Azerbaijan Journal of High Performance Computing, Vol 6, Issue 1, 2023, pp. 49-76
https://doi.org/10.32010/26166127.2023.6.1.49.76

A Unified Paradigm of Classifying GI Tract Diseases in Endoscopy Images Using Multiple Features Fusion

Muhammad Afraz, Abdul Muiz Fayyaz, Abdul Haseeb

COMSATS University Islamabad Wah Campus, Pakistan, afrazee.rcs@gmail.com, abdulmuizfayyaz@gmail.com, haseeb.rch@gmail.com

*Correspondence: Muhammad Afraz, COMSATS University Islamabad Wah Campus, Pakistan, afrazee.rcs@gmail.

Abstract

The automatic identification of Gastrointestinal (GI) tract diseases in endoscopy images has been associated with the domain of medical imaging and computer vision. Its classification has various challenges, including color, low contrast, lesion shape, and complex background. A Deep features-based method for the classification of gastrointestinal disease is implemented in this article. The method suggested involves four significant steps: preprocessing, extraction of handcrafted, and deep Convolutional neural network features (Deep CNN), selection of solid features, fusion, and classification. 3D-Median filtering in the preprocessing stage increases the lesion contrast. The second stage extracts the features centered on the shape. The extracted features are of three types: HOG features, ResNet50, and Xception. Principal Component Analysis (PCA) is chosen to select extracted features, combined by concatenating them in a single array. A support vector system eventually categorizes fused features into multiple classes. The Kvasir dataset is used for the proposed model. The SVM has outstanding efficiency reached 96.6 percent, showing the proposed system's robustness.

Keyword: GI Tract Diseases, WCE, Feature Extraction, Deep Features, Feature Selection, Classification

1. Introduction

Nowadays, gastrointestinal (GI) Tract disease detection and classification are dynamic research areas in medical imaging (Feng, R., et al., 2020, April). Multiple diseases may occur in the entire GI Tract, including Ulcers, Polyps, Esophagitis, and Bleeding (Wickstrøm, K., Kampffmeyer, M., & Jenssen, R., 2018, September; Høiland, T. N., 2017). Polyps, bleeding, and ulcers are widespread diseases around the globe (Chen, T., et al., 2020). In the human digestive system, a polyp is a cluster of cells formed inside the colon or intestine, possibly leading to colon cancer (Poorneshwaran, J. M., Kumar, S. S., Ram, K., Joseph, J., & Sivaprakasam, M., 2019, July; Khan, M. A., Sharif, M., Akram, T., Yasmin, M., & Nayak, R. S., 2019). Colon cancer is divided into two main categories, neoplastic and non-neoplastic. Adenoma and Serrated are types of neoplastic polyps with a significant risk of colorectal cancer. Hyperplastic polyps

and inflammatory polyps are non-neoplastic and are not much cancerous (Tjalma, J. J., et al., 2016). In mature adults, the seventh prevalent disease is the esophagus. The intensified diagnosis and screening of esophagus disorders may increase the probability of recovery between 19 and 80 percent. Consequently, the rate of human death will be reduced if disorders of the GI tract can be identified and treated earlier.

To detect different types of GI Tract diseases, as shown in Figure 1, healthcare practitioners use the endoscopy process. Traditionally, the stomach, esophagus, and large intestine were inspected by push endoscopy. It was impossible to examine the small intestine for bleeding, ulcers, and polyps (Yu, L., Yuen, P. C., & Lai, J., 2012, November). For the first time, as an alternative, Wireless Capsule Endoscopy (WCE) was introduced to inspect the patient's GI Tract, especially the small intestine, directly because it was impossible to reach the small intestine through push endoscopy. Wireless capsule endoscopy (WCE) is a painless and non-invasive way to inspect a complete digestive system (Sharif, M., et al., 2021). The capsule is swallowed by the patient, who moves throughout the GI Tract and takes photos of over fifty thousand (Majid, A., et al., 2020). The captured frames are forwarded to a receiver tied with a belt on the waist of the patient. The camera's battery drops dead within 8 hours, and the capsule is conceded in the patient's feces and is wasted. Transmitted images are then fetched from the belt and passed through different algorithms for diagnosis. In 2000, WCE was initiated and approved by Food and Drug Administration (FDA) in 2002. It is an essential procedure for slight bowel penetration. (Berens, J., Mackiewicz, M., & Bell, D., 2005, April).

Currently, the primary analysis and testing techniques are colonoscopy, gastroscopy, and computed tomography (CT) (Hwang, S., Oh, J., Cox, J., Tang, S. J., & Tibbals, H. F., 2006, March; Saba, T., Mohamed, A. S., El-Affendi, M., Amin, J., & Sharif, M., 2020) tests. All these are costly, time-sensitive, and inconvenient for the person (de Lange, T., Larsen, S., & Aabakken, L., 2005). A certain amount of patient distress has also been expected due to endoscopic procedures. They all involve intensive tools and medical experts, making screening entire populations difficult. There are several methods to screen the GI tract, but the most effective method for early identification is population-wide examinations (El-Matary, W., 2008). However, there are restrictions on the sensitivity, specificity, and cost of the current processes. The suggested colon testing technique is endoscopy, but it is a challenging operation that needs plenty of actions from medical staff. Besides, polyps are often missed because of the physician's tiredness or because an endoscope may be unable to reach a particular portion of the colon (Sharif, M., et al., 2021; Tajbakhsh, N., Gurudu, S. R., & Liang, J., 2013). Precise identification of GI tract diseases is complicated due to the complex features of endoscopic images. The contextual nature of polyp images, such as structure, scale, and correlation with the adjacent ones, differ, and physicians may locate anomalies in all situations with considerable reliability.

In comparison, endoscopic video generates more than 55000 frames per patient

test. The filmed bizarre scenes only consume 5 percent of the total scenarios obtained, so it is stressful for health professionals to view an abnormal scenario manually for the entire set of images (Yuan, Y., Wang, J., Li, B., & Meng, M. Q. H., 2015; Khan, M. A., et al., 2020). It is helpful for physicians to assist them by automatically reviewing videos during manual examinations and giving live feedback. One way to achieve this might be through computer vision. Computer-Aided Diagnosis (CAD) and Automated Computer Diagnosis (ACD) have emerged more recently, both of which could make the process more automated (Mughal, B., Sharif, M., & Muhammad, N., 2017; de Souza-Filho, E. M., & de Amorim Fernandes, F., 2021; Mughal, B., Muhammad, N., & Sharif, M., 2019). CAD seeks to assist physicians during examinations by getting both the doctor and a disease search detection scheme, generating a synergistic

Impact where the machine can provide a second opinion (Li, P., et al., 2015). ACD tries to optimize the method so that a doctor is unnecessary during the initial processing. This could allow patients to conduct their initial testing, placing less pressure on the healthcare system, which improves the scalability by reducing costs and enabling a more considerable proportion of the population to be screened.

Many of the CAD systems developed recently helped doctors in their clinics significantly. Most CAD detection and classification methods for GI diagnosis are based on supervised algorithms (Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L., 2018). Such CAD systems automatically diagnose GI diseases by extracting features from WCE images. Handcrafted characteristics, including shape and color, for treating abnormalities were extracted from different CAD systems (Ali, H., Yasmin, M., Sharif, M., & Rehmani, M. H., 2018; Khan, M. A., Javed, M. Y., Sharif, M., Saba, T., & Rehman, A., 2019, April). Recent profound learning developments perform well on classical features of many previously underperforming CAD systems. The high-level features of the fully connected neural network are calculated (Sharif, M. I., Khan, M. A., Alhussein, M., Aurangzeb, K., & Raza, M., 2021). Developing an automatic disease recognition CAD system is therefore essential in time. While several methods have been developed for detecting infections in the GI Tract, focusing on feature selection and optimization techniques (Shahzad, A., Raza, M., Shah, J. H., Sharif, M., & Nayak, R. S., 2022), using and optimizing features remain challenging. The new advancement of machine vision and analytical procedures influences the transition of Diagnostic imaging from specialist doctor's manual examination to the extensive use of integrated diagnostic systems (Saba, T., Sameh, A., Khan, F., Shad, S. A., & Sharif, M., 2019). It is essential to monitor objects within the human body. Surfaces and anomalies may have different shapes, forms, textures, patterns, and viewpoints. They may obscure particular objects or barriers. The principle aspect of the approach is to determine the identity of GI tract diseases preliminarily by quantifying global image patterns and, if verified, to locate them using fully Convolutional neural networks.

The GI tract disease classification in endoscopy images usually consists of four significant stages: preprocessing, extraction of the features, selection of solid features,

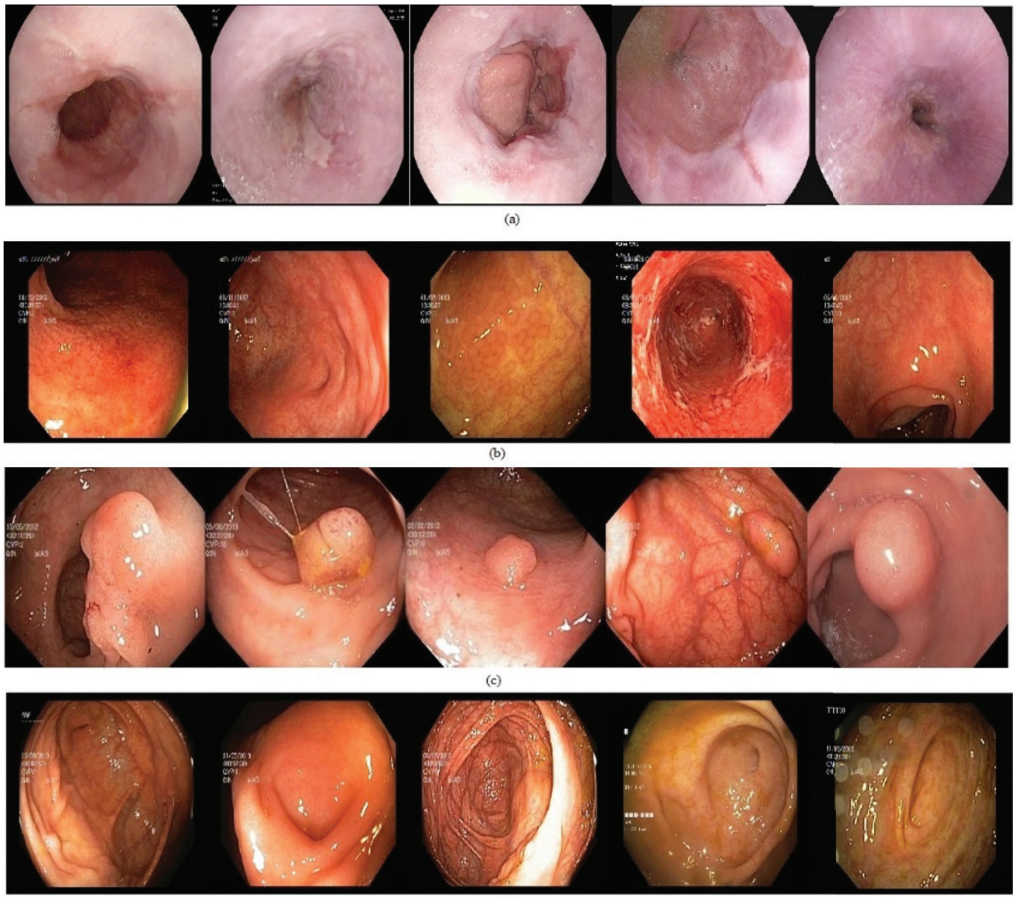

Fig. 1: Abnormalities in GI Tract having different sizes and locations (a) Esophagitis (b) Ulcerative-colitis (c) Polyps (d) Healthy images

fusion, and classification (Amin, J., Sharif, M., Yasmin, M., Saba, T., Anjum, M. A., & Fernandes, S. L., 2019; Nida, N., et al., 2016; Khan, M. A., et al., 2020). Owing to many challenges, this is a complicated and time-consuming technique, and the primary one faces many regions that are not easily visible due to a comprehensive series of videos. The CAD programs can help physicians to overcome these problems. The anomalous area can be a different size and type, making identification difficult. In particular, this area has significant issues, such as a stable and robust selection of the appropriate data (Umer, M. J., & Sharif, M. I., 2022), the spectrum of lesion texture, variance, color, form and scalability, and accuracy results.

A novel paradigm for classifying GI tract diseases is suggested to overcome these problems. The suggested technique consisted of four key steps: (a) preprocessing the image, (b) extracting features, (c) features selection (d) fusion and classification. Below are the key contributions:

(i) The image is scaled into a particular size during the preprocessing stage, and the 3D-median mask is used to enhance contrast to the lesion.

(ii) In the second phase, shape-based features (HOG) and deep features (Xception and ResNet50) are extracted to recognize high-level image endoscopy characteristics that focus on low-image features (color, shape, and texture indicators).

(iii) Principal component analysis (PCA) and Entropy reduce the selected features to provide greater precision following other approaches. The methodology of this study uses the core features to classify various types of images accurately. The reduced features are supplied to the high-value correlation coefficient to specify the stable and robust features.

(iv) The selected solid features and then fused by a sequential procedure and finally classified by MC-SVM.

The most critical task is to extract and select appropriate features. The emphasis of this paper is more on improving the features identified and reducing recognition time.

The background concepts needed for understanding the paper be explained in section 2. In Section 3, the GI tract disease classification methodology is described. Section 4 provides the experiments and their performance with a detailed analysis. The paper ends in section 5 by summarizing the methods suggested.

2. Related Work

Classification (Redwan, S. M., Uddin, M. P., Ulhaq, A., & Sharif, M. I., 2022; Hasan, S. M., et al., 2022; Fayyaz, A. M., et al., 2023; Ramzan, M., Raza, M., Sharif, M. I., & Kadry, S., 2022) and segmentation (Li, J., & Wang, J. Z., 2003) of lesions based on extracted features have received much consideration in computer visualization and graphics for various purposes like medical imaging, classification of scenes, agriculture, and biometrics (Berens, J., Mackiewicz, M., & Bell, D., 2005, April). WCE is the current medical imaging technology that passes through the human body to visualize the entire GI Tract. However, large numbers of video frames make the scheme more complicated and make it harder to explore the diagnosis for physicians (Iddan, G., Meron, G., Glukhovsky, A., & Swain, P., 2000). This chapter summarizes and explores the previous polyp segmentation and features extraction techniques and the classification of GI tract diseases. Identifying and classifying three main types of GI diseases include ulcers, bleeding, and polyps (Amin, J., Sharif, M., Raza, M., & Yasmin, M., 2018). In terms of their characteristics, classification and identification were performed. Figure 2 demonstrates the overall graphical representation of the literature review.

Preprocessing of images requires enhancing local and global contrast between images in the data to enhance the image's visual properties. Based on its many applications, such as monitoring, biometrics, and medical, the preprocessing stage significantly impacts computer vision (Amin, J., Sharif, M., Rehman, A., Raza, M., & Mufti, M. R., 2018; Amin, J., Sharif, M., Raza, M., Saba, T., & Anjum, M. A., 2019;

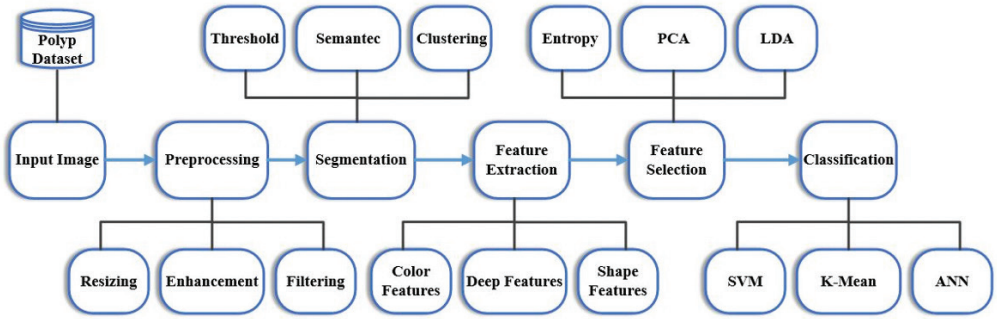

Fig. 2: Graphical Representation of the Literature Review

Nisa, M., et al., 2020). Such approaches as comparison, binary image equalization, and morphological techniques use basic image processing algorithms. Segmentation performs a significant role in extracting the critical region. Segmentation means the image is divided by pixel information into several parts (Khan, M. A., et al., 2022; Khan, M. A., et al., 2020; Sharif, M., Amin, J., Raza, M., Yasmin, M., & Satapathy, S. C., 2020). It is used to detect the infectious disease in a particular image. Feature extraction is the essential step, which is further processed after preprocessing. High precision relies entirely on a good selection of features. The reduction of dimensional space is the same as feature extraction (Sharif, M., et al., 2020; Rashid, M., et al., 2019). Classification is among the most common procedures in computer vision, pattern recognition, and image processing. It determines the feature space of unknown patterns. Classification is generally divided into two main categories: supervised and unsupervised (Amin, J., et al., 2020; Amin, J., et al., 2020; Yuan, Y., Li, B., & Meng, M. Q. H., 2015). For preprocessing, different filters, like the Gaussian filter, guided filter, Gaussian kernel., were used by Yuan et al. (2015), X. et al. (2016), Eskandari et al. (2012, December), and Zou et al. (2016, July) for bleeding images, hookworm, polyp images, and esophagus thus achieved the classification accuracy of 95.75%, 98.4%, 90.91%, and 95.52% respectively. Liaqat et al. (2018) implemented a novel segmentation system for ulcers and bleeding images by separating HSV channels and extracting geometric features of the S channel, thus achieving 98.3% accuracy. Charfi et al. (2018) suggested a new way focused on LBP and DWT for segmenting bleeding and polyp regions. This system achieved an accuracy of 97.0%. Another polyp region was defined by Sanchez-Gonzalez et al. (2019) using shape, color, and curve of margin. The effect is a polyp identification score of 90.53 percent for polyp segmentation. Y. And I, Shin. Balasingham et al. (2018) utilized a dictionary learning method that included a histogram of gradient and hue histogram to distinguish polyp presentation through a 95% classification through SVM for regular versus polyp images. In (Sánchez-González, A., García-Zapirain, B., Sierra-Sosa, D., & Elmaghraby, A., 2018), the researchers used an edge detection system along with color and shapes

region for the polyps segmentation, thus attaining 90.53% accuracy. Wang et al. (2015) implemented Polyp-Alert, which used the previous visual cross-section and rules classifier. It was able to run on non-shelf machines and was used for colonoscopies. It could reach 95.7% accuracy, which means 4.3% false-positive. Mehshan et al. (2020) proposed a new polyp, ulcer, and bleeding detection system by fusing FCN and ResNet50. For classification purposes, Cubic Kernel MSVM attained an accuracy of 99.13%. The researchers proposed another technique using Med-Net (Nguyen, N. Q., Vo, D. M., & Lee, S. W., 2020) for in-depth features to detect polyps. The proposed system through Med-Net features showed 99.0% accuracy. In (Hajabdollahi, M., et al., 2018), the authors provided a convenient and efficient method for the segmentation from WCE images of the bleeding region. The multi-perceptron method used HSI and LAB color transformations, resulting in enhanced results. For detecting polyps, the researchers proposed a two-stage technique (Akbari, M., et al., 2018, July). In the first stage, the image dataset was augmented through patch selection, while in the 2nd step; the processed images were passed through the FCN-8S deep network and thus showed 97.7% accuracy for detecting polyps. Another approach in (Yuan, Y., & Meng, M. Q. H., 2017) extracted deep features using SSAEIM for detecting polyps. Input images were passed through a deep network for feature selection, and at the end, the SVM classifier was used and obtained 98.0% accuracy. Souaidi M., et al. (2019) suggested an ulcer detection algorithm using Cr and Green components of YCbCr and RGB space, respectively. The tests obtained an average accuracy of 95.11 percent by the SVM classification. The authors suggested ResUNet++ (Jha, D., et al., 2019, December) and modern ResUNet architecture for colonoscopy image classification to build a fully automatic paradigm for pixel-specific polyp detection. Experimental investigations found that the proposed architecture provided strong effects on data sets that were publicly accessible.

3. Proposed System

A significant, diverse, and dynamic study in computer visualization is the automated diagnosis of GI tract diseases in endoscopy images in medical imaging. The example of WCE image plays a vital role in detecting GI diseases, including bleeding, ulcer, and polyp. A classification method of infected areas for image processing is introduced in this chapter consisting of four main steps: i) preprocessing the image; ii) features extraction; iii) features selection; iv) reduction of features and classification. The brightness effects are first reduced with methods of preprocessing that generate an efficient image. The improved image is then used for the extraction of features. The optimal value characteristics are chosen from the original features extracted and fed to a multiclass support vector (MC- SVM) for classification, which gives different sets of WCE images, including esophagitis, ulcer, polyp, and healthy. Figure 3 illustrates the proposed method's detailed flow diagram that suggests that each step involves multiple phases.

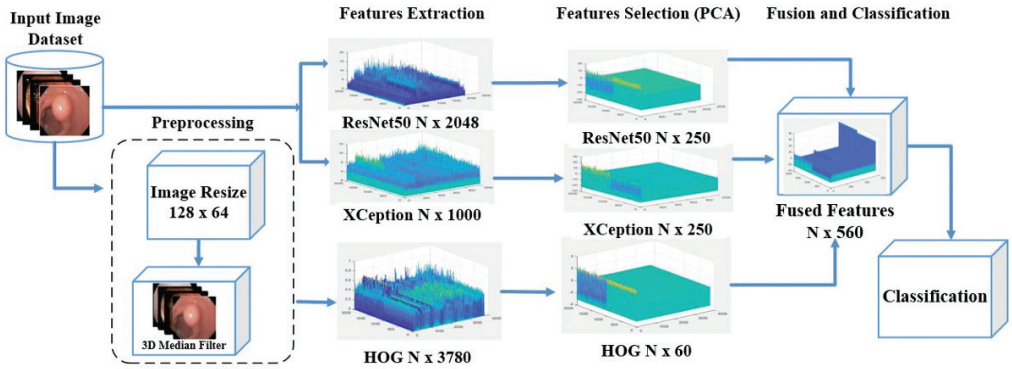

Fig. 3: Proposed framework of GI tract disease classification

3.1. Image Preprocessing

The preprocessing of images ensures that the quality of the reference image increases the object's optical value locally and globally. The preprocessing stage significantly affects computer vision, depending on its numerous uses, including tracking, medical imaging, and biometric identification. It poses many medical issues, such as luminosity variations, illumination characteristics, and the nature of the context. The critical goal for image processing applications is to eliminate harmful material and retrieve the most excellent features. Hence the suggested contrast enhancement approach has been applied to resolve these mentioned problems, which involves two phases: a) Redesign the image to a suitable size for extraction of features; b) 3D-Median filtering. The image is initially resized to a resolution of 128×64. Since the GI tract images contain much chromatic aberration, a 3D-Median filter on the query image is performed to stabilize the image based on the same weights as the pixels joined. Owing to its nonlinearity, the 3D-Median filter best retains the edges. The image also blurs with a moderate pixel value. In the following steps, a 3D-Median filter is implemented. At first, the image is transformed from RGB pixel P(x,y) with three channels, shows a 3D median filter applied to a polyp image. Here is the concept of the median filter:

$$Mask(\tau,\varsigma) = \sum_{\iota=1}^{3}\sum_{\zeta=1}^{3}[\mathcal{R}_{\iota}][\mathcal{C}_{\zeta}] \quad (1)$$

Where $\iota \in \tau$, $\zeta \in \varsigma$, and $\tau, and\ \varsigma$ indicate the rows and columns of the filter created. The generated filter size is 3×3=9, executed on each mode separately as:

$$\gamma_{mask}(\chi,\psi) = Mask(\tau,\varsigma)[\gamma(\chi,\psi)] \quad (2)$$

$$\delta_{mask}(\chi,\psi) = Mask(\tau,\varsigma)[\delta(\chi,\psi)] \quad (3)$$

$$\beta_{mask}(\chi,\psi) = Mask(\tau,\varsigma)[\beta(\chi,\psi)] \quad (4)$$

Where $\gamma_{mask}(\chi,\psi)$ Is the red mask filter, $\delta_{mask}(\chi,\psi)$ is the green mask filter, and $\beta_{mask}(\chi,\psi)$ is the blue mask filter. Moreover, $\gamma(\chi,\psi)$, $\delta(\chi,\psi)$, and $\beta(\chi,\psi)$ signify the red, green, and blue channels comparatively determined.

$$\gamma(\chi,\psi) = \frac{red}{\mathrm{I}(\chi,\psi)} \quad (5)$$

$$\delta(\chi,\psi) = \frac{green}{\mathrm{I}(\chi,\psi)} \quad (6)$$

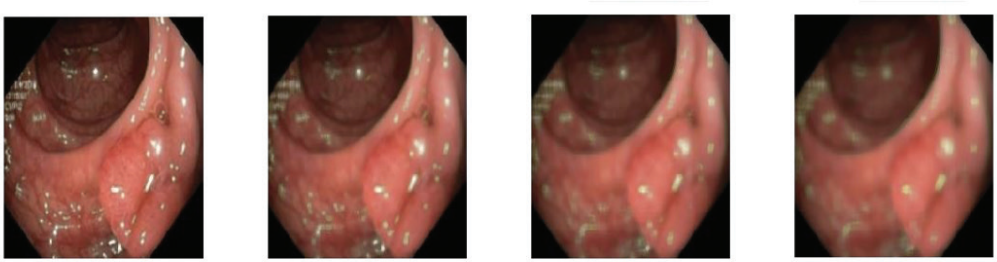

Fig. 4: 3D Median filter on polyp image (a) Original Image (b) 3D Median Filter with 3 Neighboring Pixels (c) 3D Median Filter with 5 Neighboring Pixels (d) 3D Median Filter with7 Neighboring pixels

$$\beta(\chi,\psi) = \frac{blue}{\mathrm{I}(\chi,\psi)} \tag{7}$$

Afterward, the Convolution process is carried out with the mask filter and pixel values for every channel individually. As a result of $\tau = 3$, $\varsigma = 3$, and $\tau_1 = 256$, $\varsigma_1 = 256$, a null matrix $Z(\chi,\psi)$of size 256×256 is generated. A new matrix is produced by utilizing τ, ς, τ_1, and $\varsigma_1$given as under:

$$F_{\varsigma}(\chi,\psi) = \sum_{\iota=1}^{\tau_M} \sum_{\zeta=1}^{\varsigma_\eta} [Z_{\iota,\zeta}(\chi,\psi) + Mask(\tau,\varsigma)] \tag{8}$$

Where $\iota \in \tau_u, \zeta \in \tau_\varsigma, \tau_u = 3, \varsigma_u = 3, \tau_M = \tau_u + \tau_1 - 1$, and$\varsigma_\eta = \varsigma_u + \varsigma_1 - 1$. The τ_u and ς_u symbols signify the modified rows and column pixels, up to 256 times iterated. Finally, $F_{\varsigma}(\chi,\psi)$ has been used, and for each channel, a 2D image is created, gradually coupled to each other, and a new enhanced image is obtained.

$$\mathrm{M}F(\chi,\psi) = \sum_{\iota=1}^{M} \sum_{\zeta=1}^{\eta} \mathrm{I}_{\iota\times\zeta} \tag{9}$$

$$F\eta(\chi,\psi) = \sum_{\iota} \mathrm{M}F_{\iota}(\chi,\psi) \tag{10}$$

Where $F\eta(\chi,)$ shown in (a), while 4 (b), (c), and (d) are the images after 3D median filtering.

3.2. Feature Extraction

Extraction of features is the essential procedure in machine learning (Zafar, M., et al., 2023) and computer vision in monitoring systems, robotics, surgery, farming, and much more. Features are extracted for machine learning to develop a reformation of the input dataset, which preserves its most appropriate data. It would be hard to efficiently classify some classes if we used exclusively traditional features to distinguish medical images. Deep learning has been one of the most important areas of exploration in information science and computer applications in the last several years. Due to advancements in deep learning, several researchers also sought to use such a design framework to work with non-medical images. Various profound models to overcome image difficulties have been suggested. Classification of medical images is one of the main issues in the field of computer vision, and its purpose is to recognize medical images into multiple categories to assist healthcare professionals in the evaluation of disease or further examination. The critical goal of extracting and

reducing features is to increase machine accuracy and running time. Therefore, the most critical features of selected diseases' classification were attempted to be extracted in this section. The feature extraction and deployment of the extracted features are given in detail below.

3.3. Shape Features

the histogram oriented gradients (HOG) correspondingly recognized as shape features are taken from the input frames, To discover the shapes, gradients, and orientation. Combining handcrafted feature vectors improved the modularity and efficiency of deep models. HOG properties are helpful for specific applications as they are essential in various scenarios, such as security, medical imaging, and agri-business-economics. The attributes are interpreted in a row index, where N is the range of the HOG feature. The obtained elements encrypt information on the local shape in an image from regions. Therefore, HOG features are used for the classification of GI tract diseases. The distribution of gradients in the HOG feature collection clarifies the lesion regions. Therefore, 2D gradients D_χ and D_ψ are determined by the mask $[-1, 0, 1]$ and $[-1, 0, 1]^{\mathrm{T}}$ correspondingly.

$$D_\chi = \frac{\partial F_s(\chi,\psi)}{\partial \chi} \tag{11}$$

$$D_\psi = \frac{\partial F_s(\chi,\psi)}{\partial \psi} \tag{12}$$

Then calculate the orientation and weights measures as follows:

$$W_F(\chi,\psi) = \sqrt{{D_\chi}^2 + {D_\psi}^2} \tag{13}$$

$$\theta(\chi,\psi) = tan^{-1}\left(\frac{{D_\psi}^2}{{D_\chi}^2}\right) \tag{14}$$

Where $W_F(\chi,\psi)$ are a weighing parameter, and lesion field orientation is indicated by $\theta(\chi,\psi)$.

3.4. Deep Features using ResNet50 Model

Since CNN has demonstrated practical perception training, extracting features by Convolutional mask filters and redirecting the parameters, we use ResNet50 neural network, pre-trained in the ImageNet dataset (Rezende, E., et al., 2017, December). For high-level image feature extraction, deep features are used. The characteristics are prepared by placing the last fully connected layer to obtain the vector dimension 2048. Without much computing power, these feature vectors are easily obtained. As Resnet50 is able for most images to have an appropriate feature, we did not adjust it to the dataset and instead used the pre-trained weights of the ImageNet dataset. A comprehensive architecture of Resnet50, shown in Figure 6, demonstrates that the model includes one input layer, 4 Convolution blocks containing 3 Convolution layers, 4 Batch-Normalization layers, and 3 ReLU layers. 12 Identity blocks containing 3 Convolution layers, 3 Batch-Normalization, and 3 ReLU layers. In the pooling process, global average pooling is used. A fully-connected layer (FC) and softmax layer are used for classification purposes. The size of the image input layer is 224x224x3 as an RGB image. As an activation function, all hidden layers use the linear activation unit.

CNN consists of a general model with multiple Convolution layers and provides a valuable way to select closely connected features automatically. It includes the input of H×W to the matrix in 3 dimensions. Kernel K is associated with the Convolutional layer h×w×3, while the Convolutional layer threshold is t. There are specified principal formulas for Convolutional layers as

$$\mathrm{X}_{\eta} = \mathrm{X} \times \mathrm{K} = \sum_{\iota=1}^{3} (\chi_{\iota} + \kappa_{\iota}) + \tau \tag{15}$$

$$\mathscr{g}_{\eta} = \frac{\mathscr{g} - \delta + 2 \times \zeta}{D} + 1 \tag{16}$$

$$\mathrm{K}_{\eta} = \frac{\mathrm{K} - \kappa + 2 \times \zeta}{D} + 1 \tag{17}$$

Where K denotes the kernel size, D denotes the number of Convolutions, and τ is a threshold value.

The initialization layer of ReLU is then carried out as follows:

$$\Upsilon = \max(0, \mathrm{X}) \tag{18}$$

This equation shows that all negative values are transformed into zero after applying the function. The term pooling layer is another layer used to condense the number of features derived after the previous layers. Three different types of layers are used primarily as min, average, and max pooling. The strength of pooling is that the most relevant features are collected.

Also, the formula is defined mathematically as follows:

$$\rho_{\max = Max} \mathscr{f}(.\ ,\ .\ ,\ .) \tag{19}$$

$$\rho_{\min = Min} \mathscr{f}(.\ ,\ .\ ,\ .) \tag{20}$$

$$\rho_{avg} = \frac{1}{\eta} \sum \mathscr{f}(.\ ,\ .\ ,\ .) \tag{21}$$

Where $\rho_{\max}$ indicates the features of max-pooling layers, minimum pooling features are defined by $\rho_{\min}$, and average pooling features by ρ_{avg} The final layers of the CNN model, also known as fully connected layers, are described by the formulation below.

$$\mathcal{L}^{\iota\eta} = X \times W + D \tag{22}$$

$$\mathcal{L}^{out} = ReLU(L^{\iota\eta}) \tag{23}$$

The FC layer is defined as follows from these generalized equations:

$$\mathcal{L}_0{}^{out} = X \tag{24}$$

$$\mathcal{L}_{\iota}{}^{\iota\eta} = \mathcal{L}_{\iota-1}{}^{out} \times W + D_{\iota} \tag{25}$$

$$\mathcal{L}_{\iota}{}^{out} = \mathrm{F}_{\iota}(\mathcal{L}_{\iota}{}^{\iota\eta}) \tag{26}$$

Where ι is the number of layers implied, and F indicates the activation function.

Residual (ResNets) networks are highly complex networks where blocks of the Convolution layer can be eliminated using shortcut links to other layers. The simple blocks referred to as 'Bottleneck' blocks are constructed. ResNet50 has 177 layers and 50 fully Convolutional strips divided into 18 segments, except the first and last modules being linear residual links. Down-sampling is carried out directly by the Convolution layer with a 2-step normalization and is carried out immediately after any Convolutions and before ReLU activation. An identity Shortcut is used when the input layers have the exact dimensions as the output layer. Once the dimension rises, the projection shortcut is used for 1-1 Convolution. Both situations are carried out with a stride two if

the shortcuts move across two-size maps. The network accomplishes a fully connected softmax layer (FC).

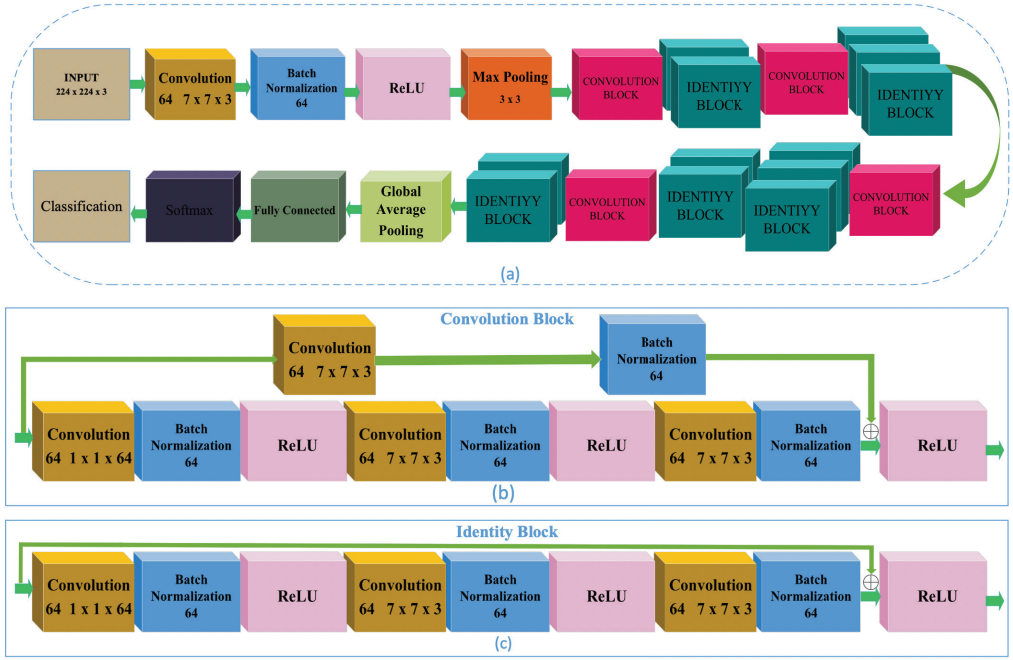

Fig. 5: ResNet 50 (a) Deep CNN Architecture (b) Convolution Block (c) Identity Block

3.5. Deep Features using XCeption Model

Xception is an enhancement of the Inception architecture, introduced by François Chollet (Kaiser, L., Gomez, A. N., & Chollet, F., 2017). This model is a continuous stack of layers of residual links. The thoroughly segregated process focuses on reducing memory and computational cost deserves. Xception has 170 layers and 36 fully Convolutional strips divided into 14 segments, except the first and last modules being linear residual links. Therefore, it allows the description and alteration of the architecture. Down-sampling is carried out directly by the Convolution layer with a 2-step normalization immediately after any Convolutions and before ReLU activation. The independent activation function in Xception distinguishes the analysis of channel and spatial features. A comprehensive architecture of Xception is shown in Figure 7. A fully-connected layer (FC) and softmax layer are used for classification purposes. The size of the image input layer is 299x299x3 as an RGB image. As an activation function, all hidden layers use the linear activation unit.

3.6. Features Selection and Fusion

The selection of the features shortens the feature vector and arranges the best features for the best classification results. Various tests check the elimination of feature vector functions. It re-assigns the chosen characteristics of a feature matrix from the most common features. After extracting shape-based features and in-depth features

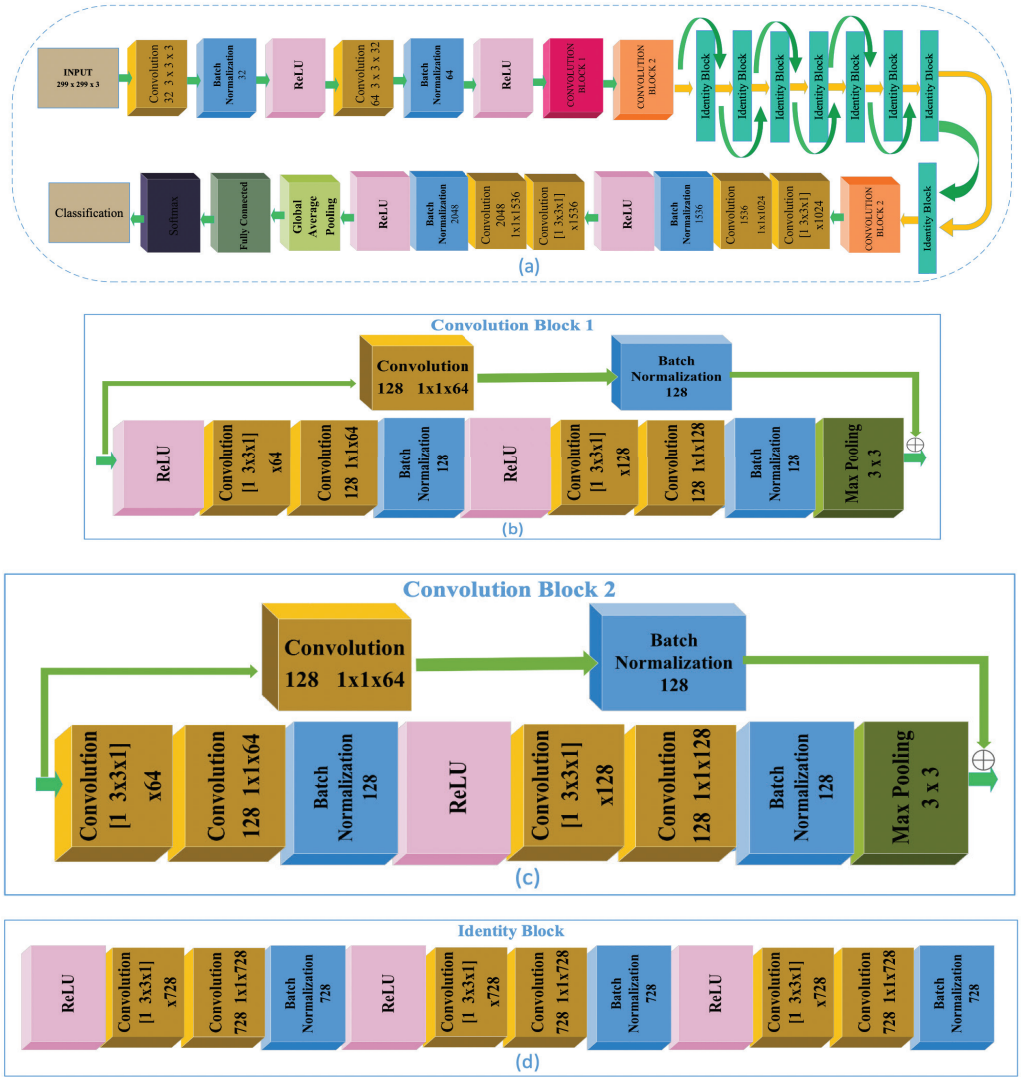

Fig. 6: Xception (a) Deep CNN Architecture (b) Convolution Block 1 (c) Convolution Block 2 (d) Identity Block

using different models, PCA is executed on the feature matrix because it is an unsupervised technique that ignores the class labels to obtain the extracted features' outcomes. Then the average score is measured by defining a selection threshold for higher scores characteristics. The threshold is chosen accordingly:

$$\mathcal{S}(Thr) = \begin{cases} fv(\iota+1) & if\ \ fv(\iota) \geq Mean(score) \\ fv(-1) & if\ fv(\iota) < Mean(score) \end{cases} \quad (27)$$

The function $fv(\iota+1)$indicates the feasible selected features, where the index is

more significant than their mean, and the features not selected are denoted by $fv(-1)$. The correlation is then determined between selected $fv(\iota+1)$features and high-correlation features.

$$\rho\big(fv(\iota)\big) = Corr(\iota,\zeta) = \frac{Cov(\iota,\zeta)}{\sigma\iota\sigma\zeta} \tag{28}$$

$$Cov(\iota,\zeta) = E[(\iota-\mu\iota)(\zeta-\mu\zeta)] \tag{29}$$

Where the mean of rows and columns vector is $\mu\iota$ and $\mu\zeta$, and σ denotes the default standard deviation. Fusion is an essential step in computer vision in which various features are merged in a string to produce the final classified feature vector. This phase is primarily motivated to collect all of the information from multiple descriptors in one vector that could be useful for the minimal error rate. The extracted elements above have been fused using a serial process, and a final vector fused using dimension N×560 has been obtained. Finally, the fused feature vector is fed to the classifier to classify the disease. One versus-one multiclass SVM was used for the chosen features.

4. Experimental Results and Analysis

A detailed analysis and evaluation of the experimental data is presented in this section. The proposed system uses the Kvasir dataset. This dataset contains four different categories: (a) healthy, (b) ulcer, (c) polyp, and (d) esophagitis. The results are collected through six classifications, including Medium Tree, Linear Discriminant Analysis, Kernel Naive Bayes, Linear SVM, Cosine KNN, And Ensemble Subspace KNN, to make a sustainable analysis and comparison of the system proposed. All strategies are evaluated on various features, as outlined in Table 1. Classification measurements like recall, accuracy, precision, FNR, F1 score, number of observations per second, and execution time of each classifier are calculated. All results are computed through 10-fold cross-validation.

4.1. Dataset

This research uses the Kvasir dataset to classify GI tract diseases. Kvasir contains 10,000 images, originally recorded and verified by doctors (professional endoscopists), including a healthy class and eight classes with lesion sites, pathologies, or GI endoscopy processes. To be used for different purposes, for example, image recovery, machine learning, deep learning, and transfer learning. The number of images is adequate. Anatomical features such as Z line, pylorus, and cecum, while esophagitis, polyps, and Crohn's disease-like ulcers are pathological observations[73].

Moreover, two images dataset relevant to the extraction of polyps, "dyed and lifted polyp" and the "dyed resection margins," are also included. The dataset contains images from 720x576 to 1920x1072 pixels with different resolutions and is designed in a manner that is sorted in a different folder with their respective content names. This research classifies four classes: Esophagitis, Healthy, Polyp, and Ulcer. Each class contains 1000 images of their respective class. Polyp images are also obtained from two other datasets. One hundred ninety-six polyp images are obtained from

ETIS-LaribPolypDB [74], while 612 polyp images are obtained from CVC-ColobDB (Hajabdollahi, M., et al., 2018) dataset. Entire images are increased up to 9616 by the data augmentation technique.

4.2. Overview of Experiments

We performed six experiments using distinct criteria for qualitative analysis. The description of all the tests, including the number of features and average accuracy, are listed in Table 1. All the tests are evaluated on the Kvasir dataset having four different classes. Nine thousand six hundred sixteen images are used for classification purposes. Tests 1, 3 & 5 are performed using PCA for feature reduction, while tests 2, 4 & 6 use Entropy for the feature reduction process. Table 1 describes the detailed evaluations of each test.

Table 1 Summary of Test Settings

Experiment	Number of Features			Fused Feature Vector	Feature Reduction	Average Accuracy	Time Consumed
	HOG	Xception	ResNet50				
Test 1	60	250	250	560	PCA	96.6	65.17
Test 2	60	250	250	560	Entropy	95.9	36.53
Test 3	120	500	500	1120	PCA	96.4	150.08
Test 4	120	500	500	1120	Entropy	96.2	58.96
Test 5	180	750	750	1680	PCA	95.9	254.09
Test 6	180	750	750	1680	Entropy	95.2	88.45

4.2.1. Test 1

In test 1, PCA reduces the feature vector's dimensionality. The maximum classification rate for Linear SVM, as shown in Table 2, is 96.60 percent with recall (96.87%) and precision (96.53%), as shown in Table 2. The minimum FNR achieved by the LSVM is 3.4, while the F1 score is 96.54. Figure 8 shows the performance of the Linear SVM analysis by the confusion matrix. The second highest achieved accuracy is 94.1 percent on Cosine KNN using the same feature set. The worst accuracy in Test 1 is 86.5%, achieved by Medium Tree. The minimum time consumed in test case 1 is 21.92sec by Kernal Naïve Bayes, but the performance of Kernal Naïve Bayes in terms of accuracy is deficient. The time consumed by each classifier is shown in Figure 9. According to the above analysis, the proposed methodology is well implemented for the feature vector size N×560 of the function.

Table 2. Test 1 Classification Results

Classifier	Acc (%)	Pre (%)	Rec (%)	FNR (%)	F1 Score (%)	Speed (obs/sec)	Time (sec)

Medium Tree	86.5	86.4	86.6	13.5	86.43	7600.00	23.85
Linear Discriminant Analysis	95.5	95.37	95.33	4.5	95.39	2800.00	28.287
Kernel Naive Bayes	86.9	85.85	87.35	13.2	86.85	4400.00	21.92
Linear SVM	96.6	96.53	96.87	3.4	96.54	4000.00	65.17
Cosine KNN	94.1	94.90	95.37	6.0	94.02	450.00	**39.09**
Ensemble Subspace KNN	91.1	90.45	91.13	8.9	91.13	760.00	105.37

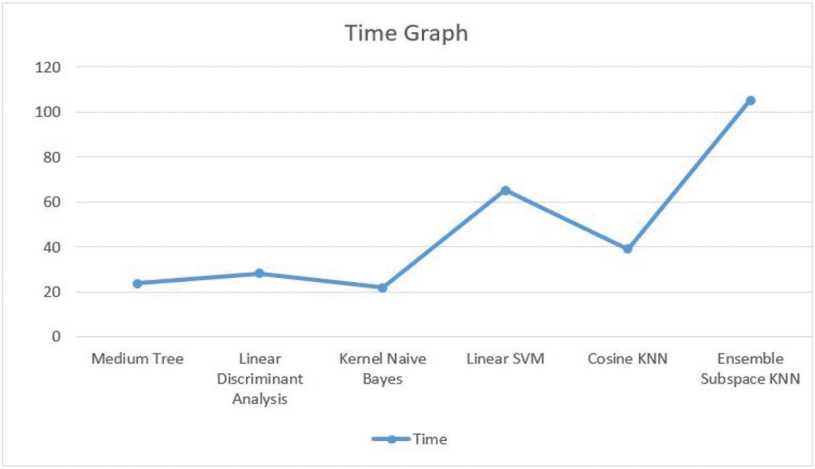

Fig. 7: Test 1 Confusion Matrix of Linear SVM

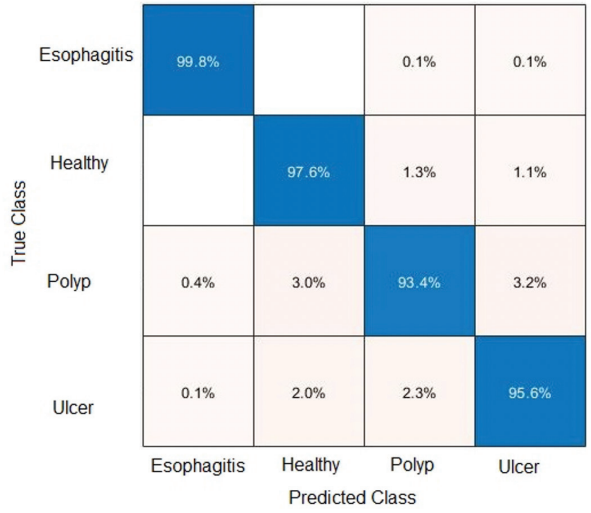

Fig. 8: Time Analysis of Test Case 1

4.2.2. Test 2

Entropy is used to reduce the feature vector's dimensionality in test case 2. The maximum classification rate for Linear SVM, as set out in Table 2, is 95.90 percent with recall (95.95%) and precision (96.0%), as shown in Table 3. The minimum FNR achieved by the LSVM is 4.1, while the F1 score is 95.94. The second highest achieved accuracy is 92.0 percent on Cosine KNN using the same feature set. The worst accuracy in Test 2 is 71.4%, achieved by Medium Tree. The minimum time consumed in test case 2 is 24.10sec by LDA. The time consumed by each classifier is shown in Figure 10. According to the above analysis, the proposed methodology is well implemented for the feature vector size N×560 of the function.

Table 3. Test 2 Classification Results

Classifier	Acc (%)	Pre (%)	Rec (%)	FNR (%)	F1 Score (%)	Speed (obs/ sec)	Time (sec)
Medium Tree	71.4	71.3	71.5	28.6	71.50	5300.00	49.86
Linear Discriminant Analysis	90.1	89.35	90.6	9.9	89.42	3000.00	24.10
Kernel Naive Bayes	90.3	89.33	89.25	9.7	90.33	3900.00	25.78
Linear SVM	95.9	96.0	95.95	4.1	95.94	3600.00	36.53
Cosine KNN	92.0	91.87	92.1	8.0	92.10	450.00	40.63
Ensemble Subspace KNN	88.0	87.95	87.95	12.1	87.95	730.00	109.09

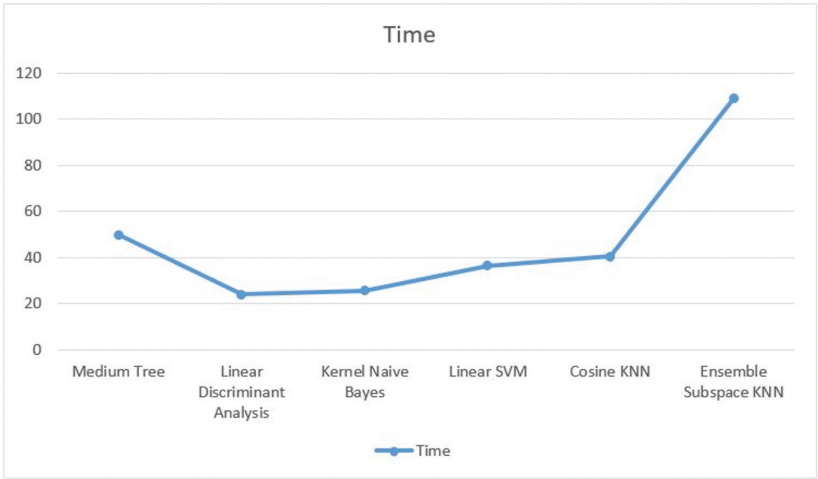

Fig. 9: Time analysis of Test Case 2

4.2.3. Test 3

In test 3, PCA is applied on a feature vector of size N×1120 extracted from the

entire dataset. The maximum classification rate for Linear SVM, as set out in Table 4, is 96.40 percent with recall (96.87%) and precision (96.53%), as shown in Table 4. The minimum FNR achieved by the LSVM is 3.4, while the F1 score is 96.43. The second highest achieved accuracy is 93.8 percent on Cosine KNN using the same feature set. The worst accuracy in Test 1 is 83.9%, achieved by Kernal Naïve Bayes. The minimum time consumed in test case 3 is 24.10sec by LDA. The time consumed by each classifier is shown in Figure 9.

Table 4. Test 3 Classification Results

Classifier	Acc (%)	Pre (%)	Rec (%)	FNR (%)	F1 Score (%)	Speed (obs/ sec)	Time (sec)
Medium Tree	86.5	86.4	86.57	13.5	86.55	2200.00	55.38
Linear Discriminant Analysis	90.1	89.35	90.6	9.9	89.42	3000.00	24.10
Kernel Naive Bayes	83.9	83.25	84.50	16.1	83.93	1500.00	57.27
Linear SVM	96.4	96.3	96.23	3.6	96.43	1900.00	150.08
Cosine KNN	93.8	93.6	93.72	6.2	93.70	230.00	77.92
Ensemble Subspace KNN	92.6	92.36	92.61	7.4	92.60	200.00	292.54

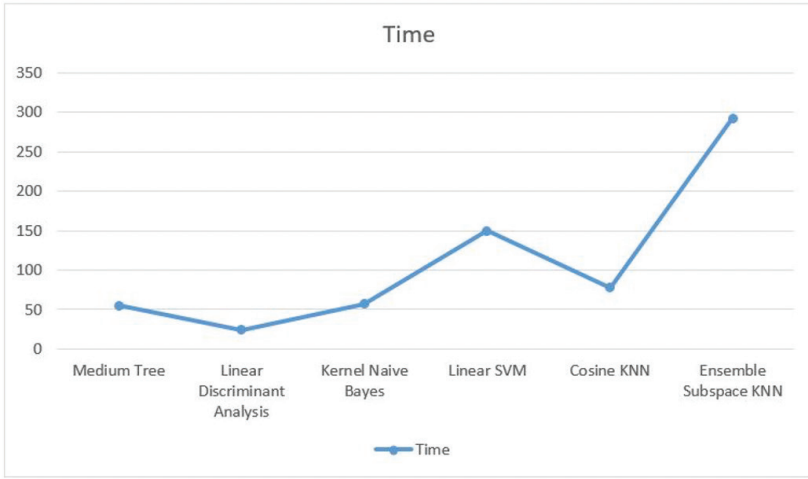

Fig. 10: Time Analysis of Test Case 3

4.2.4. Test 4

In test 4, Entropy is applied on a feature vector of size N×1120. The maximum classification rate for Linear SVM, as set out in Table 4, is 96.20 percent with recall (96.2%) and precision (96.12%), as shown in Table 5. The minimum FNR achieved by the LSVM is 3.8, while the F1 score is 96.2. The second highest achieved accuracy

is 96.1 percent on Ensemble Subspace KNN using the same feature set. The worst accuracy in Test 4 is 80.3%, achieved by Medium Tree. The minimum time consumed in test case 4 is 47.94sec by LDA. The time consumed by each classifier is shown in Figure 9.

Table 5. Test 4 Classification Results

Classifier	Acc (%)	Pre (%)	Rec (%)	FNR (%)	F1 Score (%)	Speed (obs/ sec)	Time (sec)
Medium Tree	80.3	80.21	80.24	19.7	80.36	2500.00	56.38
Linear Discriminant Analysis	95.4	95.4	95.27	4.6	95.35	1100.00	47.94
Kernel Naive Bayes	92.5	91.75	93.42	7.5	92.48	1600.00	55.82
Linear SVM	96.2	96.12	96.2	3.8	96.20	1800.00	58.96
Cosine KNN	92.7	92.7	92.59	7.3	92.78	220.00	78.70
Ensemble Subspace KNN	96.1	95.31	97.23	3.9	96.10	200.00	301.38

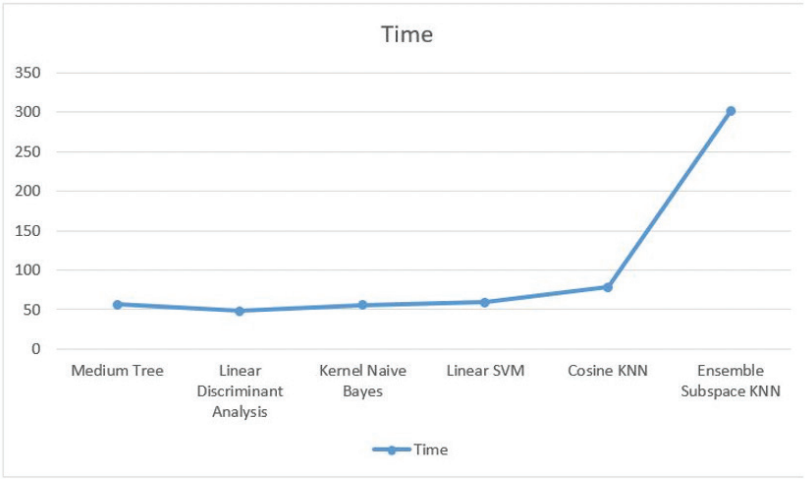

Fig. 11: Time analysis of Test Case 4

4.2.5. Test 5

In test 5, PCA is applied on a feature vector of size N×1680. The maximum classification rate for Linear SVM, as set out in Table 4, is 95.90 percent with recall (95.97%) and precision (95.7%), as shown in Table 6. The minimum FNR achieved by the LSVM is 4.1, while the F1 score is 95.71. The second highest achieved accuracy is 95.2 percent on LDA using the same feature set. The worst accuracy in Test 5 is 86.4%, achieved by Medium Tree. The minimum time consumed in test case 5 is 77.23sec by Medium Tree. The time consumed by each classifier is shown in Figure 10.

Table 6. Test 5 Classification Results

Classifier	Acc (%)	Pre (%)	Rec (%)	FNR (%)	F1 Score (%)	Speed (obs/ sec)	Time (sec)
Medium Tree	86.4	86.3	86.38	13.7	86.42	1600.00	77.23
Linear Discriminant Analysis	95.2	95.34	95.3	4.8	95.10	710.00	99.96
Kernel Naive Bayes	81.0	80.42	81.37	19.0	81.00	1100.00	79.16
Linear SVM	95.9	95.7	95.97	4.1	95.71	1000.00	254.09
Cosine KNN	93.2	93.2	93.31	6.8	93.17	150.00	122.98
Ensemble Subspace KNN	93.5	93.15	93.61	6.5	93.45	110.00	658.09

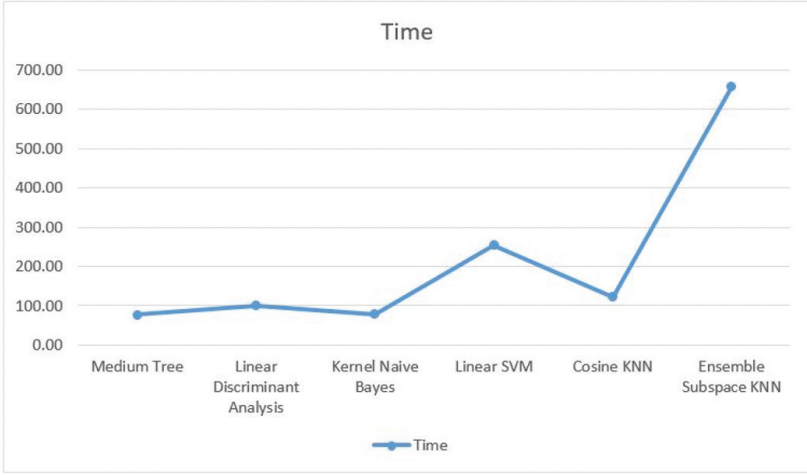

Fig. 12: Time analysis of Test Case 5

4.2.6. Test 6

In test 6, Entropy is applied on a feature vector of size N×1680. The maximum classification rate for Linear SVM, as set out in Table 4, is 95.2 percent with recall (95.09%) and precision (95.14%), as shown in Table 7. The minimum FNR achieved by the LSVM is 4.8, while the F1 score is 95.32. The second highest achieved accuracy is 94.6 percent on LDA using the same feature set. The worst accuracy in Test 6 is 79.9%, achieved by Medium Tree. The minimum time consumed in test case 6 is 88.45sec by LSVM. The time consumed by each classifier is shown in Figure 10.

Table 7. Test 6 Classification Results

Classifier	Acc (%)	Pre (%)	Rec (%)	FNR (%)	F1 Score (%)	Speed (obs/ sec)	Time (sec)

Medium Tree	79.9	79.7	79.72	20.1	79.77	1400.00	92.09
Linear Discriminant Analysis	94.6	94.6	94.69	5.4	94.51	730.00	99.05
Kernel Naive Bayes	92.2	93.28	91.57	7.8	92.20	800.00	95.45
Linear SVM	95.2	95.14	95.09	4.8	95.32	1100.00	88.45
Cosine KNN	92.8	92.8	92.7	7.2	92.87	150.00	119.45
Ensemble Subspace KNN	88.1	87.87	88.53	12.0	88.17	120.00	583.45

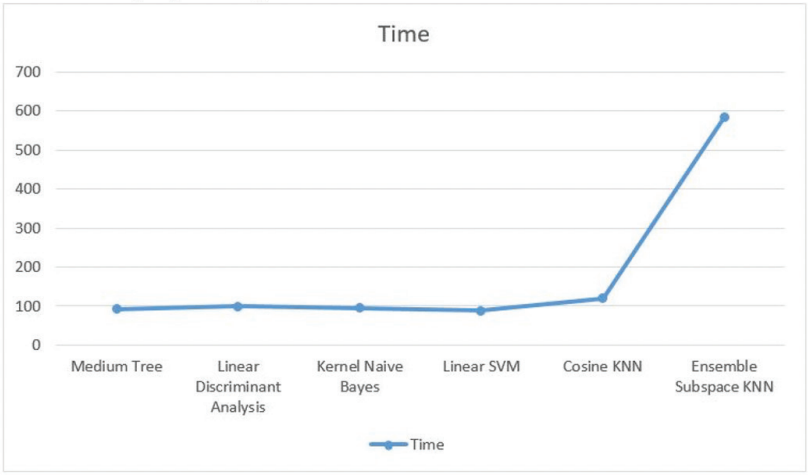

Fig. 13: Time Analysis of Test Case 6

4.3. Comparison with Existing Techniques

We also equated the performance of the proposed algorithm with current GI tract disease classification techniques, as seen in Table 8. The graphical description of contrasts with current GI tract disease classification methods described in Figure 14 shows that the proposed scheme acquires the best recall, accuracy, and precision classification outcomes. All the given outcomes are trained on Kvasir Dataset. 96.6 percent of Kvasir data set accuracy achieved by the proposed approach is better than the existing methods. Asperti et al. achieved 91.5% accuracy by classifying GI tract diseases through data augmentation. Pozdeev et al. attained 88.0% accuracy by binary classifying the disease by fusing global features of the input images. Poudel et al. accomplished 95.7% accuracy by dilation in CNN using ResNet50 architecture.

Table 8. Comparison with Existing Techniques

Ref	Classifier	Year	Accuracy (%)	Recall (%)	Precision (%)
[75]	Ensemble (Asperti et al)	2018	91.5	91.5	91.5

[76]	SVM (Pozdev et al)	2019	88.0	93.0	82.0
[77]	CNN-Res-Net50 (Poudel et al)	2020	95.7	92.2	86.8
Proposed	Linear SVM	2022	96.6	96.87	96.53

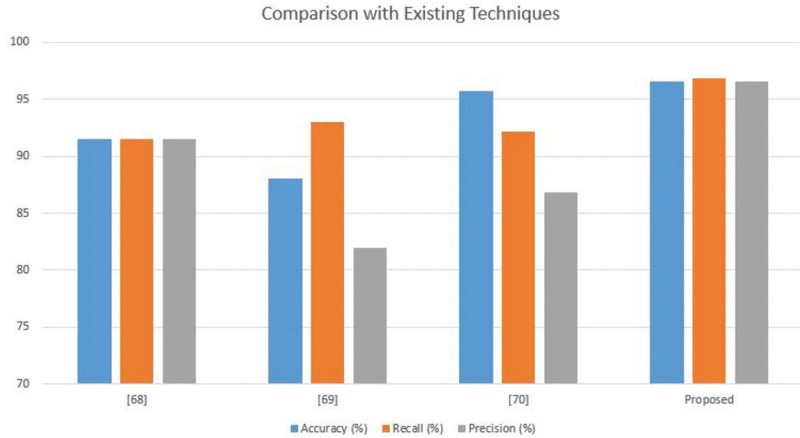

Fig. 14: Comparison with Existing Techniques

4.4. Discussion

This research evaluates the experimental findings in detail using the suggested methodology focused on recall, precision, and accuracy. There are four main steps in the proposed algorithm, with preprocessing as the first step and extraction of features as the second one. The third step is the selection of features, then fusion, and finally, classification. The lesion contrast is enhanced in a preprocessing step with a 3D Median Filter. Furthermore, the shape characteristics of the preprocessed image are obtained. Deep CNN features are also extracted. The features extracted are fused in a single matrix and selected by the PCA and Entropy later. We also used six classifications: Medium Tree, Linear Discriminant Analysis, Linear SVM, Kernel Naive Bayes, Cosine KNN, and Ensemble Subspace KNN. In this approach, we have used four classes: Esophagitis, polyps, ulcers, and healthy, with features vectors: N×560, N×1120, and N×1680; however, the best accuracy is obtained by LSVM having feature vector size N×560. The accuracies are described in Table 2, Table 3, Table 4, Table 5, Table 6, and Table 7: 96.6%, 95.9%,96.4%, 96.2%, 95.9%, and 95.2%, respectively. Figure 9, Figure 10, Figure 11, Figure 12, Figure 13, and Figure 14 shows the time consumed by each classifier graphically.

From the debate above, it is evident that the PCA performs best on minimal characteristic vectors and is more reliable than the high dimensional vectors. Entropy outperforms best by utilizing less time.

Conclusion

In this document, we suggested a system to classify GI tract diseases image composed of four pipeline systems, including preprocessing, extraction, selection, the fusion of features, and classification. We have inferred from the above findings and debate that the classification of GI tract diseases in the WCE dataset is tackled through the recommended Deep CNN characteristics fusion method. This technique can be used for GI tract disease images of WCE. We have also concluded that for classifying target classes like esophagitis, ulcers, polyps, and every day, shape characteristics are also essential. Because of the representation of WCE diseases, shape features are essential, and the extracted features are efficiently handled. Besides, the selection and fusion approaches also remain essential in the context of accuracy and sensitivity to enhance system performance. The analysis findings indicate that the new system's efficiency concerning the current methods exceeded 96%. Results show that combining in-depth features with handcrafted features improves accuracy but does not ultimately enhance the desired efficiency by adding unnecessary features. Selecting the most robust features reduces redundant data, significantly improving identification accuracy. In future research, we will concentrate on more GI diseases and study how to improve the quality of the detection system.

References

Akbari, M., et al. (2018, July). Polyp segmentation in colonoscopy images using fully convolutional network. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 69-72). IEEE.

Ali, H., Yasmin, M., Sharif, M., & Rehmani, M. H. (2018). Computer assisted gastric abnormalities detection using hybrid texture descriptors for chromoendoscopy images*. Computer methods and programs in biomedicine, 157,* 39-47.

Amin, J., et al. (2020). Brain tumor detection: a long short-term memory (LSTM)-based learning model. *Neural Computing and Applications, 32,* 15965-15973.

Amin, J., et al. (2020). Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI. *Cognitive Systems Research, 59,* 304-311.

Amin, J., Sharif, M., Raza, M., & Yasmin, M. (2018). Detection of brain tumor based on features fusion and machine learning. *Journal of Ambient Intelligence and Humanized Computing,* 1-17.

Amin, J., Sharif, M., Raza, M., Saba, T., & Anjum, M. A. (2019). Brain tumor detection using statistical and machine learning method. *Computer methods and programs in biomedicine, 177,* 69-79.

Amin, J., Sharif, M., Rehman, A., Raza, M., & Mufti, M. R. (2018). Diabetic retinopathy detection and classification using hybrid feature set. *Microscopy research and technique, 81*(9), 990-996.

Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. (2018). Big data analysis for brain

tumor detection: Deep convolutional neural networks. *Future Generation Computer Systems, 87,* 290-297.

Amin, J., Sharif, M., Yasmin, M., Saba, T., Anjum, M. A., & Fernandes, S. L. (2019). A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning. *Journal of medical systems, 43,* 1-16.

Asperti, A., & Mastronardo, C. (2017). The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images. *arXiv preprint arXiv:1712.03689.*

Berens, J., Mackiewicz, M., & Bell, D. (2005, April). Stomach, intestine, and colon tissue discriminators for wireless capsule endoscopy images. In *Medical Imaging 2005: Image Processing* (Vol. 5747, pp. 283-290). SPIE.

Berens, J., Mackiewicz, M., & Bell, D. (2005, April). Stomach, intestine, and colon tissue discriminators for wireless capsule endoscopy images. In *Medical Imaging 2005: Image Processing* (Vol. 5747, pp. 283-290). SPIE.

Brandao, P., et al. (2018). Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks. *Journal of Medical Robotics Research, 3*(02), 1840002.

Charfi, S., & Ansari, M. E. (2018). Computer-aided diagnosis system for colon abnormalities detection in wireless capsule endoscopy images. *Multimedia Tools and Applications, 77,* 4047-4064.

Chen, T., et al. (2020). Computer-aided diagnosis of gallbladder polyps based on high resolution ultrasonography. *Computer methods and programs in biomedicine, 185,* 105118.

de Lange, T., Larsen, S., & Aabakken, L. (2005). Image documentation of endoscopic findings in ulcerative colitis: photographs or video clips?. *Gastrointestinal endoscopy, 61*(6), 715-720.

de Souza-Filho, E. M., & de Amorim Fernandes, F. (2021). Deep Learning and Artificial Intelligence in Nuclear Cardiology. *Nuclear Cardiology: Basic and Advanced Concepts in Clinical Practice,* 741-762.

El-Matary, W. (2008). Wireless capsule endoscopy: indications, limitations, and future challenges. *Journal of pediatric gastroenterology and nutrition, 46*(1), 4-12.

Eskandari, H., Talebpour, A., Alizadeh, M., & Soltanian-Zadeh, H. (2012, December). Polyp detection in Wireless Capsule Endoscopy images by using region-based active contour model. In *2012 19th Iranian Conference of Biomedical Engineering (ICBME)* (pp. 305-308). IEEE.

Fayyaz, A. M., et al. (2023). Analysis of Diabetic Retinopathy (DR) Based on the Deep Learning. *Information, 14*(1), 30.

Feng, R., et al. (2020, April). SSN: A stair-shape network for real-time polyp segmentation in colonoscopy images. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (pp. 225-229). IEEE.

Hajabdollahi, M., et al. (2018). Segmentation of bleeding regions in wireless capsule

endoscopy images an approach for inside capsule video summarization. *arXiv preprint arXiv:1802.07788.*

Hasan, S. M., et al. (2022). A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders. *IEEE Access.*

Høiland, T. N. (2017). *Automatic Analysis of Endoscopic Videos* (Master›s thesis).

Hwang, S., Oh, J., Cox, J., Tang, S. J., & Tibbals, H. F. (2006, March). Blood detection in wireless capsule endoscopy using expectation maximization clustering. In *Medical Imaging 2006: Image Processing (Vol. 6144, pp. 577-587). SPIE.*

Iddan, G., Meron, G., Glukhovsky, A., & Swain, P. (2000). Wireless capsule endoscopy. *Nature, 405(6785),* 417-417.

Jha, D., et al. (2019, December). Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)* (pp. 225-2255). IEEE.

Kaiser, L., Gomez, A. N., & Chollet, F. (2017). Depthwise separable convolutions for neural machine translation. *arXiv preprint arXiv:1706.03059.*

Khan, M. A., et al. (2020). An automated system for cucumber leaf diseased spot detection and classification using improved saliency method and deep features selection. *Multimedia Tools and Applications, 79,* 18627-18656.

Khan, M. A., et al. (2020). Gastrointestinal diseases segmentation and classification based on duo-deep architectures. *Pattern Recognition Letters, 131,* 193-204.

Khan, M. A., et al. (2020). Lungs cancer classification from CT images: An integrated design of contrast based classical features fusion and selection. *Pattern Recognition Letters, 129,* 77-85.

Khan, M. A., et al. (2020). StomachNet: Optimal deep learning features fusion for stomach abnormalities classification. *IEEE Access, 8,* 197969-197981.

Khan, M. A., et al. (2022). Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection. *Expert Systems, 39(7),* e12497.

Khan, M. A., Javed, M. Y., Sharif, M., Saba, T., & Rehman, A. (2019, April). Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification. In *2019 international conference on computer and information sciences (ICCIS)* (pp. 1-7). IEEE.

Khan, M. A., Sharif, M., Akram, T., Yasmin, M., & Nayak, R. S. (2019). Stomach deformities recognition using rank-based deep features selection. *Journal of medical systems, 43,* 1-15.

Li, J., & Wang, J. Z. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on pattern analysis and machine intelligence, 25*(9), 1075-1088.

Li, P., et al. (2015). Using circular RNA as a novel type of biomarker in the screening of gastric cancer. *Clinica chimica acta, 444, 132-136.*

Liaqat, A., et al. (2018). Automated ulcer and bleeding classification from WCE

images using multiple features fusion and selection. *Journal of Mechanics in Medicine and Biology, 18*(04), 1850038.

Majid, A., et al. (2020). Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection. *Microscopy research and technique, 83*(5), 562-576.

Mughal, B., Muhammad, N., & Sharif, M. (2019). Adaptive hysteresis thresholding segmentation technique for localizing the breast masses in the curve stitching domain. *International journal of medical informatics, 126,* 26-34.

Mughal, B., Sharif, M., & Muhammad, N. (2017). Bi-model processing for early detection of breast tumor in CAD system. *The European Physical Journal Plus, 132,* 1-14.

Nguyen, N. Q., Vo, D. M., & Lee, S. W. (2020). Contour-aware polyp segmentation in colonoscopy images using detailed upsampling encoder-decoder networks. *IEEE Access, 8,* 99495-99508.

Nida, N., et al. (2016). A framework for automatic colorization of medical imaging. *IIOAB J, 7,* 202-209.

Nisa, M., et al. (2020). Hybrid malware classification method using segmentation-based fractal texture analysis and deep convolution neural network features. *Applied Sciences, 10*(14), 4966.

Pogorelov, K.,et al. (2017, June). Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 164-169).

Poorneshwaran, J. M., Kumar, S. S., Ram, K., Joseph, J., & Sivaprakasam, M. (2019, July). Polyp segmentation using generative adversarial network. In *2019 41St annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 7201-7204). IEEE.

Poudel, S., Kim, Y. J., Vo, D. M., & Lee, S. W. (2020). Colorectal disease classification using efficiently scaled dilation in convolutional neural network. *IEEE Access, 8,* 99227-99238.

Pozdeev, A. A., Obukhova, N. A., & Motyko, A. A. (2019, January). Automatic analysis of endoscopic images for polyps detection and segmentation. In *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (pp. 1216-1220). IEEE.

Ramzan, M., Raza, M., Sharif, M. I., & Kadry, S. (2022). Gastrointestinal Tract Polyp Anomaly Segmentation on Colonoscopy Images Using Graft-U-Net. *Journal of Personalized Medicine, 12*(9), 1459.

Rashid, M., et al. (2019). Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and SIFT point features. *Multimedia Tools and Applications, 78,* 15751-15777.

Redwan, S. M., Uddin, M. P., Ulhaq, A., & Sharif, M. I. (2022). Power Spectral Density-Based Resting-State EEG Classification of First-Episode Psychosis. *arXiv*

preprint arXiv:2301.01588.

Rezende, E., et al. (2017, December). Malicious software classification using transfer learning of resnet-50 deep neural network. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1011-1014). IEEE.

Saba, T., Mohamed, A. S., El-Affendi, M., Amin, J., & Sharif, M. (2020). Brain tumor detection using fusion of hand crafted and deep learning features. *Cognitive Systems Research, 59,* 221-230.

Saba, T., Sameh, A., Khan, F., Shad, S. A., & Sharif, M. (2019). Lung nodule detection based on ensemble of hand crafted and deep features. *Journal of medical systems, 43,* 1-12.

Sánchez-González, A., García-Zapirain, B., Sierra-Sosa, D., & Elmaghraby, A. (2018). Automatized colon polyp segmentation via contour region analysis. *Computers in biology and medicine, 100,* 152-164.

Sánchez-González, A., García-Zapirain, B., Sierra-Sosa, D., & Elmaghraby, A. (2018). Automatized colon polyp segmentation via contour region analysis. *Computers in biology and medicine, 100,* 152-164.

Shahzad, A., Raza, M., Shah, J. H., Sharif, M., & Nayak, R. S. (2022). Categorizing white blood cells by utilizing deep features of proposed 4B-AdditionNet-based CNN network with ant colony optimization. *Complex & Intelligent Systems,* 1-17.

Sharif, M. I., Khan, M. A., Alhussein, M., Aurangzeb, K., & Raza, M. (2021). A decision support system for multimodal brain tumor classification using deep learning. *Complex & Intelligent Systems,* 1-14.

Sharif, M., Amin, J., Raza, M., Yasmin, M., & Satapathy, S. C. (2020). An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor. *Pattern Recognition Letters, 129,* 150-157.

Sharif, M., et al. (2020). Brain tumor detection based on extreme learning. *Neural Computing and Applications, 32,* 15975-15987.

Sharif, M., et al. (2021). Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. *Journal of Experimental & Theoretical Artificial Intelligence, 33*(4), 577-599.

Sharif, M., et al. (2021). Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. *Journal of Experimental & Theoretical Artificial Intelligence, 33*(4), 577-599.

Shin, Y., Qadir, H. A., Aabakken, L., Bergsland, J., & Balasingham, I. (2018). Automatic colon polyp detection using region based deep CNN and post learning approaches. *IEEE Access, 6,* 40950-40962.

Souaidi, M., et al. (2019). Multi-scale completed local binary patterns for ulcer detection in wireless capsule endoscopy images. *Multimedia Tools and Applications, 78,* 13091-13108.

Tajbakhsh, N., Gurudu, S. R., & Liang, J. (2013). A classification-enhanced vote accumulation scheme for detecting colonic polyps. In *Abdominal Imaging. Computation*

and Clinical Applications: 5th International Workshop, Held in Conjunction with MICCAI 2013, Nagoya, Japan, September 22, 2013. Proceedings 5 (pp. 53-62). Springer Berlin Heidelberg.

Tjalma, J. J., et al. (2016). Molecular fluorescence endoscopy targeting vascular endothelial growth factor a for improved colorectal polyp detection. *Journal of Nuclear Medicine, 57*(3), 480-485.

Umer, M. J., & Sharif, M. I. (2022). A Comprehensive Survey on Quantum Machine Learning and Possible Applications. *International Journal of E-Health and Medical Communications (IJEHMC), 13*(5), 1-17.

Wang, Y., et al. (2015). Polyp-alert: Near real-time feedback during colonoscopy. *Computer methods and programs in biomedicine, 120*(3), 164-179.

Wickstrøm, K., Kampffmeyer, M., & Jenssen, R. (2018, September). Uncertainty modeling and interpretability in convolutional neural networks for polyp segmentation. In *2018 ieee 28th international workshop on machine learning for signal processing (mlsp)* (pp. 1-6). IEEE.

Wu, X., et al. (2016). Automatic hookworm detection in wireless capsule endoscopy images*. IEEE transactions on medical imaging, 35*(7), 1741-1752.

Yu, L., Yuen, P. C., & Lai, J. (2012, November). Ulcer detection in wireless capsule endoscopy images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (pp. 45-48). IEEE.

Yuan, Y., & Meng, M. Q. H. (2017). Deep learning for polyp recognition in wireless capsule endoscopy images. *Medical physics, 44*(4), 1379-1389.

Yuan, Y., Li, B., & Meng, M. Q. H. (2015). Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images. *IEEE Transactions on automation science and engineering, 13*(2), 529-535.

Yuan, Y., Wang, J., Li, B., & Meng, M. Q. H. (2015). Saliency based ulcer detection for wireless capsule endoscopy diagnosis. *IEEE transactions on medical imaging, 34*(10), 2046-2057.

Zafar, M., et al. (2023). Skin Lesion Analysis and Cancer Detection Based on Machine/Deep Learning Techniques: A Comprehensive Survey. *Life, 13*(1), 146.

Zou, Y., et al. (2015, July). Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network. In *2015 IEEE International Conference on Digital Signal Processing (DSP)* (pp. 1274-1278). IEEE.

Submitted: 02.02.2023
Accepted: 03.05.2023