Human-centric Computing and Information Sciences

Gastrointestinal Diseases Recognition: A Framework of Deep Neural Network and Improved Moth-Crow Optimization with DCCA Fusion

Muhammad Attique Khan[1], Khan Muhammad[2], Shui-Hua Wang[3], Shtwai Alsubai[4], Adel Binbusayyis[4],
Abdullah Alqahtani[4], Arnab Majumdar[5], Orawit Thinnukool[6,*]

Abstract

Wireless capsule endoscopy (WCE), the most efficient technology, is used in the endoscopic department for the examination of gastrointestinal (GI) diseases such as a poly and ulcer. WCE generates thousands of frames for a single patient's procedure, and the manual examination is time-consuming and exhausting. In the WCE frames, computerized techniques make the manual inspection process easier. Deep learning has been used by researchers to introduce a variety of techniques for the classification of GI diseases. Some of them have concentrated on ulcer and bleeding classification, while others have classified ulcers, polyps, and bleeding. In this paper, we proposed a deep learning and Moth-Crow optimization-based method for GI disease classification. There are a few key steps in the proposed framework. Initially, the contrast of the original images is increased, and three operations based on data augmentations are performed. Then, using transfer learning, two pre-trained deep learning models are fine-tuned and trained on GI disease images. Features are extracted from the middle layers using both fine-tuned deep learning models (average pooling). On both extracted deep feature vectors, a hybrid Crow-Moth optimization algorithm is proposed and applied. The resultant selected feature vectors are later fused using the distance-canonical correlation (D-CCA) approach. For classifying GI diseases, the final fused vector features are classified using machine learning algorithms. The experiments are carried out on three publicly available datasets titled CUI Wah WCE imaging, Kvasir-v1, and Kvasir-v2, providing improved accuracy with less computational time compared with recent techniques.

Keywords

Stomach Cancer, Wireless Capsule Endoscopy, Contrast Enhancement, Deep Learning, Optimization, Features Fusion

*Corresponding Author: Orawit Thinnukool (orawit.t@cmu.ac.th)
[1]Department of Computer Science, HITEC University-Taxila, Taxila, Pakistan
[2]Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Applied Artificial Intelligence, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul, Korea
[3]Department of Mathematics, University of Leicester, Leicester, UK
[4]College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia
[5]Faculty of Engineering, Imperial College London, London, UK
[6]College of Arts, Media, and Technology, Chiang Mai University, Chiang Mai, Thailand

1. Introduction

A lot of research is being carried out by researchers in the field of medical imaging by utilizing computer vision and deep learning (DL) on different types of imaging technology such as magnetic resonance imaging (MRI), computed tomography (CT) [1], dermoscopy, X-ray [2], and capsule endoscopy [3]. A common type of cancer that affects both men and women is colorectal cancer [4]. These types of cancers can be a polyp, bleeding, or ulcer. Stomach disease affects approximately 3.6 million children annually [5]. In the United States, the number of registered colorectal cancer cases since 2015 tallies 1.32 million. The total number of bowl infection cases tallies 1.6 million and 200,000 new cases are added each year. Early-stage diagnosis is very difficult due to a high mortality rate. In 2021, the number of reported stomach cancer cases tallied 149,500, while the numbers of deaths tallied 52,980 [6].

Some infections can be linked to colorectal cancer such as short bowl and hemorrhoids. The diagnosis of these infections can be done by using colonoscopy techniques [7]. The drawback of this method is its high time consumption and a limited number of specialists available [3]. Moreover, this method is not suitable for detecting small bowls due to their complex build. This problem was addressed by introducing a new technique called wireless capsule endoscopy (WCE) [8]. WCE is a widely used method for detecting gastrointestinal (GI) diseases. In this method, a tiny camera with a diameter of 11 mm × 30 mm is used to capture the region of GI. The whole procedure may take more than 2 hours [9]. After capturing, all frames are compressed by utilizing the JPEG technique. In WCE, there is no need for an external wire and patients are asked to swallow the camera to record the video. After that, the radio telemetry method is used to transfer the videos to an external recorder. An array based on eight aerials is attached to the patient and a capsule in GI tract is located. This method can detect diseases and perform a small bowl diagnosis, and so it is popular in hospitals. Approximately 1 million people are successfully treated by this method in the last year [10]. However, this method has a few limitations such as time consumption and a lack of experts. The main concern with this method is the time constraint because it takes a long time for a manual diagnosis. Also, an ulcer in the WCE images is not clearly visible due to the low contrast [11]. Therefore, there is a chance that a physician may miss the region of ulcer during the detection process. Furthermore, there is another problem that occurs during a diagnosis with the naked eyes with regard to similarity of color, texture, and variations of the shape [12].

Therefore, researchers introduced various CV techniques for the diagnosis and classification of medical infections such as stomach cancers [13], skin cancers [14], and brain tumors. These techniques are based on some basic steps such as increasing the contrast and removing the noise from the original image, segmenting infected regions in the image, extracting important features of each image, selecting the best features, and finally classifying them into relevant classes. Contrast enhancement is an important step of a computerized method. The main purpose of this step is to improve the intensity range of an infected region to get better segmentation accuracy and extract relevant features [15]. In the segmentation part, infected regions are detected through several techniques such as saliency based [16], and named a few more [17]. The resultant images of this step are passed to the next step for feature extraction; however, this step has several challenges (i.e., change in shape of infected lesion, similarity in color of healthy and infected parts, presence of an infected region on the border) that reduce the segmentation accuracy. The reduction in segmentation accuracy later produces a misclassification of a disease into the relevant class.

Recently, a deep convolutional neural network (CNN) has shown improved performance for both detection and classification of medical infections [18–20]. A CNN is a form of DL that includes several layers such as convolutional, ReLu, fully connected, and pooling. In a CNN model, the raw images are normally processed for features extraction and classification. Compared to the classical techniques, the CNN-based techniques produce much better and reliable results. Several DL techniques are utilized recently for automated recognition of multiple stomach diseases, as illustrated in Fig. 1 [21]. These

methods utilized the pre-trained CNN models for features extraction that are later employed for features optimization. The pre-trained CNN models are trained through transfer learning (TL) due to the shortage of memory and time.
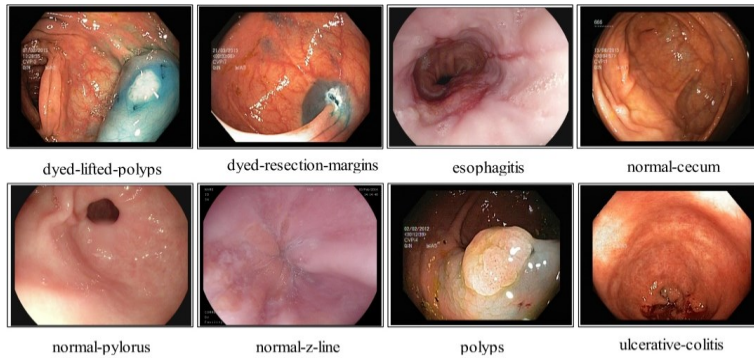
Fig. 1. Sample WCE images of multiple stomach diseases collected from Kvasir-v1 [21].

As an example, Ayyaz et al. [22] presented a hybrid CNN-based approach for stomach infection detection and classification. For feature extraction, they used several CNN pre-trained models, including VGG and AlexNet. Later on, they used a genetic algorithm to select the best features, which were then classified using machine learning methods. By the same token, Lee at al. [23] utilized DL models such as ResNet-50, VGG16, and Inception-V3 for classifying normal and ulcer GI images. In this method, Resnet-50 outperformed the rest of the deep networks. Khan et al. [24] introduced a saliency-based method to segment the GI infections, whereas a DL architecture is used for classification. They used a YIQ color space along with an HSI color space that is later feed to a contour-based approach for segmentation. Suman et al. [25] used several color spaces for feature extraction such as CMYK, LAB, YUV, RGB, XYZ, and HSV for non-ulcer and ulcer detections. These feature vectors are fused by utilizing a cross-correlation method, and a final classification is done by using support vector machine (SVM), attaining an accuracy of 97.89%. Yuan et al. [26] introduced an automated method for detecting an ulcer from the WCE frames. Initially, a saliency approach based on super pixel is implemented to draw an ulcer region boundary. Next, each level texture and color attributes are computed and fused to get the final map of saliency. After that, a saliency max-pooling (SMP) technique is introduced and merged with locality-constrained linear coding (LLC) to attain a recognition rate of 92.65%. Rustam et al. [27] proposed a bleedy image recognizer (BIR) DL architecture for classifying bleeding infected frames. They trained with using BIR and two custom deep models. They used 1,650 WCE images in the evaluation process and achieved an improved accuracy. The primary goal of this study was to perform an automatic analysis of WCE bleeding images. Jain et al. [28] presented an attention-based DL architecture for stomach disease classification and localization from WCE images. Initially, they performed an efficient CNN-based classification of stomach diseases. Later on, they combined Grad-CAM++ and a custom SegNet for the localization of infected regions. The presented method is evaluated on a KID dataset and achieved an improved accuracy. Lan and Ye [29] introduced a combination of an unsupervised DL method for WCE video summarization. They used several networks such as LSTM, autoencoder to name a few more and then performed summarization. The main purpose of this work was to help doctors in the analysis of the entire WCE video. Naz et al. [30] introduced a hybrid sequential framework for classifying stomach diseases from WCE images. They initially performed contrast enhancement through filtering techniques, followed by feature extraction using a hand-engineered method and VGG. Finally, a serial-based fusion is used, while the classification is performed. The main purpose of this work was to improve the current classification accuracy of stomach diseases. Several other techniques were also presented such as graph convolutional based [31], CNN-batch normalization [32], to name a few more [33].

The features extracted from a single deep network extract a single characteristic; however, recent work has shown a decrease in accuracy when dealing with complex images [34]. Furthermore, the features extracted from the pre-trained models contain a significant amount of irrelevant and redundant information [13]. To deal with the redundant information, the researchers used several fusion and feature selection techniques, which thereby improved the accuracy. However, they were still confronted with the issue of computational time and room for improvement in the accuracy. We proposed a new automated end-to-end framework for stomach disease classification using a DL and hybrid moth-crow optimization algorithms in this article. The following are the major contributions of this work:

- A contrast stretching technique based on the maximum intensity value of the infected region and a combination of local-global information are proposed.
- For feature extraction, two pre-trained models, called MobileNet-V2 and NasNet Mobile, are fine-tuned and trained with using TL.
- Instead of deep layers, deep learning features are extracted from average pooling layers and refined, using an entropy-based function. The features are then combined with the help of a serial-based average threshold function.
- An amalgam crow-moth optimization algorithm is proposed based on the cross-entropy loss function for the best DL feature selection.

The remainder of the paper is structured as follows: Section 2 provides a detailed mathematical formulation of the proposed DL and crow-moth flame optimization algorithms. The experimental results are presented in Section 4, which follows the conclusion in Section 5 with our key findings.

2. The Proposed Methodology

The proposed DL and hybrid crow-moth optimization algorithm-based frameworks consist of a few important steps, as illustrated in Fig. 2. The contrast of the original images is increased, and three operation-based data augmentations are performed. After that, two pre-trained DL models are fine-tuned and trained on GI disease images using TL. By employing both fine-tuned DL models, features are extracted from the middle layers (average pooling). A hybrid crow-moth optimization algorithm is proposed and applied on both extracted deep feature vectors that are later fused using the distance-canonical correlation (D-CCA) approach. The final fused vector features are classified using machine learning algorithms for classifying GI diseases.
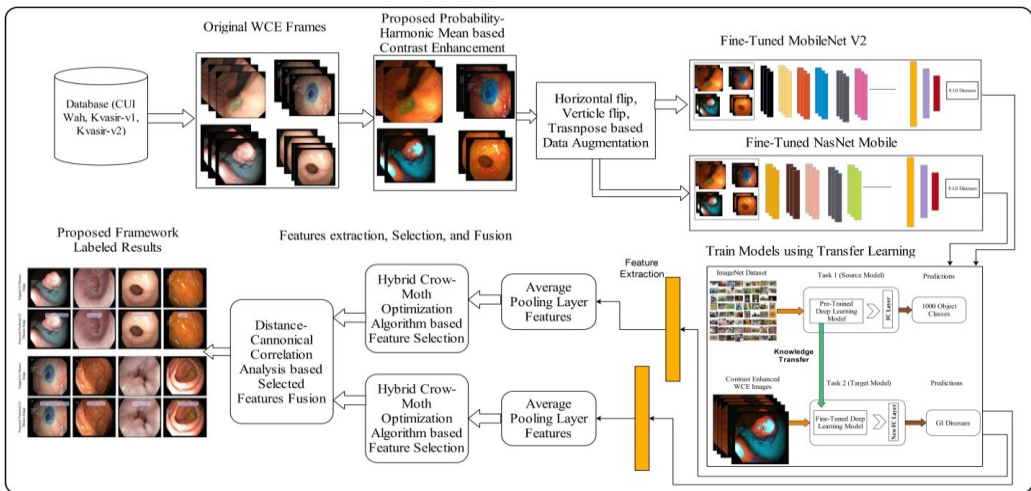
Fig. 2. The proposed framework of GI diseases classification using deep learning and moth-crow-based features optimization.

2.1 Contrast Enhancement and Augmentation

Let $\tilde{\Delta}$ be a database having three datasets of WCE images. Suppose $\phi(i,j)$ is a WCE image of dimensions $M \times N \times K$, where $M = N = 256$ and $K = 3$. The $K = 3$ represents that the nature of input image is RGB. The contrast of each image is computed by the following mathematical formulation expressed as:

$$\phi_{c1}(i,j) = \tilde{B}_t + \tilde{T}_h \tag{1}$$

$$\tilde{B}_t = B_t(\phi(i,j) \cdot S) + Ct \tag{2}$$

$$\tilde{T}_h = T_h(\phi(i,j) \circ S) + Ct \tag{3}$$

$$\tilde{\phi}_c(i,j) = (\phi_{c1}(i,j) - \phi(i,j)) * \tilde{B}_t \tag{4}$$

where, $\phi_{c1}(i,j)$ is initial contrast enhanced resultant image, $\tilde{B}_t$ is bottom hat transformed values, $\tilde{T}_h$ denotes top-hat transformed values, $\tilde{\phi}_c(i,j)$ is updated contrast enhanced values, Ct is constant of value 1, and S is a structuring element value of 11. Two operators are also applied such as opening and closing denoted by $(\circ)$ and $(\cdot)$. The Gaussian function is applied on the updated contrast image $\tilde{\phi}_c(i,j)$ to remove the noisy pixels. Mathematically, it is defined as follows:

$$\mathbb{G} = \frac{1}{2\pi\sigma} e^{\frac{-1}{2}\left(\frac{\tilde{\phi}_c - Mean}{\sigma}\right)} \tag{5}$$

$$Mean = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\tilde{\phi}_c(i,j)\right) \tag{6}$$

$$\sigma = \sqrt{E\left(\tilde{\phi}_c(i,j)\right) - E\left(E\tilde{\phi}_c(i,j)\right)^2} \tag{7}$$

$$\tilde{\phi}_G(i,j) = \mathbb{G}\left(\tilde{\phi}_c(i,j)\right) \tag{8}$$

where, $\mathbb{G}$ denotes the Gaussian function, σ is a standard deviation, and $\tilde{\phi}_G$ is Gaussian updated image. After that, the minimum and maximum intensity values are computed denoted by α from $\tilde{\phi}_G$ and the Gaussian image is updated and expressed as follows:

$$\alpha = \frac{\max(\tilde{\phi}_G(i,j)) + \min(\tilde{\phi}_G(i,j))}{2} \tag{9}$$

$$\tilde{\phi}_{\tilde{G}}(i,j) = \tilde{\phi}_G(i,j) \times \alpha \tag{10}$$

where, $\tilde{\phi}_{\tilde{G}}(i,j)$ is updated Gaussian image. This image is divided into three different channels and histogram of each channel is computed. Based on the histograms, the higher frequency pixel value is computed that is embedded in the threshold functions expressed as follows:

$$Ch_k(x,y) = \tilde{\phi}_{\tilde{G}}(M,N,k), k = 1,2,3 \tag{11}$$

$$Hist_k = \sum_{k=1}^{3} H(Ch_k) \tag{12}$$

where, H is the histogram of each channel. Based on $Hist_k$, three probability values p_1, p_2, and p_3 are obtained and are individually put in the tumor region brightness increasing function expressed as follows:

$$Tr = \begin{cases} Ch_k & for & Ch_k(x,y) \leq p_{\breve{k}}, \\ Not\ Updated & Otherwise \end{cases}, \quad \breve{k} \in (1,2,3) \tag{13}$$

This function describes that if the pixel value of each channel ($\breve{k}$) is less than or equal to the corresponding threshold value, and then the pixel value is updated with that probability value, otherwise the same pixel value will be considered. After that, all three resultant channels are combined and multiplied with a harmonic mean value. The new updated image is obtained as follows:

$$\tilde{\phi}_{cn}(i,j) = \psi(3, Ch_k), k = 1,2,3 \tag{14}$$

$$\tilde{\phi}_{fI}(i,j) = \tilde{\phi}_{cn}(i,j) \times \widehat{HM} \tag{15}$$

where, $\widehat{HM}$ is a harmonic mean value defined as $\widehat{HM} = \frac{N}{\Sigma\left(\frac{1}{\tilde{\phi}_{cn}}\right)}$, $\tilde{\phi}_{cn}$ is a concatenated image of three channels, and $\tilde{\phi}_{fI}(i,j)$ is the final contrast enhanced image. Visually, the resultant image is shown in Fig. 3. After this, three operations (horizontal flip, vertical flip, and rotate 90) are performed two times on all images of each dataset to increase the training images.
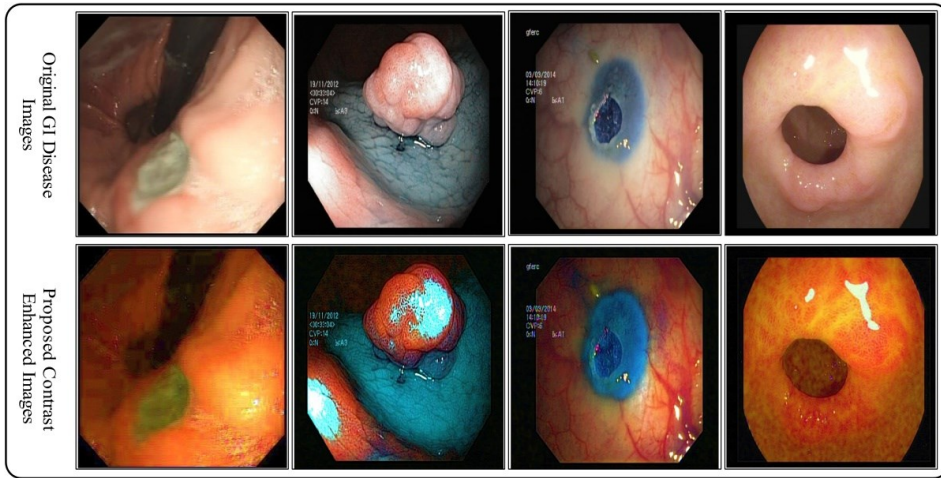
Fig. 3. Proposed tumor contrast enhancement using WCE images.
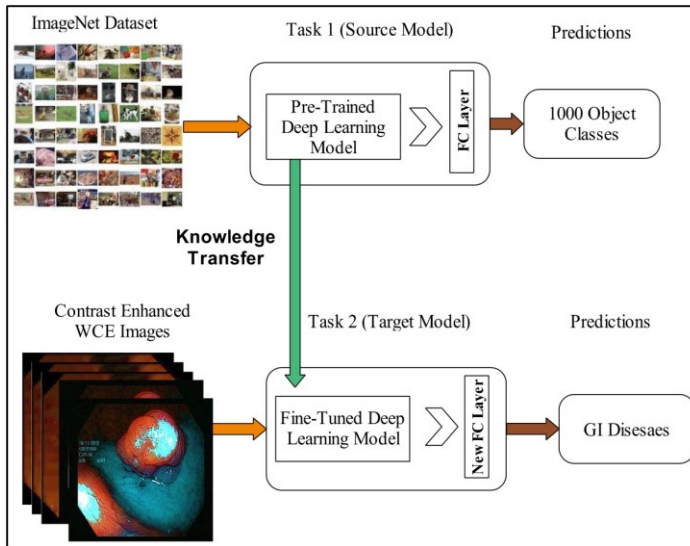
2.2 Transfer Learning

TL is a technique to reuse a pre-trained deep CNN model for another task. As shown in Fig. 4, it is described that the originally pre-trained DL models are trained on an ImageNet dataset having 1,000 object classes. Through knowledge transferring, the fine-tuned models are trained on Stomach dataset without having a maximum 8 GI diseases. This shows that the target data is less than the source data. Hence, the TL main purpose is to re-train a fine-tuned DL model on a smaller dataset. Mathematically, we can define this process shown as follows.

A source domain is provided $S_{dom} = \{(u_1^S, v_1^S), \ldots, (u_i^S, v_i^S), \ldots \ldots (u_n^S, v_n^S)\}$, where $(u_n^S, v_n^S) \in \mathbb{R}$; with a specific learning objective, S_l, and target domain $T_{dom} = \{(u_1^t, v_1^t), \ldots, (u_i^t, v_i^t), \ldots \ldots (u_m^t, v_m^t)\}$, along with learning task $T_{dom}, (u_n^t, v_n^t) \in \mathbb{R}$. The size of the training data is $((m,n)|n \ll m)$, and the labels are Z_1^D and Z_1^T. TL's major role is to enhance the target function T_{dom} learning ability and leveraging the information from the source S_{dom} and target T_{dom}.

2.3 Deep Features Extraction

MobileNet-V2 [35] is a new pre-trained DL light weight model used for classification. Compared to the V1 version [36], the MobileNet-V2 has a better capability to address the problem of gradient vanishing and due to the addition of an inverted residual block and linear bottleneck frame. The new features of MobileNet-V2 are the addition of expansion layer of 1×1 convolution and expending the channels before going to the depth wise convolution operation. This network is known as the "DagNetwork," and accepts an input of dimensions $224 \times 224 \times 3$. This network includes a total of 154 layers and 163×2 connections. Originally, this network is trained on an ImageNet dataset and the output of the last layer has 1,000 object classes. The loss function named cross-entropy is employed for classification. In this work, fine-tuning is performed to reuse this model for feature extraction of GI WCE images. To accomplish this, the last three layers are removed with three new layers added known as New FC, Softmax, and Classification Output. Furthermore, we have only frozen the first 50 layers and retrained the remaining layers on the target dataset using TL. Following the fine-tuning of the model, features are extracted from the global average pooling layer, and the dimension of extracted features on this layer is $N \times 1280$.

NasNet Mobile [37] is a DAGNet CNN architecture, consisting of basic building blocks that are optimized through reinforcement learning. Each cell consists of several layers such as convolutions, pooling, and recurrent as per the size of the network. This network has 5.3 M parameters and a total of 12 cells. The first layer of this network called the "input layer" accepts an input of dimensions $224 \times 224 \times 3$. The total layers in this network are 913 with 1072×2 connections. We replaced the last three layers with the New FC, Softmax, and Classification Output layers during the fine-tuning process. Except for the first input layer, the first 500 layers are frozen, and the rest are retrained on GI WCE datasets. The training is done with TL, along with a learning rate of 0.001, mini batch size of 64, 100 epochs, and an ADAM optimization method. Other hyperparameter values include a dropout factor of 0.5, weight decay of $4e^{-3}$, and norm decay of 0.8. Following the training of a fine-tuned model, features are extracted from the global average pooling layer, yielding a feature vector of dimensions $N \times 1056$.

Fig. 4. Transfer learning based retrained a pre-trained deep learning model for GI disease classification.

2.4 Moth-Crow Features Optimization

In this work, a moth-crow optimization algorithm is utilized and modified through the Renyi entropy (RE) activation function and single layered feed forward neural network. Consider, we have two feature

vectors denoted by $\tilde{\psi}_1$ and $\tilde{\psi}_2$ having dimensions $N \times 1280$ and $N \times 1056$, respectively. The crow search algorithm [38] is applied on each vector separately and an RE activation function is added at the end. Features passed from the activation function are considered for the fitness calculation and process is continued for the initialized iterations. After that, a best crow search best feature vector is obtained and further passed to the moth flame optimization algorithm [39]. Similar to the crow search, RE activation function is selected and checked through fitness function for final features selection.

Consider, C_n crows that means flock size is C_n. At the i_{th} iteration, the current location of crow C may be described as vector $A^{l,m}$.

$$A^{l,m} = \left[A_1^{l,m}, A_2^{l,m}, A_3^{l,m}, \dots \dots, A_n^{l,m} \right] \tag{16}$$

Each crow has its own memory that stores information about where it hides its food. $P_{k,i}$ represents the Crow C food hiding position at the i_{th} iteration. Crow C has discovered the finest place yet. Crows visit and investigate various sites in order to find the better food hiding sites. For this purpose, the position mechanism is selected. Assume that during the i_{th} iteration, crow C must go to its food hiding place $P_{k,i}$. At the same moment (iteration), crow D chooses to pursue crow k in order to get access to crow $C's$ food hiding spot. In this circumstance, two update conditions are selected:

$$x^{D,i+1} = x^{D,i} + D_{rand} \times \mathrm{L}^{D,i} \times (P_{k,i} - x^{D,i}) \tag{17}$$

where, D_{rand} denotes the random numbers between 0 and 1 and $\mathrm{L}^{D,i}$ indicates the crow D at i^{th} iteration's flight length. The $\mathrm{L}^{D,i}$ has a substantial impact on the algorithm's capacity to search. Hence, the lower L values favor local search, whereas larger L values favor global search. The second update equation is defined as follows:

$$x^{D,i+1} = \begin{cases} x^{D,i} + D_{rand} \times \mathrm{L}^{D,i} \\ \times \left(P_{k,i} - x^{D,i} \right) & r_D \geq aprob^{C,i} \\ a\ random\ location & otherwise \end{cases} \tag{18}$$

where r_D donates uniformly distributed random number with range between 0 and 1 and $aprob^{C,i}$ donates the crow D awareness probability at i^{th} iteration. We selected Eq. (18) update criteria and compute the RE value as follows:

$$REv = Et(x^{D,i+1}) \tag{19}$$

$$Et = -\log \sum_{i=1}^{NT} P_i^2 \tag{20}$$

where, REv denotes the RE value, Et is RE function, $x^{D,i+1}$ is input initial crow search feature vector, NT denotes the total number of features, and P is the probability of each value. Based on the REv, an activation function is proposed that selects the features for further processing expressed as follows:

$$At = \begin{cases} \tilde{x}^{D,i+1} & for\ \ x^{D,i+1} \geq REv \\ Remove, & Otherwise \end{cases} \tag{21}$$

The selected features $\tilde{x}^{D,i+1}$ from this function are evaluated through single layered feedforward neural network (SLFFNN) [40] fitness function. The MSER is selected as a loss function of the fitness function. This process is executed 100 times and after that, a possible best feature vector is obtained that is further refined through moth flame algorithm [39]. The moth flame works based on the following three steps of creating the population, updating the positions, and updating the final amount of flame. Initially, the group of moths can be expressed as follows:

$$B = \begin{bmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,z} \\ B_{2,1} & B_{2,2} & \ldots & B_{2,z} \\ \vdots & \vdots & \vdots & \vdots \\ B_{m,1} & B_{m,2} & \cdots & B_{m,z} \end{bmatrix} \tag{22}$$

where, $B \in \tilde{x}$, m denotes the number of moths, and z denotes the number of dimensions. The fitness value of each moth is stored in the following array manner:

$$OB = \begin{bmatrix} OB_1 \\ OB_2 \\ \vdots \\ OB_m \end{bmatrix}, Q = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,z} \\ q_{2,1} & q_{2.2} & \ldots & q_{2,z} \\ q_{3.1} & q_{3.2} & \cdots & q_{3,z} \\ \vdots & \vdots & \vdots & \vdots \\ q_{m,1} & q_{m,2} & \cdots & q_{m,z} \end{bmatrix}, OQ = \begin{bmatrix} OQ_1 \\ OQ_2 \\ \vdots \\ OQ_m \end{bmatrix} \tag{23}$$

After that, the position of the moths is updated to get the global best solutions. To find the global best solution of the optimization challenge, the following function is selected:

$$MFO = (R, S, T) \tag{24}$$

where R denotes the initial moths' random positions defined by $(R: P \rightarrow \{P, OP\})$, S denotes the movement of moths in the search space defined by $(S: P \rightarrow P)$, and T denotes the complete search space defined by $(T: P \rightarrow True, False)$. The P is utilized to implement the random distribution and is defined as follows:

$$P(i, k) = (u(i) - l(k) * r() + lb(i)) \tag{25}$$

$$U(P_i, Q_j) = H_i . e^{a1c} . \cos(2\pi t) + Q_j \tag{26}$$

where l and u are the variable's lower and upper limits, respectively. The gap between the i_{th} moth and the j_{th} flame is referred to as H_i i.e., $(H_i = |Q_j - P_i|)$. The symbol $a1$ is fixed for determining the logarithmic spiral's form, where c is an integer between -1 and 1. The moths update the positions until they get the local optima and the optimal solutions have been retained in each iteration. Finally, the selected flames (features) are obtained by the following equation shown as:

$$F_n^{i+1} = round \left(Max - * \frac{Max - I}{V} \right) \tag{27}$$

where, Max denotes the highest possible amount of flames, I denotes the number of iterations currently in progress, and V denotes the amount of total iterations. The selected flames F_n^{i+1} are passed to the Eqs. (22)–(24) and their fitness is checked through SLFFNN. This above entire process is applied on both deep extracted feature vectors through which two best selected vectors are obtained in the output with dimensions $N \times 752$ and $N \times 660$.

2.5 Distance Canonical Correlation based Fusion

Consider that we have two selected feature vectors A_1 and A_2 having dimensions $N \times 752$ and $N \times 660$, respectively. Initially, CCA [41] based fusion is performed and then the distance among pair of features is computed for the final fusion. CCA looks for transforming the two variates to such transformed variates, except for the maximum relationship across the two features vectors. In the two given feature vectors $A_1 \in \boldsymbol{R}^{i \times m}$ and $A_2 \in \boldsymbol{R}^{i \times n}$, CCA finds the linear combinations **A1U1** and **A2U2** that maximize the feature vectors across the pairwise correlation. The **Z**1 and $Z_2 \in \boldsymbol{R}^{i \times n}$, $c \leq min(\text{rank}(\mathbf{A1}, \mathbf{A2}))$, these are called recognized variates and $U_1 \in \boldsymbol{R}^{m \times c}$ and $U_2 \in \boldsymbol{R}^{n \times c}$ and these are

the canonical coefficients vectors. The first pair of canonical coefficient vectors $u_1^{(1)}$ is found by the deflationary approach method. The linear combinations of two feature vectors are maximized and expressed as follows:

$$max_{u_1^{(1)}, u_2^{(1)}} \; corr \; (A_1 u_1^{(1)}, A_2 \; u_2^{(1)})$$

(28)

Based on this equation, the maximum combination is computed among the features. By considering the rest of the steps in CCA, we skipped all the others and selected the distance formula. Through the distance formula, features are fused based on the minimum distance value shown as follows:

$$Dis = \sqrt{(f_{i+1} - f_i)^2 - \left(f_{j+1} - f_j\right)^2}$$

(29)

Based on this formula, the distance is computed among each feature, and only those feature pairs are considered whose distance is a minimum. The final fused vector is passed to machine learning classifiers for the final classification results. A few labeled results of the proposed framework are illustrated in Fig. 5.
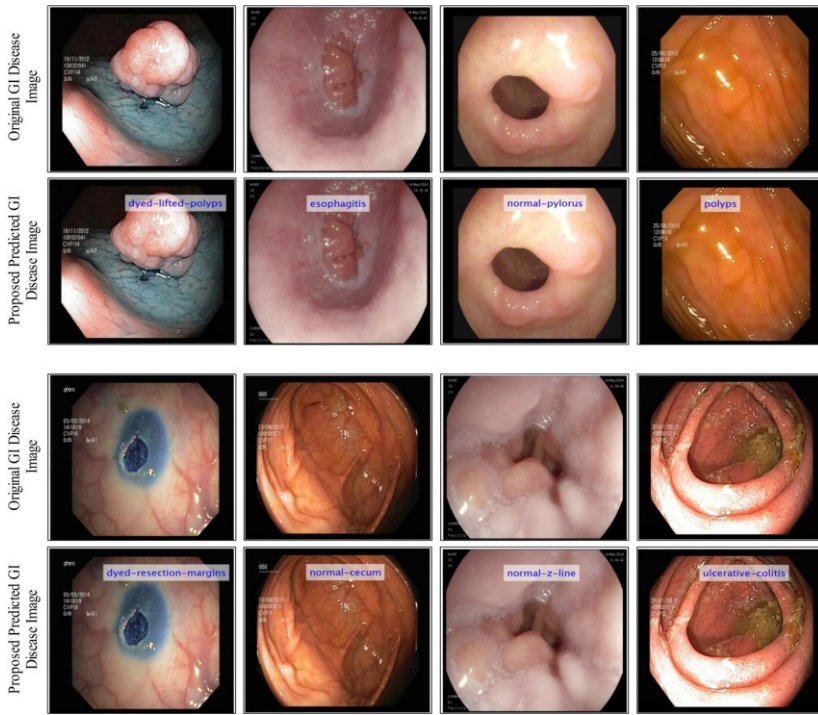
Fig. 5. Sample labeled GI disease prediction results of the proposed framework.

3. Results and Discussion

The proposed framework is evaluated on three publically available WCE images datasets having multiple GI diseases. The selected datasets are CUI Wah WCE images [34], Kvasir-v1, and Kvasir-v2 [21]. The 50% of the images of each dataset are utilized for the training of the models, while 50% is employed for the testing process. In this work, the cross-validation value is set to 10-fold for the entire experimental process. The hype parameters are selected for the training of DL models based on a learning rate of 0.001, mini batch size of 64, 100 epochs, optimization method through ADAM, dropout

factor of 0.5, weight decay of $4e^{-3}$, and norm decay is 0.8. The sigmoid activation function is employed for features extraction, whereas the action loss is a cross-entropy max. Several classification methods have been employed for the evaluation applying a fine tree, weighted k-nearest neighbors (KNN), ensemble baggage tree, and multiclass SVM. The performance of each classifier is computed based on accuracy and computational time. The entire framework is simulated on MATLAB 2021b equipped with a personal computer and 8 GB of graphics card and 32 GB of RAM.

3.1 Results

The proposed framework is evaluated based on the five middle step experiments consisting of (1) features extraction from a fine-tuned MobileNet-V2; (2) features extraction from a fine-tuned NasNet mobile; (3) crow-moth flame modified optimization algorithm applied on MobileNet-V2 features; (4) crow-moth flame modified optimization algorithm applied on NasNet mobile features, and (5) fusion of best features using a D-CCA-based approach. The results are computed for each step using all datasets to analyze their importance in the proposed framework, accordingly.

Table 1. Classification results of the proposed framework on CUI WCE images dataset

Classifier	Feature					Performance measure	
	MobileNet	NasNet Mobile	CMFO-MobileNet	CMFO-NasNet Mobile	Fusion	Accuracy (%)	Time (s)
Fine tree	√					89.64	336.8834
		√				86.52	259.4367
			√			92.93	154.9987
				√		90.64	136.6625
					√	94.15	107.2105
Weighted KNN	√					91.46	362.7745
		√				88.27	310.32267
			√			94.18	194.7725
				√		91.59	180.0024
					√	96.52	145.1178
Ensemble baggage tree	√					91.95	405.1178
		√				87.47	367.9935
			√			94.85	226.7843
				√		90.43	210.5776
					√	96.78	163.0078
MCSVM	√					94.65	276.7734
		√				92.87	251.1730
			√			97.64	142.0783
				√		96.79	128.7723
					√	**99.42**	**103.8901**

The best results are given in bold.

CUI WCE Dataset Results: Table 1 shows the outcomes of the proposed framework on the CUI WCE dataset. This table displays the classification accuracy of each step for the selected classifiers. The fine tree achieved the highest accuracy of 94.15% and shortest execution time of 107.2105 seconds for the D-CCA-based selected features fusion. The weighted KNN achieved the highest accuracy of 96.52% for the D-CCA fusion, with the shortest time of 145.1178 seconds. With a computational time of 163.0078 seconds, the ensemble tree achieved the highest accuracy of 96.78%. For the fusion process, the MCSVM classifier had a higher accuracy of 99.42%. This classifier's best execution time was 103.8901 seconds. Fig. 6 also illustrates the MCSVM (fusion) confusion matrix. According to the

results, the fusion process produces superior results; however, selecting the best features improves classification accuracy when compared to fine-tuned model features. Furthermore, the selection process reduces the computational time that was later improved in the fusion step. Fig. 7 illustrates the overall time of this dataset, which demonstrates the strength of the proposed framework.
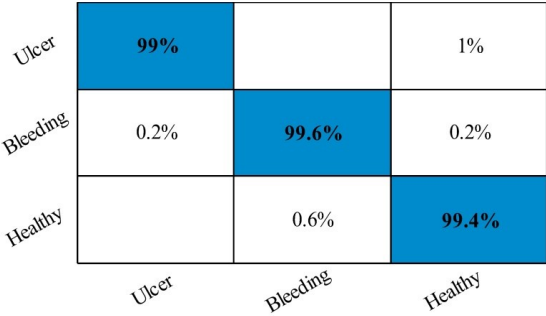

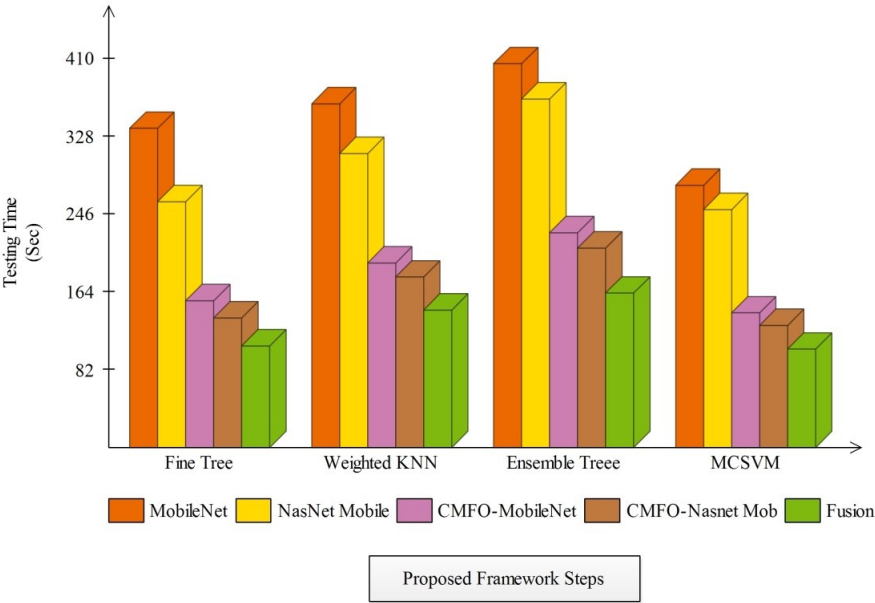**Fig. 6.** Confusion matrix of MCSVM for CUI WCE dataset.

Fig. 7. Comparison of middle steps of the proposed framework based on computational time.

Kvasir-v1 Dataset Results: Table 2 presents the classification results of the proposed framework on the Kvasir-v1 dataset. Similar to the CUI Wah WCE dataset, this dataset was also evaluated on middle steps for each selected classifier. The fine tree achieved the highest accuracy of 92.57% and the minimum testing time of 91.9465 seconds for the D-CCA-based selected features fusion. The weighted KNN classifier obtained the best accuracy of 93.28% for the fusion process with the shortest time of 106.7678 seconds. Performance of the ensemble tree is not improved than the previous classifiers, with this classifier achieving an accuracy of 92.89%. For the D-CCA-based features fusion, the MCSVM classifier had a higher accuracy of 97.85% on this dataset. The shortest time of this classifier was 71.0315 seconds that was smaller than the rest of the classifiers. The confusion matrix of MCSVM is also presented as shown in Fig. 8. This figure illustrated the true prediction rate of each class. According to the results provided in this Table, it is observed that the initial fine-tuned models' accuracy is improved after employing the features selection algorithm. Also, it is noteworthy that the selection process reduced the overall testing time of the proposed framework. The performance of

MCSVM is better than the rest of the classifiers for all middle and fusion steps. Fig. 9 illustrated the time plot of this dataset and shows the strong point of the feature selection and D-CCA fusion.
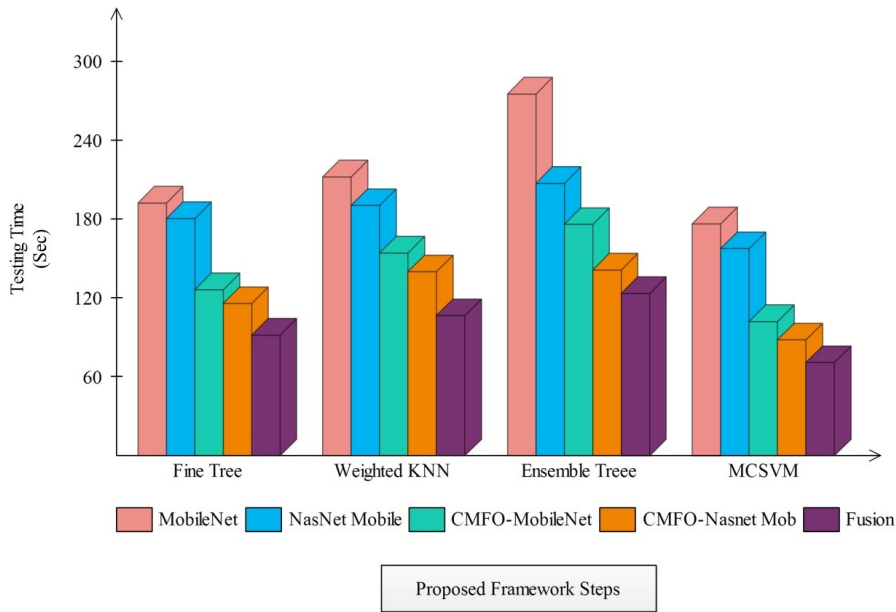
Table 2. Classification results of the proposed framework on Kvasir-v1 dataset

Classifier	Feature					Performance measure	
	MobileNet	NasNet Mobile	HMFO-MobileNet	HMFO-NasNet Mob	Fusion	Accuracy (%)	Time (s)
Fine tree	√					87.43	192.5783
		√				85.29	180.7443
			√			90.75	126.5521
				√		88.49	116.0078
					√	92.57	91.9465
Weighted KNN	√					89.14	212.5976
		√				86.97	190.7654
			√			92.52	154.5895
				√		90.30	140.2574
					√	93.28	106.7678
Ensemble baggage tree	√					88.50	275.7158
		√				85.71	207.5754
			√			90.50	176.3113
				√		89.83	141.6432
					√	92.89	123.7386
MCSVM	√					91.98	176.7832
		√				90.54	157.8594
			√			94.24	102.3899
				√		93.90	88.3243
					√	**97.85**	**71.0315**

The best results are given in bold.
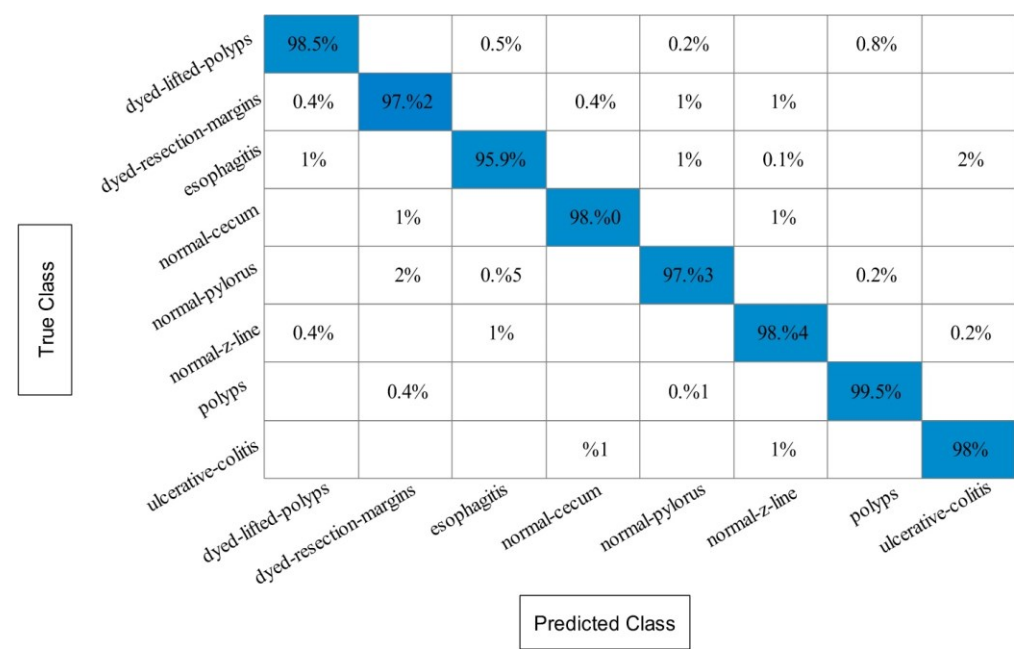
Fig. 8. Confusion matrix of the proposed framework for Kvasir-v1 dataset.

Fig. 9. Comparison of middle steps of the proposed framework based on computational time on Kvasir-v1 dataset.

Table 3. Our framework's classification results on the Kvasir-v2 dataset

Classifier	Feature					Performance measure	
	MobileNet	NasNet Mobile	HMFO-MobileNet	HMFO-NasNet Mob	Fusion	Accuracy (%)	Time (s)
Fine tree	√					87.43	260.6804
		√				85.29	287.9583
			√			90.75	168.8992
				√		88.49	152.7328
					√	92.57	121.0488
Weighted KNN	√					89.14	382.2346
		√				86.97	292.4334
			√			92.52	274.2535
				√		90.30	189.3324
					√	93.28	146.7282
Ensemble baggage tree	√					88.50	402.3328
		√				85.71	309.4543
			√			90.50	236.3453
				√		89.83	190.2324
					√	92.89	151.3453
MCSVM	√					91.98	220.2332
		√				90.54	197.4324
			√			94.24	152.9864
				√		93.90	138.3332
					√	**97.20**	**111.1375**

The best results are given in bold.

Kvasir-v2 Dataset Results: Table 3 shows the outcomes of the proposed framework on the Kvasir-v3 dataset. This classification performance of the selected classifiers is computed for all the middle

steps. The fine tree classifier obtained the highest accuracy of 92.57% for the fusion step. The weighted KNN and ensemble tree classifiers also obtained the best accuracy of 93.28% and 91.98%, respectively, for the D-CCA-based fusion step. For the D-CCA based fusion process, the MCSVM classifier had a higher accuracy of 97.20% that is improved than the rest of the selected classifiers. Fig. 10 also illustrates the confusion matrix of MCSVM. For this dataset, the shortest execution time was 111.1375 seconds for MCSVM, which is also plotted in Fig. 11. According to the results, the accuracy of the proposed framework jumped and computational time decreased after the optimization step. However, the D-CCA fusion step gave a better performance based on both accuracy and time.
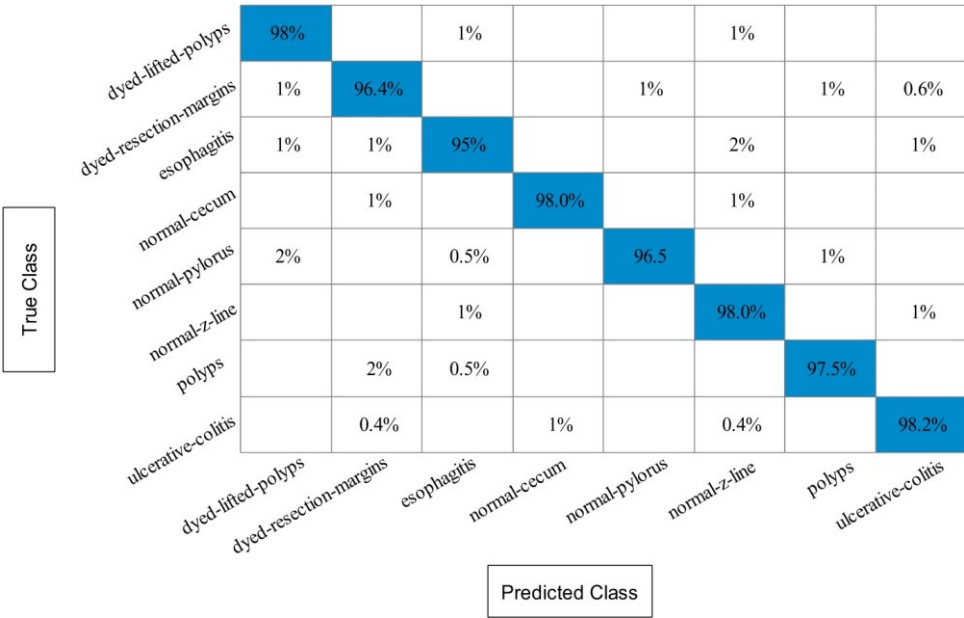

Fig. 10. Confusion matrix of the proposed framework for Kvasir-v2 dataset.
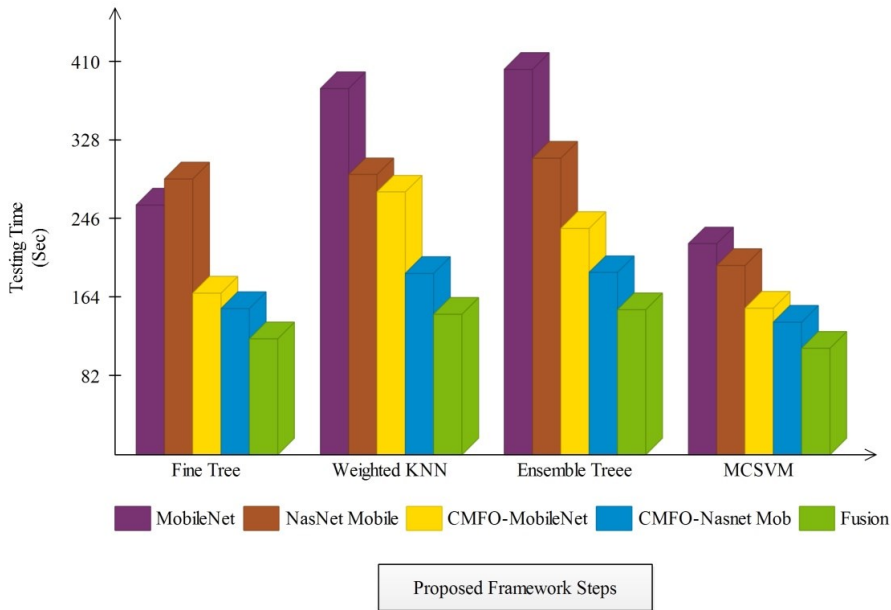

Fig. 11. Comparison of middle steps of the proposed framework based on computational time on Kvasir-v2 dataset.

3.2 Discussion

The detailed analysis of the proposed framework is conducted in this section. The results of each selected dataset are given in Tables 1–3. MCSVM performed better for all three datasets. The confusion matrixes of MCSVM are illustrated in Figs. 6, 8, and 10. The testing times are also plotted in Figs. 7, 9, and 11 for each dataset, respectively, showing the shortest time consumed by the D-CCA fusion step. As shown in Fig. 2, the proposed framework consists of several important steps. The contrast enhancement step effects the classification accuracy of the entire framework. Therefore, we conducted a comparison based on the following two steps comprised of (1) importance of contrast enhancement in the proposed framework based on the accuracy value and (2) difference in the computational time without employing the contrast enhancement step in the proposed framework. Tables 4–6 show the effect of the contrast enhancement step in the proposed classification framework of GI diseases. The values given in these tables show that accuracy of the proposed framework is improved after employing the contrast enhancement technique; however, without contrast enhancement, accuracy is degraded by an average of 6%. On the other hand, the computational time is minimized without employing the contrast enhancement step. Based on this observation, it is clear that the addition of some important steps (i.e., enhancement) increases the accuracy of the system, provided that the system's computational time is affected as well.

Table 4. Comparison of our classification results with or without contrast enhancement step on the CUI WCE dataset

Classifier	Feature					Performance measure	
	MobileNet	NasNet Mobile	HMFO-MobileNet	HMFO-NasNet Mob	Fusion	Accuracy (%)	Time (s)
Contrast	√					94.65	222.8767
enhanced		√				92.87	212.3240
dataset			√			97.64	122.8343
				√		96.79	108.3323
					√	**99.42**	91.0131
Without	√					88.62	201.8579
contrast		√				83.55	191.9343
enhanced			√			89.88	102.8553
dataset				√		85.56	89.0684
					√	92.28	**73.9044**

The best results are given in bold.

Table 5. Comparison of our classification results with or without contrast enhancement step on Kvasir-v1 dataset

Classifier	Feature					Performance measure	
	MobileNet	NasNet Mobile	HMFO-MobileNet	HMFO-NasNet Mob	Fusion	Accuracy (%)	Time (s)
Contrast	√					91.98	176.7832
enhanced		√				90.54	157.8594
dataset			√			94.24	102.3899
				√		93.90	88.3243
					√	**97.85**	71.0315
Without	√					86.73	150.4759
contrast		√				81.87	131.0352
enhanced			√			88.20	92.8050
dataset				√		84.07	73.6894
					√	91.69	**61.1135**

The best results are given in bold.

Table 6. Comparison of our classification results with or without contrast enhancement step on Kvasir-v2 dataset

Classifier	Feature					Performance measure	
	MobileNet	NasNet Mobile	HMFO-MobileNet	HMFO-NasNet Mob	Fusion	Accuracy (%)	Time (s)
Contrast	√					91.98	220.2332
enhanced		√				90.54	197.4324
dataset			√			94.24	152.9864
				√		93.90	138.3332
					√	**97.20**	111.1375
Without	√					85.67	192.9044
contrast		√				79.99	170.2432
enhanced			√			86.01	122.0352
dataset				√		83.78	112.5065
					√	92.96	**98.6849**

The best results are given in bold.

At the end, the proposed framework's accuracy is compared with state-of-the-art (SOTA) techniques, as given in Table 7 [21, 34, 42–47]. In [21], authors used the Kvasir-v2 dataset and achieved an accuracy of 96.33%. Later on, authors in [42] obtained an accuracy of 96.33% on the Kvasir-v2 dataset. The proposed framework obtained an improved accuracy of 97.2% on the Kvasir-v2 dataset. In [43], authors used the Kvasir-v1 dataset and attained an accuracy of 94.46% that was further improved in [44] up to 97%. The proposed framework achieved an improved accuracy of 97.85% on the Kvasir-v1 dataset. Researchers in [34] used the CUI Wah WCE dataset and attained an accuracy of 96.50% that was further improved in [45] up to 98.40%. The proposed framework achieved an accuracy of 99.42% which is better than any SOTA method. Overall, the proposed framework showed an improved accuracy, dominating any SOTA method.

Table 7. Comparison of our results with SOTA techniques

Study	Year	Dataset	Accuracy (%)
Pogorelov et al. [21]	2017	Kvasir-v2	94.20
Gamage et al. [46]	2019	Kvasir-v2	90.74
Majid et al. [34]	2020	CUI Wah WCE	96.50
Khan et al. [45]	2020	CUI Wah WCE	98.40
Kumar et al. [43]	2021	Kvasir-v1	94.46
Yogapriya et al. [42]	2021	Kvasir-v2	96.33
Al-Adhaileh et al. [44]	2021	Kvasir-v1	97.00
Ahmed [47]	2022	Kvasir-v1	90.17
Proposed		CUI Wah WCE	**99.42**
		Kvasir-v1	**97.85**
		Kvasir-v2	**97.20**

The best results are given in bold.

4. Conclusion

A deep learning and moth-crow modified optimization algorithm-based frameworks are proposed for GI disease classification. The contrast enhancement-based data augmentation step is employed for better training of the fine-tuned models. Features are extracted from global average pooling layers that are optimized through the moth-crow modified optimization algorithm. The best selected features are

fused using the D-CCA-based fusion technique. The final fused vector is then passed on to machine learning classifiers for classification. The experimental process is conducted on three datasets of CUI Wah WCE, Kvasir-v1, and Kvasir-v2, resulting in improved accuracy through our framework. Based on the results, we arrived at the following conclusions:

- The proposed contrast enhancement technique shows an improvement in the classification accuracy.
- The learning of fine-tuned models from the middle layers takes some time during the training process but do return better results later.
- The proposed moth-crow modified optimization algorithm improves the classification accuracy and reduces the testing time due to a fewer number of predictors.
- The D-CCA fusion method removes the recurrent features and improves the overall accuracy.

The limitation of this work is the addition of a contrast enhancement step that increases the computational time. In the future, the "EfficientNet" pre-trained model will be considered for feature extraction. Furthermore, the issue of irrelevant and redundant features will be resolved by employing a search point algorithm based on the Newton Raphson. Moreover, our framework will be fine-tuned to work smoothly with the Internet of medical things and blockchain-assisted systems [48, 49].

Author's Contributions

Conceptualization, MAK, KM, SHW. Methodology, MAK, KM, OT, SHW. Software, MAK, KM, AM, OT, SHW. Validation, SA, AB, AA. Formal analysis, SA, AB, AA. Investigation, SA, AB, AA. Resources, SA, AB, AA. Data curation, SA, AB, OT, AA. Writing—original draft preparation, MAK, KM, SHW. Writing—review and editing, AM, OT, SHW. Visualization, AM, OT. Supervision, KM, AM. Project administration, AM, MAK. Funding acquisition, SHW, AM, OT. All authors have read and agreed to the published version of the manuscript.

Funding

Competing Interests

The authors declare that they have no competing interests.

References

[1] M. A. Khan, N. Hussain, A. Majid, M. Alhaisoni, S. A. C. Bukhari, S. Kadry, Y. Nam, and Y. D. Zhang, "Classification of positive COVID-19 CT scans using deep learning," *Computers, Materials, & Continua*, vol. 66, no. 3, pp. 2923-2938, 2021.

[2] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *Applied Intelligence*, vol. 51, no. 2, pp. 854-864, 2021.

[3] M. A. Khan, M. S. Sarfraz, M. Alhaisoni, A. A. Albesher, S. Wang, and I. Ashraf, "StomachNet: optimal deep learning features fusion for stomach abnormalities classification," *IEEE Access*, vol. 8, pp. 197969-197981, 2020.

[4] L. H. Biller and D. Schrag, "Diagnosis and treatment of metastatic colorectal cancer: a review," *JAMA*, vol. 325, no. 7, pp. 669-685, 2021.

[5] M. A. Khan, A. Majid, N. Hussain, M. Alhaisoni, Y. D. Zhang, S. Kadry, and Y. Nam, "Multiclass stomach diseases classification using deep learning features optimization," *Computers, Materials & Continua*, vol. 67, no. 3, pp. 3381-3399, 2021.

[6] P. H. Viale, "The American Cancer Society's facts & figures: 2020 edition," *Journal of the Advanced Practitioner in Oncology*, vol. 11, no. 2, pp. 135-136, 2020.

[7] Y. Hazewinkel and E. Dekker, "Colonoscopy: basic principles and novel techniques," *Nature Reviews Gastroenterology & Hepatology*, vol. 8, no. 10, pp. 554-564, 2011.

[8] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, no. 6785, pp. 417-417, 2000.

[9] A. G. Ionescu, A. D. Glodeanu, M. Ionescu, S. I. Zaharie, A. M. Ciurea, A. L. Golli, N. Mavritsakis, D. L. Popa, and C. C. Vere, "Clinical impact of wireless capsule endoscopy for small bowel investigation," *Experimental and Therapeutic Medicine*, vol. 23, no. 4, article no. 262, 2022. https://doi.org/10.3892/etm.2022.11188

[10] A. Mir, V. Q. Nguyen, Y. Soliman, and D. Sorrentino, "Wireless capsule endoscopy for diagnosis and management of post-operative recurrence of Crohn's disease," *Life*, vol. 11, no. 7, article no. 602, 2021. https://doi.org/10.3390/life11070602

[11] P. Muruganantham and S. M. Balakrishnan, "Attention aware deep learning model for wireless capsule endoscopy lesion classification and localization," *Journal of Medical and Biological Engineering*, vol. 42, pp. 157-168, 2022.

[12] M. A. Khan, M. A. Khan, F. Ahmed, M. Mittal, L. M. Goyal, D. J. Hemanth, and S. C. Satapathy, "Gastrointestinal diseases segmentation and classification based on duo-deep architectures," *Pattern Recognition Letters*, vol. 131, pp. 193-204, 2020.

[13] M. A. Khan, M. Rashid, M. Sharif, K. Javed, and T. Akram, "Classification of gastrointestinal diseases of stomach from WCE using improved saliency-based method and discriminant features selection," *Multimedia Tools and Applications*, vol. 78, no. 19, pp. 27743-27770, 2019.

[14] M. A. Khan, M. Sharif, T. Akram, R. Damasevicius, and R. Maskeliūnas, "Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization," *Diagnostics*, vol. 11, no. 5, article no. 811, 2021. https://doi.org/10.3390/diagnostics11050811

[15] Z. Amiri, H. Hassanpour, and A. Beghdadi, "Feature extraction for abnormality detection in capsule endoscopy images," *Biomedical Signal Processing and Control*, vol. 71, article no. 103219, 2022. https://doi.org/10.1016/j.bspc.2021.103219

[16] M. A. Khan, T. Akram, M. Sharif, K. Javed, M. Rashid, S. A. C. Bukhari, "An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection," *Neural Computing and Applications*, vol. 32, no. 20, pp. 15929-15948, 2020.

[17] Y. Yao, S. Gou, R. Tian, X. Zhang, and S. He, "Automated classification and segmentation in colorectal images based on self-paced transfer network," *BioMed Research International*, vol. 2021, article no. 6683931, 2021. https://doi.org/10.1155/2021/6683931

[18] M. A. Khan, I. Ashraf, M. Alhaisoni, R. Damasevicius, R. Scherer, A. Rehman, and S. A. C. Bukhari, "Multimodal brain tumor classification using deep learning and robust feature selection: a machine learning application for radiologists," *Diagnostics*, vol. 10, no. 8, article no. 565, 2020. https://doi.org/10.3390/diagnostics10080565

[19] G. Cicceri, F. De Vita, D. Bruneo, G. Merlino, and A. Puliafito, "A deep learning approach for pressure ulcer prevention using wearable computing," *Human-centric Computing and Information Sciences*, vol. 10, article no. 5, 2020. https://doi.org/10.1186/s13673-020-0211-8

[20] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, and T. Sato, "Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 54-61, 2021.

[21] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, et al., "Kvasir: a multi-class image dataset for computer aided gastrointestinal disease detection," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, Taipei, Taiwan, 2017, pp. 164-169.

[22] M. S. Ayyaz, M. I. U. Lali, M. Hussain, H. T. Rauf, B. Alouffi, H. Alyami, and S. Wasti, "Hybrid deep learning model for endoscopic lesion detection and classification using endoscopy videos," *Diagnostics*, vol. 12, no. 1, article no. 43, 2021. https://doi.org/10.3390/diagnostics12010043

[23] J. H. Lee, Y. J. Kim, Y. W. Kim, S. Park, Y. I. Choi, Y. J. Kim, D. K. Park, K. G. Kim, and J. W. Chung, "Spotting malignancies from gastric endoscopic images using deep learning," *Surgical Endoscopy*, vol. 33, no. 11, pp. 3790-3797, 2019.

[24] M. A. Khan, M. I. U. Lali, M. Sharif, K. Javed, K. Aurangzeb, S. I. Haider, A. S. Altamrah, and T. Akram, "An optimized method for segmentation and classification of apple diseases based on strong correlation and genetic algorithm based feature selection," *IEEE Access*, vol. 7, pp. 46261-46277, 2019.

[25] S. Suman, F. A. Hussin, A. S. Malik, S. H. Ho, I. Hilmi, A. H. R. Leow, and K. L. Goh, "Feature selection and classification of ulcerated lesions using statistical analysis for WCE images," *Applied Sciences*, vol. 7, no. 10, article no. 1097, 2017. https://doi.org/10.3390/app7101097

[26] Y. Yuan, J. Wang, B. Li, and M. Q. H. Meng, "Saliency based ulcer detection for wireless capsule endoscopy diagnosis," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 2046-2057, 2015.

[27] F. Rustam, M. A. Siddique, H. U. R. Siddiqui, S. Ullah, A. Mehmood, I. Ashraf, and G. S. Choi, "Wireless capsule endoscopy bleeding images classification using CNN based model," *IEEE Access*, vol. 9, pp. 33675-33688, 2021.

[28] S. Jain, A. Seal, A. Ojha, A. Yazidi, J. Bures, I. Tacheci, and O. Krejcar, "A deep CNN model for anomaly detection and localization in wireless capsule endoscopy images," *Computers in Biology and Medicine*, vol. 137, article no. 104789, 2021. https://doi.org/10.1016/j.compbiomed.2021.104789

[29] L. Lan and C. Ye, "Recurrent generative adversarial networks for unsupervised WCE video summarization," *Knowledge-Based Systems*, vol. 222, article no. 106971, 2021. https://doi.org/10.1016/j.knosys.2021.106971

[30] J. Naz, M. Sharif, M. Raza, J. H. Shah, M. Yasmin, S. Kadry, and S. Vimal, "Recognizing gastrointestinal malignancies on WCE and CCE images by an ensemble of deep and handcrafted features with entropy and PCA based features optimization," *Neural Processing Letters*, 2021. https://doi.org/10.1007/s11063-021-10481-2

[31] S. Adewole, P. Fernandes, J. Jablonski, A. Copland, M. Porter, S. Syed, and D. Brown, "Graph convolutional neural network for weakly supervised abnormality localization in long capsule endoscopy videos," 2021 [Online]. Available: https://arxiv.org/abs/2110.09110.

[32] D. Ezzat, H. M. Afify, M. H. N. Taha, and A. E. Hassanien, "Convolutional neural network with batch normalization for classification of endoscopic gastrointestinal diseases," in *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*. Cham, Switzerland: Springer, 2021, pp. 113-128.

[33] M. R. A. Fujiyoshi, H. Inoue, Y. Fujiyoshi, Y. Nishikawa, A. Toshimori, Y. Shimamura, M. Tanabe, H. Ikeda, and M. Onimaru, "Endoscopic classifications of early gastric cancer: a literature review," *Cancers*, vol. 14, no. 1, article no. 100, 2022. https://doi.org/10.3390/cancers14010100

[34] A. Majid, M. A. Khan, M. Yasmin, A. Rehman, A. Yousafzai, and U. Tariq, "Classification of stomach infections: a paradigm of convolutional neural network along with classical features fusion and selection," *Microscopy Research and Technique*, vol. 83, no. 5, pp. 562-576, 2020.

[35] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 4510-4520.

[36] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: efficient convolutional neural networks for mobile vision applications," 2017 [Online]. Available: https://arxiv.org/abs/1704.04861.

[37] M. Cakmak and M. E. Tenekecı, "Melanoma detection from dermoscopy images using Nasnet mobile with transfer learning," in *Proceedings of 2021 29th Signal Processing and Communications Applications Conference (SIU)*, Istanbul, Turkey, 2021, pp. 1-4.

[38] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm," *Computers & Structures*, vol. 169, pp. 1-12, 2016.

[39] S. Mirjalili, "Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm," *Knowledge-Based Systems*, vol. 89, pp. 228-249, 2015.

[40] P. N. Suganthan and R. Katuwal, "On the origins of randomization-based feedforward neural networks," *Applied Soft Computing*, vol. 105, article no. 107239, 2021. https://doi.org/10.1016/j.asoc.2021.107239

[41] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: an overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639-2664, 2004.

[42] J. Yogapriya, V. Chandran, M. G. Sumithra, P. Anitha, P. Jenopaul, and C. Suresh Gnana Dhas, "Gastrointestinal tract disease classification from wireless endoscopy images using pretrained deep learning model," *Computational and Mathematical Methods in Medicine*, vol. 2021, article no. 5940433, 2021. https://doi.org/10.1155/2021/5940433

[43] C. Kumar and D. M. N. Mubarak, "Classification of early stages of esophageal cancer using transfer learning," *IRBM*, 2021. https://doi.org/10.1016/j.irbm.2021.10.003

[44] M. H. Al-Adhaileh, E. M. Senan, F. W. Alsaade, T. H. H. Aldhyani, N. Alsharif, A. A. Alqarni, et al., "Deep learning algorithms for detection and classification of gastrointestinal diseases," *Complexity*, vol. 2021, article no. 6170416, 2021. https://doi.org/10.1155/2021/6170416

[45] M. A. Khan, S. Kadry, M. Alhaisoni, Y. Nam, Y. Zhang, V. Rajinikanth, and M. S. Sarfraz, "Computer-aided gastrointestinal diseases analysis from wireless capsule endoscopy: a framework of best features selection," *IEEE Access*, vol. 8, pp. 132850-132859, 2020.

[46] C. Gamage, I. Wijesinghe, C. Chitraranjan, and I. Perera, "GI-Net: anomalies classification in gastrointestinal tract through endoscopic imagery with deep learning," in *Proceedings of 2019 Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, Sri Lanka, 2019, pp. 66-71.

[47] A. Ahmed, "Classification of gastrointestinal images based on transfer learning and denoising convolutional neural networks," in *Proceedings of International Conference on Data Science and Applications*. Singapore: Springer, 2022, pp. 631-639.

[48] L. Yang, K. Yu, S. X. Yang, C. Chakraborty, Y. Lu, and T. Guo, "An intelligent trust cloud management method for secure clustering in 5G enabled Internet of Medical Things," *IEEE Transactions on Industrial Informatics*, 2021. https://doi.org/10.1109/TII.2021.3128954

[49] L. Tan, K. Yu, N. Shi, C. Yang, W. Wei, and H. Lu, "Towards secure and privacy-preserving data sharing for COVID-19 medical records: a blockchain-empowered approach," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 271-281, 2022.