis demonstrated a clear pattern, where the red pixels were highly effective in accurately identifying specific classes. On the other hand, the presence of blue pixels indicated a higher probability of being far away from the target class. In order to obtain the SHAP result, a set of test samples were randomly selected for prediction. **Figure 19** displayed SHAP results which use faint grey backgrounds combined with the original image. The red pixels in the SHAP explanation image on the top row represent the presence of Accessory tools (C-0). On the other hand, the lack of blue pixels and a reduced number of red pixels completely removed other class groupings. The second row showed a clear pattern where red pixels in the SHAP explanation images represented the Colon polyps (C-6) class. However, an excess of red pixels in the SHAP explanation image correctly indicated that the image belonged to that class. Subsequently, blue pixels in the SHAP explanation graphics for other classes indicated the lack of probability. The red pixels in the third row of the SHAP explanation image suggest strong evidence of Gastric polyps (C-16) GI class disease. Significantly, C-19 exhibited a very competitive XAI prediction compared to C-26 and C-19. Red pixels were depicted in both classes. However, there was a significant presence of blue pixels in the C-26 projection, resulting in an increased false projection for C-26. In addition, rows 4, 5, 6, and 7 correctly recognized disease classifications by emphasizing red pixels in particular areas. The visual SHAP explanations validated the model's outcomes, offering doctors a deeper understanding of specific diseases categories.

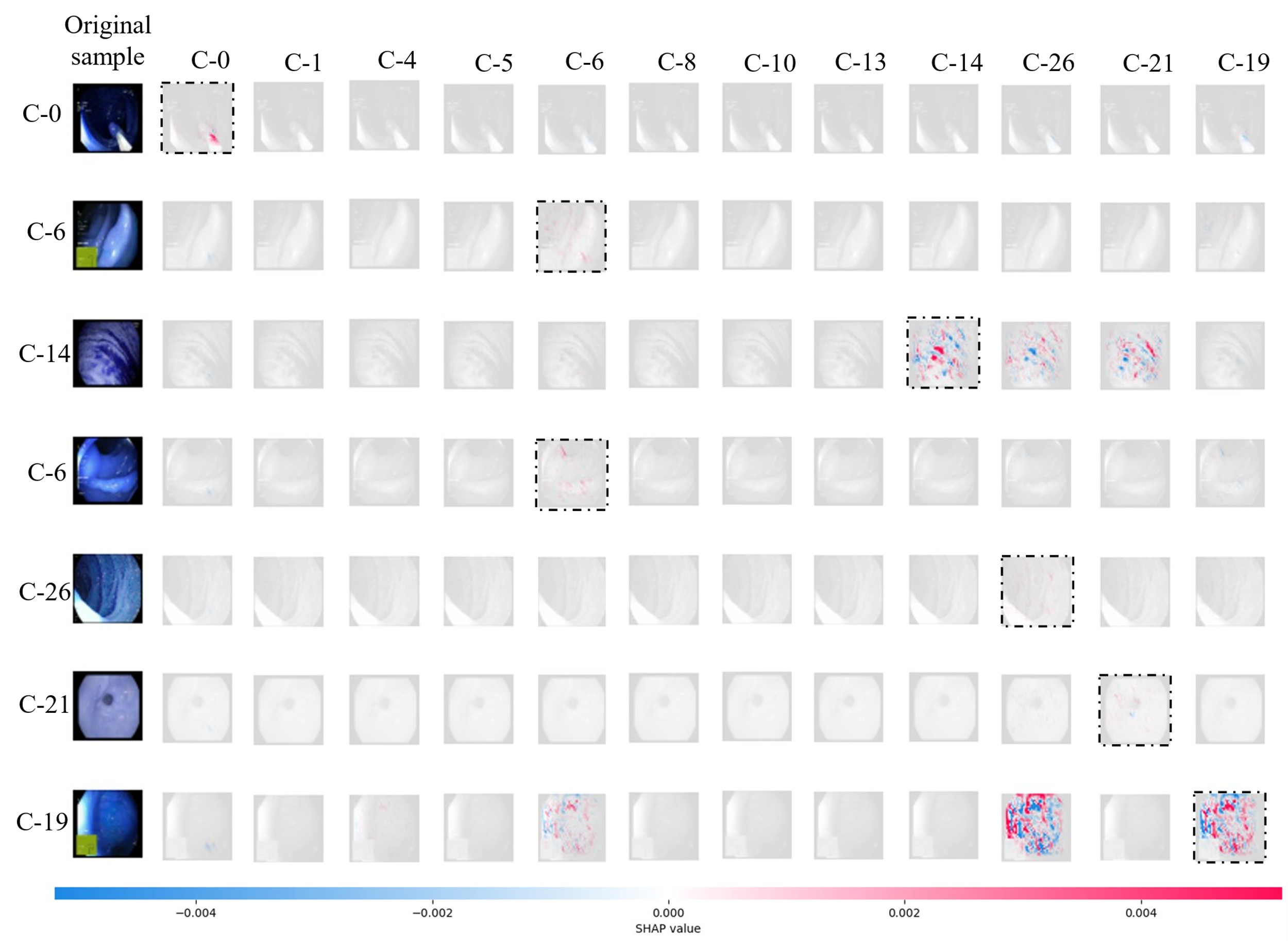


Figure 19: Sapley Additive exPlanations (SHAP) images for proposed model (Here C indicates selected Class).

## Discussion, Limitation, and Future Work

While numerous studies have focused on GI disease diagnosis, identification, and segmentation, there needs to be more research on multi-class classification encompassing a broad spectrum of GI diseases. This discussion presents a thorough review of experimental results achieved using the proposed method, which effectively categorizes 27 types of GI tract issues for the first time. The proposed approach comprises four main steps: dataset preprocessing, feature extraction, feature selection, and classification with interpretability. The preprocessing step ensures refined and standardized input data, enhancing performance and accuracy in classification tasks. The PD-CNN has extracted 200 features, but the feature selection stage retains only 39 features using the PCC. EELM performs better than other classifiers, exhibiting greater precision, recall, F1 score, accuracy, ROC-AUC and AUC-PR performance.

**Table 8** compares previous classification models with the proposed PD-CNN-PCC-EELM model. The overview indicates that, in [20], the MobileNetV2, with 3.4 million parameters and 154 layers, has attained an accuracy of 97.68% for five classes. Nouman et al. [21] also employed MobileNetV2 for the 5-class classification task, utilizing 3.4 million parameters and 1210 extracted features. Gunasekaran et al. [19] obtained a 95% accuracy rate for eight classes by employing the ensemble TL model, which comprised 66.94 million features. Öztürk et al. [65] combined a ResNet50 TL model with a residual LSTM classifier to achieve 98.05% accuracy in an eight-class classification task using the Kvasir dataset. Other researchers [66–70] utilized TL-based models for classifying gastrointestinal disorders. As previously stated, all these models contained many parameters, ranging from 5.3 to 138 million. Most of these models prioritize extracting a large number of features (943-1000), which lead to higher computational requirements for both training and inference. Large feature sets in models can result in slower inference times, making them unsuitable for practical use.

On the other hand, The PD-CNN-PCC-EELM model achieved a comparable accuracy of 86.13% across 27 classes with only 0.815 million parameters, 24 layers, and 39 features. The proposed model, which is a lightweight NN, attained notable accuracy for all 27 classes within a testing time of 0.00001 seconds. The design of the model is optimized by simultaneously running the first five CLs to improve feature extraction. The model succeeds in classification performance and computational requirement compared to SOTA TL models, as shown in **Table 5 and Table 7**. It reduces parameters, layers, and testing time while maintaining acceptable accuracy. The use of SHAP, heatmap, Grad-CAM, guided Grad-CAM and guided saliency map has improved the interpretability of the proposed model by showing it focuses on relevant image regions to extract useful features.

A few datasets like Kvasir, HyperKvasir, Kvasir-Capsule and KID provide multiple GI findings. However, Kvasir-Capsule and KID were video capsule endoscopy datasets containing a minimum number of classes. Most previous studies have demonstrated their proposed models using these datasets, classifying between 5 to 8 types of GI diseases. For the first time, GastroVision dataset contained 27 classes (highest) and covered more labeled classes of anatomical landmarks, pathological findings, and normal findings. Additionally, baseline results have been established on this dataset for GI disease detection and classification on upper, lower, and combined GI tract, offering valuable research resources for advancing GI endoscopy studies.

Although the suggested method demonstrates comparable outcomes with a lightweight model, it does have some drawbacks. The data in **Table 8** clearly demonstrates that the model's classification accuracy is inferior when compared to other existing studies. The primary cause of this lower classification accuracy is that the current study worked on a large dataset with 27 different classes. Also, the GastroVision dataset has diverse types of images of different GI diseases. Another issue affecting the classification accuracy is the resizing of images during the preprocessing stage, which results in a significant loss of resolution. Moreover, some classes of the dataset contain only very few sample images such as ulcer (class 26) has only 6 sample images and esophageal varices (class 12) has 7 sample images. Therefore, there is still potential for additional improvement in the model. Future endeavors of the authors will focus on improving the model's efficiency by balancing the GastroVision dataset. The researchers plan to gather and create a more evenly distributed dataset to enhance classification results. Additionally, as a future avenue of exploration, implementing a real-world hardware-based design is envisioned to provide enhanced visualization capabilities for the proposed model, thus improving its practical utility.

Table 8: Previous studies result in comparison with proposed PD-CNN-PCC-EELM model.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ref.** | **Dataset** | **Number of Sample Images** | **Number of Class** | **Feature Extractor** | **Parameters**  **(Million)** | **Number of Layers** | **Model Size (MB)** | **Number of Features** | **Best Classifier** | **Testing Accuracy**  **(%)** | **Testing Time**  **(Seconds)** | **Real-time XAI** |
| Nouman et al. [21] | KvasirV-2 and Hyper-Kvasir | 4854 | 5 | MobileNetV2 | 3.4 | 154 | - | 1210 | Softmax | 96.40 | --- | No |
| Noor et al. [20] | KvasirV-2 and Hyper--Kvasir | 4854 | 5 | MobileNetV2 | 3.4 | 154 | - | 810 | Softmax | 97.68 | --- | Yes (Grad-CAM) |
| Gunasekaran et al. [19] | KvasirV-2 | 8000 | 8 | Ensemble Model (InceptionV3, DenseNet201, ResNet201) | 66.94 | --- | - | --- | --- | 95 | --- | No |
| Öztürk et al. [65] | Kvasir | 6000 | 8 | ResNet50 | 23.9 | --- | - | --- | Residual LSTM | 98.05 | --- | No |
| Yogapriya et al. [70] | KvasirV-2 | 8000 | 8 | VGG16 | 138 | **16** | - | --- | --- | 96.33 | --- | No |
| Hmoud Al-Adhaileh et al. [71] | Kvasir | 5000 | 5 | AlexNet | 62.3 | 25 | - | 1000 | --- | 97.00 | --- | No |
| Lonseko et al. [66] | KvasirV-2 | 8000 | 8 | Deep CNN based Attention | 19.92 | --- | - | --- | --- | 93.19 | --- | Yes  (Heatmap) |
| Ramzan et al. [67] | Kvasir | 4000 | 8 | InceptionNetV3, GITNet | --- | 50 | - | 1000 | QSVM | 99.32 | 0.0016 | Yes (Grad-CAM) |
| Thomas Abraham et al. [68] | Kvasir | 5000 | 5 | Custom CNN with EfficientNetB0 | 5.3 | --- | - | --- | --- | 98.01 |  | Yes (Grad-CAM) |
| Khan et al. [69] | KvasirV-2 | 8000 | 8 | Darknet 53, Xception | --- | --- | - | 943 | Ensemble-based Subspace KNN (ESKNN) | 98.25 | --- | No |
| **Proposed Model** | **GastroVision** | **8000** | **27** | **PD-CNN** | **0.815** | **24** | **9.79** | **39** | **EELM** | **87.75** | **0.000001** | **Yes** (Heatmap, guided Heatmap Grad-CAM, Guided Grad-CAM, Salience Mapping and SHAP) |

\*Bold values indicate the best result.

# Conclusion

This study presented a novel method for categorizing gastrointestinal (GI) tract disorders precisely by integrating the parallel Depthwise Separable CNN (PD-CNN) feature extractor and PCC feature selector with the Ensemble ELM (EELM) classifier. The proposed model, consisting of 24 layers and 0.815 million parameters, effectively categorizes twenty-seven types of different anatomical positions of GI diseases while decreasing the computing burden. The EELM's testing duration was only 0.0001 seconds after integrating PCC by reducing irrelevant features. The hybrid EELM classifier improves classification performance by combining ELM and RELM algorithms. The proposed approach has shown excellent classification performance, with accuracy, precision, recall, f1, ROC-AUC, and AUC-PR values of 88.12±0.332%, 87.75±0.348%, 87.12±0.324%, 87.75%, 98.89%, and 92%, respectively. Also, the proposed model is compact, with a size of only 9.79MB, which makes it suitable for practical use. Moreover, its low computational requirements (parameter, layer, size) would enable its deployment on cost-effective edge devices quite easily. Furthermore, combining real-time explainable (XAI) helps the medical experts by offering a dependable explanation of the model's results in revealing the right type of GI disorder. In conclusion, the PD-CNN-PCC-EELM technique greatly enhances the accuracy of classifying 27 types of GI tract diseases and is easy to implement and evaluate in real-world scenarios. However, there is still scope for further improvement in accuracy.

**Declaration of Competing Interest:** The authors affirm that they do not possess any identifiable conflicting financial interests or personal relationships that might have been perceived to impact the findings presented in this paper.

**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.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.

[3] M.A. Khan, M.A. Khan, F. Ahmed, M. Mittal, L.M. Goyal, D.J. Hemanth, S.C. Satapathy, Gastrointestinal diseases segmentation and classification based on duo-deep architectures, Pattern Recognit. Lett. 131 (2020) 193–204.

[4] M. Sharif, M. Attique Khan, M. Rashid, M. Yasmin, F. Afza, U.J. Tanik, Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images, J. Exp. Theor. Artif. Intell. 33 (2021) 577–599.

[5] M.A. Khan, S. Kadry, M. Alhaisoni, Y. Nam, Y. Zhang, V. Rajinikanth, M.S. Sarfraz, Computer-aided gastrointestinal diseases analysis from wireless capsule endoscopy: a framework of best features selection, IEEE Access. 8 (2020) 132850–132859.

[6] 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.

[7] 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.

[8] G. Iddan, G. Meron, A. Glukhovsky, P. Swain, Wireless capsule endoscopy, Nature. 405 (2000) 417.

[9] S. Fan, L. Xu, Y. Fan, K. Wei, L. Li, Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images, Phys. Med. Biol. 63 (2018) 165001.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[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] C. Ye, J.L. Prince, A Bayesian approach to fiber orientation estimation guided by volumetric tract segmentation, Comput. Med. Imaging Graph. 54 (2016) 35–47.

[19] H. Gunasekaran, K. Ramalakshmi, D.K. Swaminathan, M. Mazzara, GIT-Net: an ensemble deep learning-based GI tract classification of endoscopic images, Bioengineering. 10 (2023) 809.

[20] M.N. Noor, M. Nazir, I. Ashraf, N.A. Almujally, M. Aslam, S. Fizzah Jilani, GastroNet: A robust attention‐based deep learning and cosine similarity feature selection framework for gastrointestinal disease classification from endoscopic images, CAAI Trans. Intell. Technol. (2023).

[21] M. Nouman Noor, M. Nazir, S.A. Khan, O.-Y. Song, I. Ashraf, Efficient gastrointestinal disease classification using pretrained deep convolutional neural network, Electronics. 12 (2023) 1557.

[22] S. Ali, M. Dmitrieva, N. Ghatwary, S. Bano, G. Polat, A. Temizel, A. Krenzer, A. Hekalo, Y.B. Guo, B. Matuszewski, Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy, Med. Image Anal. 70 (2021) 102002.

[23] D.-Y. Liu, T. Gan, N.-N. Rao, Y.-W. Xing, J. Zheng, S. Li, C.-S. Luo, Z.-J. Zhou, Y.-L. Wan, Identification of lesion images from gastrointestinal endoscope based on feature extraction of combinational methods with and without learning process, Med. Image Anal. 32 (2016) 281–294. https://doi.org/https://doi.org/10.1016/j.media.2016.04.007.

[24] D. Jha, S. Ali, S. Hicks, V. Thambawita, H. Borgli, P.H. Smedsrud, T. de Lange, K. Pogorelov, X. Wang, P. Harzig, M.-T. Tran, W. Meng, T.-H. Hoang, D. Dias, T.H. Ko, T. Agrawal, O. Ostroukhova, Z. Khan, M. Atif Tahir, Y. Liu, Y. Chang, M. Kirkerød, D. Johansen, M. Lux, H.D. Johansen, M.A. Riegler, P. Halvorsen, A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging, Med. Image Anal. 70 (2021) 102007. https://doi.org/https://doi.org/10.1016/j.media.2021.102007.

[25] P.M. Szczypiński, R.D. Sriram, P.V.J. Sriram, D.N. Reddy, A model of deformable rings for interpretation of wireless capsule endoscopic videos, Med. Image Anal. 13 (2009) 312–324. https://doi.org/https://doi.org/10.1016/j.media.2008.12.002.

[26] M. Ye, S. Giannarou, A. Meining, G.-Z. Yang, Online tracking and retargeting with applications to optical biopsy in gastrointestinal endoscopic examinations, Med. Image Anal. 30 (2016) 144–157. https://doi.org/https://doi.org/10.1016/j.media.2015.10.003.

[27] X. Guo, Y. Yuan, Semi-supervised WCE image classification with adaptive aggregated attention, Med. Image Anal. 64 (2020) 101733. https://doi.org/https://doi.org/10.1016/j.media.2020.101733.

[28] A. Perperidis, K. Dhaliwal, S. McLaughlin, T. Vercauteren, Image computing for fibre-bundle endomicroscopy: A review, Med. Image Anal. 62 (2020) 101620. https://doi.org/https://doi.org/10.1016/j.media.2019.101620.

[29] M.K. Bashar, T. Kitasaka, Y. Suenaga, Y. Mekada, K. Mori, Automatic detection of informative frames from wireless capsule endoscopy images, Med. Image Anal. 14 (2010) 449–470. https://doi.org/https://doi.org/10.1016/j.media.2009.12.001.

[30] A. Musha, R. Hasnat, A. Al Mamun, E.P. Ping, T. Ghosh, Computer-Aided Bleeding Detection Algorithms for Capsule Endoscopy: A Systematic Review, Sensors. 23 (2023) 7170.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] E. Sivari, E. Bostanci, M.S. Guzel, K. Acici, T. Asuroglu, T. Ercelebi Ayyildiz, A New Approach for Gastrointestinal Tract Findings Detection and Classification: Deep Learning-Based Hybrid Stacking Ensemble Models, Diagnostics. 13 (2023) 720.

[36] F. Rustam, M.A. Siddique, H.U.R. Siddiqui, S. Ullah, A. Mehmood, I. Ashraf, G.S. Choi, Wireless capsule endoscopy bleeding images classification using CNN based model, IEEE Access. 9 (2021) 33675–33688.

[37] 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.

[38] H. Borgli, V. Thambawita, P.H. Smedsrud, S. Hicks, D. Jha, S.L. Eskeland, K.R. Randel, K. Pogorelov, M. Lux, D.T.D. Nguyen, HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy, Sci. Data. 7 (2020) 283.

[39] 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.

[40] M. Alhajlah, M.N. Noor, M. Nazir, A. Mahmood, I. Ashraf, T. Karamat, Gastrointestinal Diseases Classification Using Deep Transfer Learning and Features Optimization, C. Mater. Contin. 75 (2023) 2227–2245.

[41] 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.

[42] 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.

[43] D. Jha, V. Sharma, N. Dasu, N.K. Tomar, S. Hicks, M.K. Bhuyan, P.K. Das, M.A. Riegler, P. Halvorsen, U. Bagci, T. de Lange, GastroVision: A Multi-class Endoscopy Image Dataset for Computer Aided Gastrointestinal Disease Detection BT - Machine Learning for Multimodal Healthcare Data, in: A.K. Maier, J.A. Schnabel, P. Tiwari, O. Stegle (Eds.), Springer Nature Switzerland, Cham, 2024: pp. 125–140.

[44] L. Kaiser, A.N. Gomez, F. Chollet, Depthwise separable convolutions for neural machine translation, ArXiv Prepr. ArXiv1706.03059. (2017).

[45] 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.

[46] S. Ding, X. Xu, R. Nie, Extreme learning machine and its applications, Neural Comput. Appl. 25 (2014) 549–556. https://doi.org/10.1007/s00521-013-1522-8.

[47] X. Shi, Q. Kang, J. An, M. Zhou, Novel L1 Regularized Extreme Learning Machine for Soft-Sensing of an Industrial Process, IEEE Trans. Ind. Informatics. 18 (2022) 1009–1017. https://doi.org/10.1109/TII.2021.3065377.

[48] 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.

[49] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: Int. Conf. Mach. Learn., PMLR, 2019: pp. 6105–6114.

[50] 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.

[51] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018: pp. 4510–4520.

[52] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, (2016) 770–778. http://image-net.org/challenges/LSVRC/2015/ (accessed July 6, 2023).

[53] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, ArXiv Prepr. ArXiv1409.1556. (2014).

[54] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017: pp. 1251–1258.

[55] J. Benesty, J. Chen, Y. Huang, I. Cohen, Pearson Correlation Coefficient BT - Noise Reduction in Speech Processing, in: I. Cohen, Y. Huang, J. Chen, J. Benesty (Eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2009: pp. 1–4. https://doi.org/10.1007/978-3-642-00296-0\_5.

[56] E. Tjoa, C. Guan, A survey on explainable artificial intelligence (xai): Toward medical xai, IEEE Trans. Neural Networks Learn. Syst. 32 (2020) 4793–4813.

[57] 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.

[58] W. Jin, X. Li, M. Fatehi, G. Hamarneh, Guidelines and evaluation of clinical explainable AI in medical image analysis, Med. Image Anal. 84 (2023) 102684. https://doi.org/https://doi.org/10.1016/j.media.2022.102684.

[59] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proc. IEEE Int. Conf. Comput. Vis., 2017: pp. 618–626.

[60] L. Chen, J. Chen, H. Hajimirsadeghi, G. Mori, Adapting grad-cam for embedding networks, in: Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis., 2020: pp. 2794–2803.

[61] S. Yang, G. Berdine, Interpretable artificial intelligence (AI)–saliency maps, Southwest Respir. Crit. Care Chronicles. 11 (2023) 31–37.

[62] D.M.W. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, ArXiv Prepr. ArXiv2010.16061. (2020).

[63] 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.

[64] S. Arora, W. Hu, P.K. Kothari, An Analysis of the t-SNE Algorithm for Data Visualization, in: S. Bubeck, V. Perchet, P. Rigollet (Eds.), Proc. 31st Conf. Learn. Theory, PMLR, 2018: pp. 1455–1462. https://proceedings.mlr.press/v75/arora18a.html.

[65] Ş. Öztürk, U. Özkaya, Residual LSTM layered CNN for classification of gastrointestinal tract diseases, J. Biomed. Inform. 113 (2021) 103638. https://doi.org/https://doi.org/10.1016/j.jbi.2020.103638.

[66] Z.M. Lonseko, P.E. Adjei, W. Du, C. Luo, D. Hu, L. Zhu, T. Gan, N. Rao, Gastrointestinal Disease Classification in Endoscopic Images Using Attention-Guided Convolutional Neural Networks, Appl. Sci. 11 (2021). https://doi.org/10.3390/app112311136.

[67] M. Ramzan, M. Raza, M.I. Sharif, F. Azam, J. Kim, S. Kadry, Gastrointestinal tract disorders classification using ensemble of InceptionNet and proposed GITNet based deep feature with ant colony optimization, PLoS One. 18 (2023) e0292601. https://doi.org/10.1371/journal.pone.0292601.

[68] J. V Thomas Abraham, A. Muralidhar, K. Sathyarajasekaran, N. Ilakiyaselvan, A Deep-Learning Approach for Identifying and Classifying Digestive Diseases, Symmetry (Basel). 15 (2023). https://doi.org/10.3390/sym15020379.

[69] Z.F. Khan, M. Ramzan, M. Raza, M.A. Khan, K. Iqbal, T. Kim, J.H. Cha, Deep Convolutional Neural Networks for Accurate Classification of Gastrointestinal Tract Syndromes, Comput. Mater. Contin. 78 (2024) 1207–1225. https://doi.org/10.32604/cmc.2023.045491.

[70] J. Yogapriya, V. Chandran, M.G. Sumithra, P. Anitha, P. Jenopaul, C. Suresh Gnana Dhas, Gastrointestinal Tract Disease Classification from Wireless Endoscopy Images Using Pretrained Deep Learning Model, Comput. Math. Methods Med. 2021 (2021). https://doi.org/10.1155/2021/5940433.

[71] M. Hmoud Al-Adhaileh, E. Mohammed Senan, F.W. Alsaade, T.H.H. Aldhyani, N. Alsharif, A. Abdullah Alqarni, M.I. Uddin, M.Y. Alzahrani, E.D. Alzain, M.E. Jadhav, Deep Learning Algorithms for Detection and Classification of Gastrointestinal Diseases, Complexity. 2021 (2021) 6170416. https://doi.org/10.1155/2021/6170416.