IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 14, 2020, accepted October 23, 2020, date of publication October 27, 2020, date of current version November 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3034217

StomachNet: Optimal Deep Learning Features Fusion for Stomach Abnormalities Classification

MUHAMMAD ATTIQUE KHAN[1], MUHAMMAD SHAHZAD SARFRAZ[2], (Member, IEEE), MAJED ALHAISONI[3], ABDULAZIZ A. ALBESHER[4], (Member, IEEE), SHUIHUA WANG[5], (Senior Member, IEEE), AND IMRAN ASHRAF[1]

[1]Department of Computer Science, HITEC University, Taxila 47080, Pakistan
[2]Department of Computer Science, National University of Computer and Emerging Sciences at Chiniot–Faisalabad Campus, Chiniot 35400, Pakistan
[3]College of Computer Science and Engineering, University of Ha'il, Ha'il 50141, Saudi Arabia
[4]College of Computing and Informatics, Saudi Electronic University, Riyadh 11673, Saudi Arabia
[5]Department of Mathematics, University of Leicester, Leicester LE1 7RH, U.K.

Corresponding authors: Imran Ashraf (imran.ashraf@hitecuni.edu.pk) and Abdulaziz A. Albesher (a.albesher@seu.edu.sa)

ABSTRACT A fully automated design is proposed in this work employing optimal deep learning features for classifying gastrointestinal infections. Here, three prominent infections – ulcer, bleeding, polyp and a healthy class are considered as class labels. In the initial stage, the contrast is improved by fusing bi-directional histogram equalization with top-hat filtering output. The resultant fusion images are then passed to ResNet101 pre-trained model and trained once again using deep transfer learning. However, there are challenges involved in extracting deep learning features including impertinent information and redundancy. To mitigate this problem, we took advantage of two metaheuristic algorithms – Enhanced Crow Search and Differential Evolution. These algorithms are implemented in parallel to obtain optimal feature vectors. Following this, a maximum correlation-based fusion approach is applied to fuse optimal vectors from the previous step to obtain an enhanced vector. This final vector is given as input to Extreme Learning Machine (ELM) classifier for final classification. The proposed method is evaluated on a combined database. It accomplished an accuracy of 99.46%, which shows significant improvement over preceding techniques and other neural network architectures.

INDEX TERMS Stomach infections, contrast stretching, deep learning, optimization, fusion.

I. INTRODUCTION

Deep learning showed much interest in the area of medical imaging for key areas like dermoscopy, Magnetic Resonance Imaging (MRI) [1], Computed Tomography (CT) [2], [3], and capsule endoscopy [4], [5]. Colorectal cancer is more common in both men and women. The common colorectal cancers are ulcer, polyp, and bleeding. Through stomach infections around 3.6 million children are affected each year [6], [7]. In the USA since 2015, approximately 132,000 new cases of colorectal cancer are registered [8]. Among these cases, 1.6 million people are facing bowel infection. In each year approximately 200,000 new cases appear. Due to the high mortality rate, it is crucial to perform diagnosis at an early stage to save human life [9].

Recently, Wireless Capsule Endoscopy (WCE) technology [10] has become a popular choice as it helps doctors see inside your small intestine - an area that isn't easily reached with more-traditional endoscopy procedures. In this technique, the patient swallows the capsule and the camera used in capsule endoscopy takes thousands of color photos as it passes through the digestive tract. The images are saved on the recorders which are transferred to a computer with special software that strings the images together to create a video. This video is analyzed by an experienced gastrologist in an offline setup to look for abnormalities within the digestive tract, requiring 2-3 hrs of manual effort. This treatment approach has an obvious drawback of the high cost of the diagnosis, due to the involvement of experienced gastrologist, as well as being time-consuming and tedious [11], therefore, automated approaches are proposed by researchers [12].

Automated treatment based on Computer Vision (CV) has been utilized by many researchers to automatically detect various protruding lesions in WCE images [13]–[15]. This method typically starts with a pre-processing step which is an important step in achieving high accuracy of the process [16], [17]. Images are enhanced in this step and then forwarded to the next stage of image segmentation [15], [18].

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

Various techniques have been proposed in the literature for image segmentation [19]. Some of the well-known techniques are K-means, Uniform approach, Normal Distribution, Saliency, and name a few more saliency-based techniques [20]. The segmented images from this step are used for feature extraction, in the next step. Recently, deep learning-based techniques have shown promising results in the field of CV. These techniques have also been successfully utilized in medical imaging [21]. The deep learning-based approach typically requires a considerable amount of data and high computational power to train a model. To tackle this problem, researchers have come up with the technique of Transfer Learning (TL) [22]. In this technique a pre-trained general model is modified by customizing some layers, typically, input and output layers to tune it to the specific problem at hand. Various pre-trained deep convolutional neural network (DCNN) models are introduced by several researchers in the CV community among which famous ones are AlexNet [23], ResNet [24], VGG [25], GoogleNet [26], and YOLO [27].

Various existing challenges in segmentation and classification phases are still not properly addressed. To address these challenges, we propose a new technique for gastrointestinal disease classification using optimal deep learning features fusion. Our technique is comprised of sevral steps; however, the main focus of this work is the optimization of deep learning features and later fusion in one matrix. In summary, our work makes the following contributions:

- Fusion of Bi-Directional Histogram Equalization (HE) and Top-Hat filtering outputs for contrast enhancement of original RGB WCE images.
- Training of the deep learning model- ResNet101 using deep transfer learning.
- Selection of the optimal deep features using two algorithms- differential evaluation (DE) and Enhanced Crow Search Algorithm (ECSA).
- Maximum correlation-based fusion of both optimal vectors to pass them to Extreme Learning Machine (ELM) for final classification.
- Comparison of the proposed method with other neural nets and existing relevant techniques.

This article is organized in the following sequence: Section 2 discusses the previous relevant techniques. Proposed work, which includes contrast enhancement step, selection of optimal solution, and fusion is presented in Section 3. Section 4 presents detailed classification results and finally conclusion of this paper is discussed in Section 5.

II. RELATED WORK

An early work on an automated system to process WCE images is presented in [28]. Authors used color features to detect bleeding and ulcers in WCE images. They employed texture information and combined all the features in one matrix. This matrix was passed to multiple classifiers like neural network, Support Vector Machine (SVM), and decision trees. The evaluation of this approach is calculated using various measures and achieved a performance of 92.86% and 93.64%, respectively. They try to resolve the issues of low contrast and change in the shape of the lesion. Authors in [19] tried to address these challenges by proposing an automated method for classification of WCE images. In this article, a novel method for automated detection and classification of stomach infection is implemented. By utilizing the proposed scheme, authors managed to get a maximum accuracy of 98.3%. In [29], authors presented a deep learning based approach for gastrointestinal diseases (ulcer, polyp, bleeding) classification. The utilized the transfer learning to fine-tune an ResNet101 pre-trained network to extract the features. Later, they applied a grasshopper approach to optimize the features, where min distance was used as a fitness function. The multi-SVM is implemented and classified features for final classification. In the experimental process, they achieved an accuracy of 99.13%, which outperformed the existing techniques.

In [17], Sharif *et al.* presented a fusion approach for classification of stomach diseases. Initially, they applied contrast enhancement approach and then two types of features were extracted, namely; geometric and deep learning features. The conditional entropy approach was considered for the fusion of both types of features. The K-Nearest Neighbor (KNN) classifier was used as a key method and achieved an accuracy of 99.42%. Work in [30] examines the performance of GoogleNet and AlexNet in terms of their ability to recognize Ulcers. These networks were trained on WCE images to detect and classify Ulcers. Authors validated their work on a dataset of 256 WCE images using a variety of performance criteria. Results show that both architectures were able to classify Ulcers with 100% accuracy. A Computer Aided Diagnosis (CAD) system has been proposed in [19]. After enhancing the input images, authors apply saliency map-based segmentation to highlight Ulcerous regions. A recognition scheme is proposed based on Hidden Markov Model (HMM) to obtain promising experimental results on two different datasets. A new CAD system is proposed in [31] which is comprised of HSI color transformation, automatic active contour segmentation, a novel saliency-based method in YIQ color space, fusion of extracted features, calculated using singular value decomposition (SVD), local binary patterns (LBP), and gray level co-occurrences matrices (GLCM), prior to final classification step. Authors evaluated the performance of their system on a dataset containing 9000 samples of Ulcer, bleeding and healthy images, showing an improved performance for the proposed CAD system.

A unified computer-aided scheme is developed in [32] for detecting multiple GI diseases from WCE images. Authors propose Least Square Saliency Transformation (LSST) and probabilistic model fitting scheme using a minimum number of pixel-level annotated images of different diseases.

Fitted PDF parameters are used as features in the proposed hierarchical classification scheme. Publicly available WCE images have been used by authors for the performance evaluation.

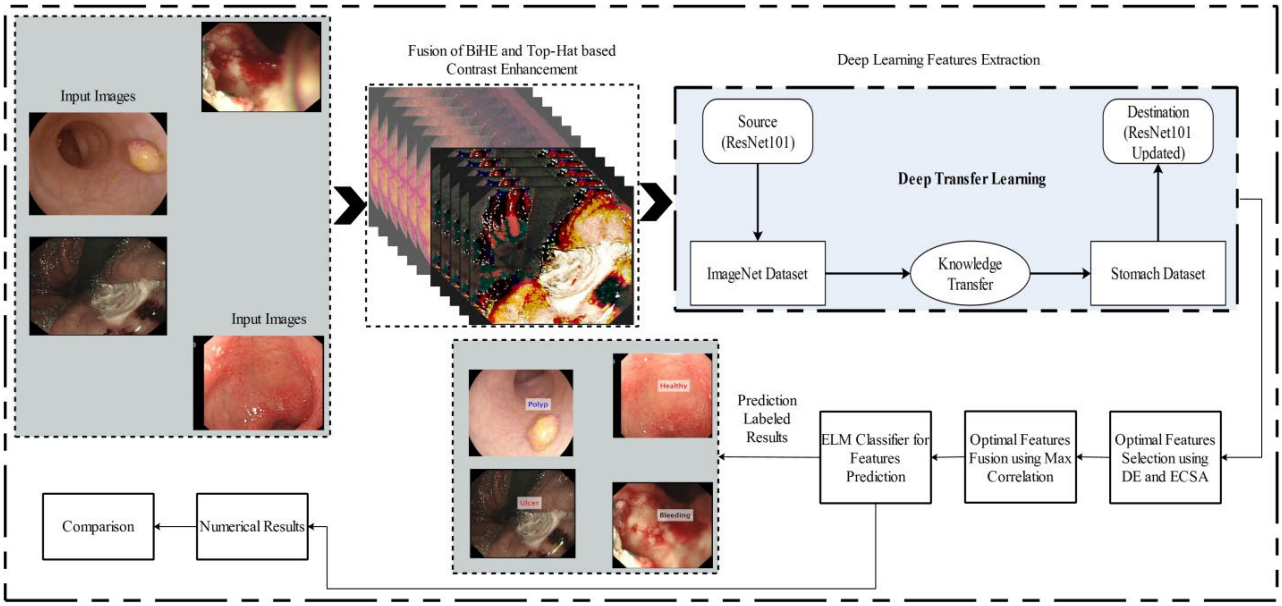

FIGURE 1. Proposed architecture diagram for stomach infections classification.

In summary, the above presented techniques mainly focused on strengthing of the extracted features for the better evaluation of the proposed techniques. Also, they showed the importance of classifiers for improved classification accuracy. The main limitations of these techniques are- i) their failure to extract good features in the classification phase; ii) overfitting problem. In this article, we focused on the contrast enhancement step and fusion of features for classification of stomach infections. We focused on the contrast stretching step to be able to extract good deep learning features. Moreover, we emphasis on the problem of overfitting and then selection of optimal features. To solve these issues, we propos a new automated technique. In summary, the proposed technique is evaluated in training and testing steps. We merge the validation step images (15%) with the testing images.

III. PROPOSED METHODOLOGY

In this work, a new fully automated sequential approach is presented for stomach infections classification. Three stomach infections and one healthy class is considered in this work. This approach is implemented in the following steps: (i) improve the contrast of input images by implementing a fusion of BiHE and Top-Hat filtering output; (ii) ResNet101 pretrained model is fine tuned using transfer learning technique to utilize it for classification of diseases; (iii) two optimal vectors are computed using DE and ECSA algorithms; (iv) fusion of both resultant optimal solutions using max correlation approach, and (v) classification of these features using ELM. The ELM classifier returns labeled output and numerical values. A detailed architecture of the proposed method is illustrated in Figure 1.

A. DATABASE PREPARATION

In this article, the following datasets are used for the evaluation of proposed technique- CUI Wah Private [31], Kvasir-SEG [33], CVC-ClinicDB [34], ETIS-Larib [35], and ASU-Mayo Clinic Colonoscopy Video Database [36]. The CUI Wah Private Dataset consists of 5000 WCE images in each class and we selected three classes- ulcer, healthy, and bleeding. All images are in RGB format of dimension 512 × 512. For polyp images, the Kvasir-SEG (13 polyp images), CVC-ClinicDB (612 images), ETIS-Larib (196), and ASU-Mayo Clinic Colonoscopy Video Database (3500 polyp images) are utilized. We combined all these polyp images in one class resulting in the total number of polyp images to be 4321. A few sample images are shown in Figure 2. A summary of overall images is given in the table below.

Class	Database Name	Total Images
Ulcer	CUI Wah	5000
Bleeding	CUI Wah	5000
Healthy	CUI Wah	5000
Polyp	Kvasir-SEG	13
	CVC-ClinicDB	612
	ETIS-Larib	196
	ASU-Mayo Clinic Colonoscopy Video Database	40

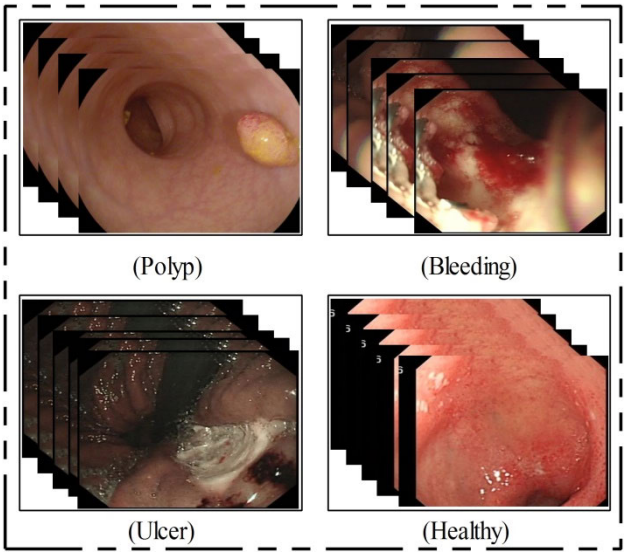

FIGURE 2. Sample database images.

B. CONTRAST ENHANCEMENT OF WCE FRAMES

The level of contrast of several medical images is not sufficient for accurate features extraction. Most of the captured images are blurred and have weak edges among adjacent tissues [37]. The features of these images can be enhanced using contrast enhancement techniques which play a great help in the accurate diagnosis. In this work, we proposed a hybrid approach for WCE images contrast enhancement. In this approach, we fused brightness preserving Bi histogram equalization (BiHE) image and top-hat filtering image to get enhanced image. This approach avoids the saturation affects and increases the intensity level of tumor region. Mathematically, this approach is formulated as follows:

Consider we have an RGB image of dimension $512 \times 512 \times 3$ and denoted by $\Psi(i, j)$. BiHE is based on the mean brightness threshold value [38]. It is divided in two parts using following histogram equation:

$$h_{low}(k) = \frac{g(k)}{N_{low}} \quad for\ k \in 0, 1, 2, \ldots \tilde{k} \tag{1}$$

$$h_{high}(k) = \frac{g(k)}{N_{high}} \quad for\ k \in \tilde{k}+1, \tilde{k}+2, \ldots, L \tag{2}$$

where, $h_{low}(k)$ and $h_{high}(k)$ represent low and high histogram pixels, respectively. The cumulative density function (CDF) of both histograms is calculated as follows:

$$CDF_{low}(k) = \sum_{j=0}^{K} h_{low}(j), \quad for\ k \in 0, 1, 2, \ldots \tilde{k} \tag{3}$$

$$CDF_{high}(k) = \sum_{j=\tilde{k}+1}^{K} h_{high}(j), \quad for\ k \in \tilde{k}+1, \tilde{k}+2, \ldots, L \tag{4}$$

In this approach CDF is utilized to reallocate the brightness value of a histogram transfer function. This transfer function is defined as follows:

$$F_{low}(k) = K_o + \left(K_{\tilde{k}} - K_o\right).CDF_{low}(k) \tag{5}$$

$$F_{high}(k) = K_{\tilde{k}+1} + (K_L - K_{k+1}).CDF_{high}(k) \tag{6}$$

Later, the HE transfer function is implemented to get the output of BBHE as:

$$\psi = \{\phi(i, j)\} = F_{low}(K_{low}) \cup F_{high}\left(K_{high}\right) \tag{7}$$

where, $F_{low}(K_{low}) = \{F_{low}(K(i, j)) | \forall K(i, j) \in K_{low}\}$, $F_{high}\left(K_{high}\right) = \left\{F_{high}(K(i, j)) | \forall K(i, j) \in K_{high}\right\}$. The output of BiHE is denoted by $\phi(i, j)$ and fused in Top-Hat filtering image. The top-hat filtering is defined as:

$$Top(i, j) = \Psi(i, j) - \Psi(i, j) \cdot s \tag{8}$$

where, s represents threshold factor initialized as 20. The resultant top-hat filtering image is fused along with $\phi(i, j)$ as follows:

$$\tilde{\Psi}(i, j) = \sum (Top(i, j), \phi(i, j)) - \Psi(i, j) \tag{9}$$

The resultant image $\tilde{\Psi}(i, j)$ is shown in Figure 3. In this figure (third column) shows that the infected region is enhanced as compared to the healthy region. This resultant image is passed to the next step for the learning of deep net model.

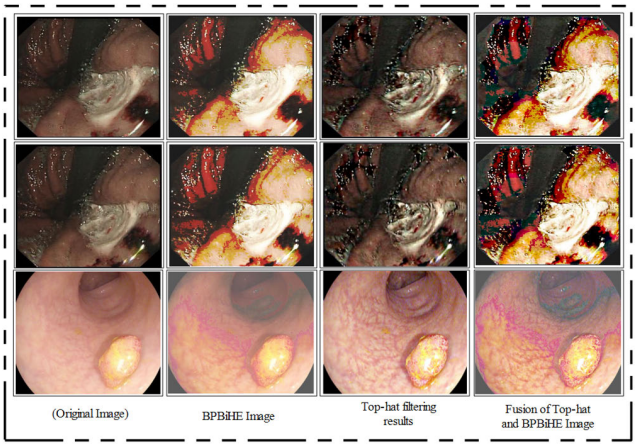

FIGURE 3. Proposed contrast enhancement results of WCE images.

C. ResNet101 DEEP MODEL

In the field of image classification, deep neural network has shown a major breakthrough in the past few years. Naturally, a deep model is the integration of low-level, mid-level, and high-level features along with a classifier. The level of features in a deep model is enriched by several stacked layers. In this work, we utilized ResNet101 [24] for deep features extraction. This architecture was inspired by VGG19 pretrained network and one of the deepest convolutional neural network (CNN) network. As mentioned earlier, a CNN model consists of many layers which are connected to each other. These layers are trained for various tasks like medical image classification etc. In the end of layers of this network, several levels of features are learned. In Resnet101, convolutional filter size is 33 and stride is 2. In the convolutional layers, downsampling is performed based on the value of stride. Originally, this network consists of 347 layers and

379 connections. The input of this network is of dimension $\mathbf{224 \times 224 \times 3}$. In the first convolution layer, filter size is $[\mathbf{7}, \mathbf{7}]$, number of channels is 3 and number of filters is 64. In the max pooling layers, the filter size is $\mathbf{3 \times 3}$ and stride is 2. For the second convolutional layer, the number of channel and filters is 64. In the last convolution layer, the number of filters is 2048 and channels equal to 512. We extract features from global average pool layer and get an output vector of dimension $\boldsymbol{N \times 2048}$, where $\boldsymbol{N}$ denotes number of features. A complete architecture of ResNet101 is illustrated in Figure 4.

Layer Name	Output Size	Architecture
conv 1	112x112	7x7, 64, stride 2
conv 2_x	56x56	3x3 maxpool, stride 2 $\begin{bmatrix} 1\times1,64 \\ 3\times3,64 \\ 1\times1,256 \end{bmatrix} \times 3$
conv 3_x	28x28	$\begin{bmatrix} 1\times1,128 \\ 3\times3,128 \\ 1\times1,512 \end{bmatrix} \times 4$
conv 4_x	14x14	$\begin{bmatrix} 1\times1,256 \\ 3\times3,256 \\ 1\times1,1024 \end{bmatrix} \times 23$
conv 5_x	7x7	$\begin{bmatrix} 1\times1,512 \\ 3\times3,512 \\ 1\times1,2048 \end{bmatrix} \times 3$
	1x1	average pool, 1000-d fc, softmax
FLOPs		$7.6\text{x}10^9$

FIGURE 4. ResNet101 architecture.

D. TRANSFER LEARNING BASED NETWORK TRAINING

Data dependence is a serious problem in deep learning. For the training of a deep model, a massive amount of data is required as compared to the traditional machine learning techniques. The main purpose behind this requirement of massive amount of training data is due to its need to learn the latent patterns. However, in few research domains, especially in medical imaging, a massive amount of data is not typically available for training a deep learning model. Transfer learning (TL) [22] is a concept of training a model with fewer amount of data. In TL, it is not essential to train the target model from scratch. Mathematically, deep transfer learning is defined as:

Given a transfer learning task defined by $\langle \mathbb{D}_s, \boldsymbol{T}_s, \mathbb{D}_t, \boldsymbol{T}_t, \boldsymbol{F}_t(\cdot)\rangle$. It is a deep transfer learning where $\mathbb{D}_s$ is source domain, $\mathbb{D}_t$ is target domain, $\boldsymbol{T}_s$ is learning task from source and destination, and $\boldsymbol{F}_t(\cdot)$ represents non-linear function that reflects a deep neural network. Visually, the process of model learning using TL is illustrated by Figure 5. In this figure, it is shown that original ResNet101 model is trained on ImageNet dataset [39] and then knowledge is transfered using deep

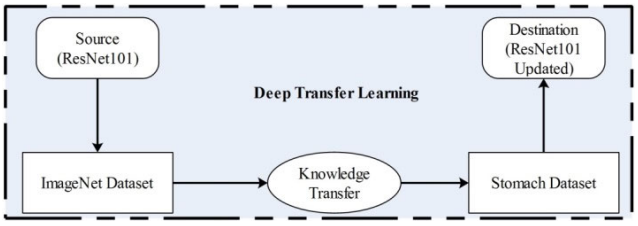

FIGURE 5. Process of deep transfer learning for stomach data learning.

TL for retraining this model on destination database. The stomach database is utilized as a destination database. After learning ResNet101 on stomach database, we extract features from global pooling layer and output a vector of dimension $\boldsymbol{N \times 2048}$. In the learning, we initialized the learning rate of 0.0001 and mini batch size of 28.

Training Process: Figure 6 shows the detailed training process of deep learning model. This figure is described as follows:

- Separate the 50% training images from each class such as polyp, ulcer, bleeding, and healthy. These images are separated by employing randomized process.
- Deep transfer learning based training of ResNet101 model for stomach infections classification.
- Extract deep learning features from global average pool layer.
- Optimze the extracted features through two optimization techniques- DE and ECSA.
- Two optimal vectors are retuned as the output.
- Maximum correlation based fusion of both optimal vectors.
- Train ELM classifier and save model for testing step.

Based on the above steps, we train and save our model for the classification of stomach infections. Details of the optimization, fusion and classification steps are presented in the following.

E. FEATURES OPTIMIZATION

The selection of most optimal set of features from the original set of features, improves the classification accuracy [40]. These features are selected from original features with the least error for learning. The key benefits are- it improves accuracy, consumes less time, and removes the problem of overfitting. The optimization process in feature selection is finding the best possible values based on the implemented objective function. For this purpose, many evolutionary techniques are presented for finding the nearest optimal solution. In this article, we implemented two algorithms- Differential Evolution (DE) and Enhanced Crow Search (ECS) algorithm.

1) DIFFERENTIAL EVOLUTION

The DE is an evolutionary algorithm which is used for global search optimization problems [41]. This algorithm is easy as compared to Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) as it is based on fewer control parameters. Due to fewer control parameters, it is much useful

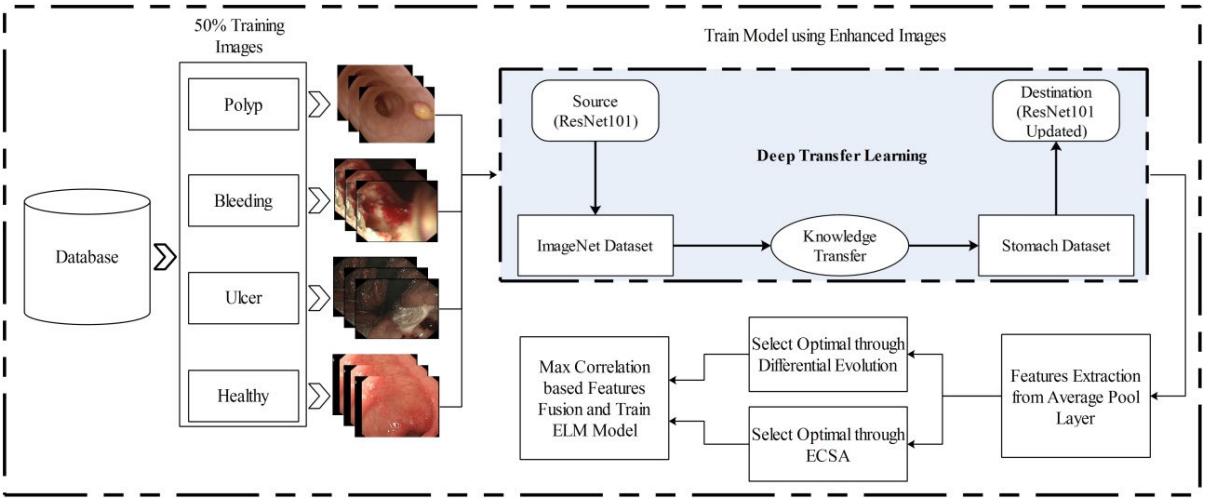

FIGURE 6. Proposed training process of deep learning model for stomach infections classification.

in the area of medical imaging. It starts with some initial values, which are randomly generated in the search space. Then mutation and crossover on input data is performed, which follows the selection process to generate a new population. The steps which are followed in this work are defined below.

Input: Original deep feature vector of dimension $N \times 2048$.
Output: Optimal feature vector of dimension $N \times 1242$.
Step 1: Initialize parameters

- Population = 50
- Minimum bound and Maximum bound
- Compute these bounds through following expression:

$$\lambda_i^j = \lambda_{min}^j + rand\ (0, 1)\ .\ \left(\lambda_{max}^j - \lambda_{min}^j\right) \tag{10}$$

Step 2: Calculate fitness function using Fine KNN
Step 3: Perform mutation

- $M_r = 0.2$

Mutation is defined through following equation:

$$v_i^j = \lambda_{r_1}^j + F\left(\lambda_{r_2}^j - \lambda_{r_3}^j\right) \tag{11}$$

- Where F represent the scaling mutation factor of range between (0, 2). Source vectors are denoted by $\lambda_{r_1}, \lambda_{r_2}, \lambda_{r_3}$ which are randomly chosen from the population. The symbols r_1, r_2, r_3, and i must be distinct to each other.

Step 4: Perform crossover

- $C_r = 0.7$

$$C_i^j = \begin{cases} v_i^j, & if\ rand\ (0, 1) \le C_r or j = j_{rand} \\ \lambda_i^j, & Otherwise \end{cases} \tag{12}$$

- Where j_{rand} represent randomly chosen integers between [0, D]
- D represents the number of variables in the problem to be solved.

Step 5: Find fitness evaluation and selection and repeat step 2, 3 and 4 until the required optimal feature vector is obtained. In the output, an optimal feature vector of dimension $N \times 1242$ is obtained.

2) ENHANCED CROW SEARCH

The ECSA [42] starts at the initial positions for a set of crows population. These positions are converted into binary solutions. The binary solutions represented as the set of selected features that are later analyzed through fitness function. Based on the fitness function, best solutions are determined and updated through local search approach. This process is continued until all the iterations are completed to get an optimal solution. Initially, values from the population are defined as:

$$Population = p_{gen}\left(gen = 0, 1, \ldots, Max_{gen}\right) \tag{13}$$

$$p_{gen} = \begin{bmatrix} u_{gen}^{(1,1)} & u_{gen}^{(1,2)} & \ldots & u_{gen}^{(1,M)} \\ u_{gen}^{(2,1)} & u_{gen}^{(2,2)} & \ldots & u_{gen}^{(2,M)} \\ \ldots & \ldots & \ldots & \ldots \\ u_{gen}^{(NP-1)} & u_{gen}^{(NP-2)} & \ldots & u_{gen}^{(NP,M)} \end{bmatrix} \tag{14}$$

where, NP denotes number of crows, the crows positions denoted by $u_{gen}^{(i,d)}$ and stored in the matrix p_{gen}, and M represents the total number of features. Each crow is placed at the position chosen randomly as:

$$u_{ij} = l_{ij} + rand \times \left(x_{ij} - l_{ij}\right) \tag{15}$$

where $i \in 1, 2, 3, \ldots, NP$, $j = 1, 2, 3, \ldots, M$. The lower and upper bounds are represented by l_{ij} and x_{ij}, respectively. Later each crow is converted into binary as follows:

$$u_{gen+1}^{(i,d)} = \begin{cases} 1 & if\ \mathbb{S}\left(u_{gen+1}^{(i,d)}\right) \ge \alpha \\ 0 & otherwise \end{cases} \tag{16}$$

Here $\alpha = 0.1$ and employs a sigmoid function. Each crow is evaluated using fitness function. The Fine KNN classifier is

employed as fitness function which calculates Mean Square Error (MSE). Based on the minimum MSE, crow positions are updated. For updation of crow positions, we utilized local search approach, formulated as follows:

$$u_{gen+1}^{(i,s)} = u_{gen}^{(i,s)} + \varphi^{gen} \times \left(m_{gen}^{(d(i),s} - u_{gen}^{(i,s)} \right) \tag{17}$$

where, φ^{gen} denote flight length of crow and $s = 1, 2, 3, \ldots, D$. This process continues until the termination condition is executed. After termination of the algorithm, an optimal feature vector is obtained which is of dimension $N \times 924$.

F. FEATURES FUSION

Consider we have two optimal feature vectors defined by $FV_{(DE)}^{k_1}$ and $FV_{(ECSA)}^{k_2}$. The vector length of each one is $N \times 1242$ and $N \times 924$, respectively. Consider $FV_{(fus)}^{k}$ is a fused feature vector based on maximum correlation coefficient of dimension $N \times K$, where K denotes the length of features which depends on the maximum correlation. The correlation coefficient between two features k_1 and k_2 is calculated as follows:

$$\rho\left(k_1, k_2\right) = \frac{COV\left(k_1, k_2\right)}{\sqrt{var(k_1)}\sqrt{var(k_2)}} \tag{18}$$

The values of $\rho\left(k_1, k_2\right)$ are always lies between $(-1, 1)$. The -1 explain the weak correlation among two features, whereas the 1 denotes the strong correlation among them. But we are interested in the maximum correlation because of strong features selection in the fused vector. Therefore, we implemented maximum correlation among both vectors as follows:

$$CC\left(k_1, k_2\right) = \mathbf{\Phi}\rho\left(g_1\left(k_1\right), g_2(k_2)\right) \tag{19}$$

Here, $\mathbf{\Phi}$ represents Supremum taken overall Borel functions $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$ which lies between $(0, 1)$. The correlation near to 1 represents a strong correlation among two features. Hence, we consider this process and compute correlation among all features and fuse them. The following steps are involved in the fusion process:

- Input both optimal feature vectors
- Perform mean padding to make the dimension of both the vectors same
- Compute maximum correlation $CC\left(k_1, k_2\right)$.
- If correlation is close to 1 then drop both features in fused vector $FV_{(fus)}^{k}$.
- If correlation is close to 0 then discard both features. The detailed numerical re (INCOMPLETE)
- In the output, a final resultant fused feature vector of dimension $N \times 861$ is obtained.

Finally, the resultant fused vector is passed to the extreme learning machine (ELM) [43] classifier to obtain two outputs-prediction results in the labeled form and numerical results. The labeled results are shown in Figure 7, whereas the numerical results are given in Section 4.

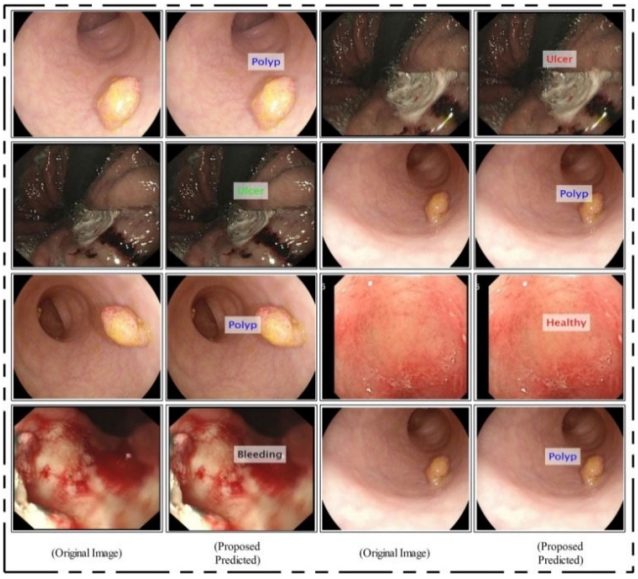

FIGURE 7. Proposed prediction results in the form of labeled images.

IV. RESULTS AND DISCUSSION

A detailed numerical prediction results are presented in this section. Originally, this dataset consists of 100 patient's data and for each patient one video was recorded which is of 8 minutes duration. In each 8 minute video, there are approximately 56,000 frames. To split data for research purpose, the researchers of CUI WAH consulted a specialist to annotate the data according to their classes like ulcer, bleeding, and healthy. In this work, we collected several datasets as mentioned in Section 3.1. We combined these datasets to train a model named Resnet101 which required a handsome amount of data. We split data only in training and testing instead of training, testing, and validation. We merge validating data (15%) in the testing step for fair results such as 50:50. All results are computed in this ratio using 10-Fold Cross Validation. The results are computed in several steps: (i) experiment 1- employed contrast stretching step, trained ResNet101 and extracted deep features; (ii) experiment 2- employed deep features and implemented DE method for optimal solution; (iii) experiment 3- implemented ECSA algorithm on deep features and selected the most optimal solution; (iv) experiment 4- fused both optimal vectors, and (v) experiment 5- compared proposed framework results with other neural nets. The cross validation is defined as 10, whereas the ratio of training model is 50%. The rest 50% images are utilized for prediction. Multiple classifiers are implemented to test the proposed prediction comparison with ELM. The implemented classifiers are Fine tree, Naïve Bayes, cubic KNN, weighted KNN, cubic SVM, and Ensemble Learning. The performance of each classifier is tested through three measures- accuracy, negative rate (100-accuracy) and the prediction time (sec). The MATLAB2020a is used as a simulation tool, whereas the hardware includes- Core-i7 desktop computer containing 16 GB of RAM and 16 GB Nvidia Graphics Processing Unit (GPU).

A. NUMERICAL RESULTS

Results are presented in this section for each experiment discussed in the last section. Table 1 presents the classification results for original ResNet deep features. These features are obtained after applying deep transfer learning. ELM classifier gives higher accuracy of 87.45% as compared to other mentioned classifiers in this table. The negative rate of ELM is 12.55% and prediction time is 176.8978 (sec). Ensemble baggage tree gives second highest accuracy of 85.74%. The negative rate of this classifier is 14.26% and prediction time is 181.4520 (sec). If we compare all these classifies in terms of prediction time, then it can be seen that cubic SVM executed in minimum time of 154.7745 (sec). But, the accuracy of this classifier is less than 3% as compared to ELM and time difference is only 26 (sec). Hence, it can be assumed that the overall performance of ELM classifier is better as compared to other methods. The performance of ELM can be validated through Table 2 (confusion matrix), which shows that bleeding class has highest correct classification rate, whereas ulcer class has highest error rate.

TABLE 1. Prediction accuracy of stomach diseases by applying ResNet original deep features.

Classifier	Performance Measure		
	Accuracy (%)	Negative Rate (%)	Prediction Time (sec)
ELM	87.45	12.55	176.8978
Fine Tree	81.02	18.98	161.6356
Naïve Bayes	82.96	17.04	182.5634
Cubic SVM	84.81	15.19	154.7745
Weighted KNN	83.68	16.32	157.0937
Cubic KNN	83.98	16.02	161.5245
E-Boosted Tree	84.60	15.4	177.4267
E-Baggage Tree	85.74	14.26	181.4520

TABLE 2. Confusion matrix of ELM after applying ResNet original deep features.

Stomach Class	Stomach Class			
	Ulcer	Polyp	Bleeding	Healthy
Ulcer	83.60%	6%	4%	>6%
Polyp	8%	88.58%	<2%	2%
Bleeding	4%	3%	92.30%	<1%
Healthy	<4%	6%	2%	88.30%

Results given in Table 3 are computed after employing DE algorithm on original deep features. Several classifiers are applied and ELM gives a highest accuracy of 91.38%. This accuracy is improved as comparing to the original ResNet features as shown in Table 1. The negative rate of ELM is 8.62% whereas the previous reported error rate

TABLE 3. Prediction accuracy of stomach diseases by applying differential evolution (DE) based features selection.

Classifier	Performance Measure		
	Accuracy (%)	Negative Rate (%)	Prediction Time (sec)
ELM	**91.38**	8.62	110.1775
Fine Tree	85.21	14.79	109.6870
Naïve Bayes	84.80	15.2	114.1957
Cubic SVM	87.16	12.84	102.2506
Weighted KNN	87.02	12.98	119.0573
Cubic KNN	83.98	16.02	105.1683
E-Boosted Tree	84.60	15.4	121.7835
E-Baggage Tree	89.31	10.69	119.2785

TABLE 4. Confusion matrix of ELM by applying DE for features selection.

Stomach Class	Stomach Class			
	Ulcer	Polyp	Bleeding	Healthy
Ulcer	87.90%	4%	3%	<6%
Polyp	4%	92.80%	<2%	2%
Bleeding	3%	<1%	94.10%	2%
Healthy	<3%	3%	4%	90.20%

was 12.55%. Ensemble baggage tree gives second highest accuracy of 89.31% and improved up to 4% as compared to the accuracy given in Table 1. The negative rate of this classifier is 10.69% and previously it was 14.26%. The execution time during the prediction is also noted and given in Table 3. According to this table, cubic SVM is quickly executed as compared to all the other classifiers. The prediction execution time of cubic SVM is 102.2506 (sec) for all the testing images. Hence, in terms of execution-time, cubic SVM performs better but in terms of accuracy, ELM performs much better as compared to others. The performance of ELM can be validated through Table 4 (confusion matrix). In this table, it is shown that the bleeding class has highest correct prediction rate of 94.10% and second highest is 92.80% for polyp class.

Table 5 presents the results of ECS algorithm. The original ResNet model features are applied on ECS algorithm to select the most optimal features. The selected features are passed to multiple classifiers as presented in Table 5. The ELM shows the highest accuracy of 93.46% which is improved as compared to the original ResNet features and DE optimal features. The error rate of ELM is 6.54% and prediction time is 98.6447 (sec). The accuracy of ELM can be verified through Table 6 which shows that the ulcer class achieved correct prediction accuracy of 90.10%, polyp 94.30%, bleeding 95.10%, and healthy class 93.50%, respectively. The second best accuracy is achieved on ensemble baggage tree of 92.10% along

TABLE 5. Prediction accuracy of stomach diseases by applying enhanced crow search algorithm based features selection.

Classifier	Performance Measure		
	Accuracy (%)	Negative Rate (%)	Prediction Time (sec)
ELM	**93.46**	6.54	98.6447
Fine Tree	88.19	11.81	97.0583
Naïve Bayes	86.27	13.73	100.2458
Cubic SVM	89.66	10.34	87.8940
Weighted KNN	89.23	10.77	101.0377
Cubic KNN	87.12	12.88	96.1789
E-Boosted Tree	88.08	11.92	104.5979
E-Baggage Tree	92.10	7.9	106.6845

TABLE 6. Confusion matrix of ELM by applying ECSA for features selection.

Stomach Class	Stomach Class			
	Ulcer	Polyp	Bleeding	Healthy
Ulcer	90.10%	4%	<2%	4%
Polyp	<1%	94.30%	3%	2%
Bleeding		3%	95.10%	<2%
Healthy	3%	<1%	3%	93.50%

with an error rate of 7.9%. In terms of prediction execution time, cubic SVM outperforms but accuracy of this classifier is not sufficient. The noted time of this classifier is 87.8940 (sec), which is better as compared to the other classifiers.

Proposed framework results are presented in Table 7. In this table, the results are calculated by fusion of both optimal vectors using max correlation technique. ELM achieved maximum accuracy of 99.46%. This accuracy is improved up to 6% as compared to the ECS algorithm, 8% as compared to DE and 12% as compared to original ResNet features, respectively. The error rate of ELM for proposed framework is just 0.54%, whereas the highest error rate is 6.38% for Naïve Bayes. The accuracy of ELM can be further verified through Table 8. In this table, it is described that the ulcer class achieved correct prediction accuracy of 98.7%, polyp 99.50%, bleeding 99.70%m and healthy class has 99.30%, respectively. The second best noted accuracy is 98.72% on ensemble baggage tree along with error rate of 1.28%. The prediction time of each classifier is also noted during the testing process and it is observed that cubic SVM executed in minimum time as compared to the other classifiers. But the accuracy of this classifier is 95.28% which is almost 4% less as comparing to ELM. Hence, based on these results, it is evident that the proposed ELM classifier gives better performance. Moreover, Figure 8 shows some important evaluation protocols like sensitivity rate, precision rate, F1-score, and accuracy. A statistical analysis is also conducted in the form of confidence interval and values are tabulated in Table 9. In this table, it is described that the confidence interval for CL 95% is $98.425 \pm 1.434(\pm 1.46\%)$.

TABLE 7. Prediction accuracy of stomach diseases by applying proposed framework after optimal features fusion.

Classifier	Performance Measure		
	Accuracy (%)	Negative Rate (%)	Prediction Time (sec)
ELM	**99.46**	**0.54**	86.6447
Fine Tree	94.96	5.04	80.0583
Naïve Bayes	93.62	6.38	88.2458
Cubic SVM	95.28	4.72	73.8940
Weighted KNN	96.48	3.52	91.0377
Cubic KNN	94.95	5.05	81.1789
E-Boosted Tree	94.48	5.52	90.5979
E-Baggage Tree	98.72	1.28	92.6845

TABLE 8. Confusion matrix of ELM by applying proposed framework.

Stomach Class	Stomach Class			
	Ulcer	Polyp	Bleeding	Healthy
Ulcer	98.70%	1%		<1%
Polyp	<1%	99.50%	<1%	
Bleeding		<1%	99.70%	
Healthy	<1%			99.30%

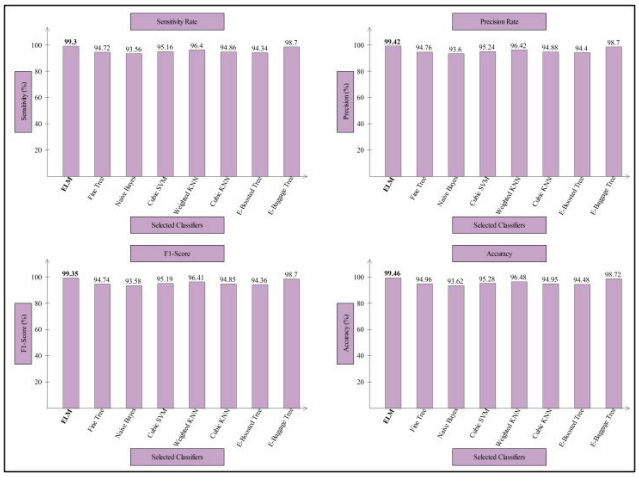

FIGURE 8. Different calculated performance measures for proposed technique.

B. DISCUSSION

A detailed discussion of the proposed framework is conducted in this section. As shown in Figure 1, the proposed framework consists of several important steps; however, it is

TABLE 9. Statistical analysis of proposed classification results.

Method	Min (%)	Avg (%)	Max (%)	σ	CI
ELM	**97.39**	**98.42**	**99.46**	**1.03**	**98.425 ± 1.434(±1.46%)**
Fine Tree	91.26	93.11	94.96	1.85	**93.11 ± 2.564(±2.75%)**
Naïve Bayes	88.97	91.29	93.62	2.32	**91.295 ± 3.22(±3.53%)**
Cubic SVM	92.04	93.66	95.28	1.62	**93.66 ± 2.245(±2.40%)**
Weighted KNN	93.14	94.81	96.48	1.67	**94.81 ± 2.315(±2.44%)**
Cubic KNN	91.16	93.05	94.95	1.89	**93.05 ± 2.62(±2.82%)**
E-Btd Tree	89.98	92.23	94.48	2.25	**92.23 ± 3.118(±3.38%)**
E-Bg Tree	95.13	96.92	98.72	1.79	**96.92 ± 2.488(±2.57%)**

important to analyze the performance of each step. The results are computed initially for original ResNet features and achieved a highest accuracy of 87.45%. After that, optimization algorithms are implemented and their separate accuracies are 91.38% and 93.46%, respectively. But the individual accuracy is not sufficient for comparison with the existing techniques; therefore, we proposed a max correlation based fusion approach and fused both optimal vectors. After fusion, accuracy jumped to 99.46%. Table 10 gives the results for classification of stomach infections without employing contrast stretching step, where best noted accuracy is 94.34%. By the comparison of this accuracy with Table 7, it is noted that on the average 5% accuracy is decreased.

Similarly, for other classifiers, the accuracy on the average is degraded by 4%. However, the computational time is minimized as compared to proposed framework. In Table 7, the best noted time was 73.8940 (sec), whereas in Table 10, the minimum noted time is 62.1667 (sec). But there is a big difference among accuracy rate.

TABLE 10. Prediction accuracy of stomach diseases after without contrast enhancement step along with ELM and other classifiers.

Classifier	Performance Measure		
	Accuracy (%)	Negative Rate (%)	Prediction Time (sec)
ELM	94.34	5.66	72.5395
Fine Tree	88.05	11.95	73.9539
Naïve Bayes	89.28	10.72	69.0473
Cubic SVM	90.68	9.32	62.1667
Weighted KNN	92.20	7.8	78.0530
Cubic KNN	91.82	8.18	70.5738
E-Boosted Tree	90.10	9.9	72.0067
E-Baggage Tree	93.94	6.06	79.2779

In addition, we compare the performance of ELM classifier with Softmax and results are given in Table 11. In this table, it is described that the results are presented with and without contrast stretching step. With contrast stretching step, the maximum noted accuracy is 94.87%, whereas without contrast strehcing step, accuracy is 91.67%. We also analyse the significance of this step by executing our methods in 200 times. After that getted minimum accuracy, average accuracy, maximm accuracy, standard deviation, and confidence interval (CI) for fusion approach. These values are plotted in Figure 9. In this figure, it is illustrated that the confidence level (CL) of sotmax classifier using contrast stretching step is (95%, $1.960\sigma_{\bar{x}}$) and confidence interval (CI) is $93.61 \pm 1.746(\pm 1.87\%)$. The CL of Softmax classifier without using contrast stretching step (WoCS) is (95%, $1.960\sigma_{\bar{x}}$) and CI is $89.93 \pm 2.412(\pm 2.68\%)$.

TABLE 11. Prediction accuracy of stomach diseases using Softmax classifier, where classifier is applied with and without contrast stretching. * M1 denotes ResNet101, M2 (DE), M3(ESCA), and M4(fusion).

Method	Validation Steps				Performance Measure		
	M1	M2	M3	M4	ACC (%)	SEN (%)	FNR (%)
Softmax with CS	✓				83.19	83.14	16.86
		✓			91.62	91.60	8.40
			✓		92.49	92.43	7.57
				✓	94.87	94.80	5.20
Softmax WoCS	✓				80.04	80.10	19.90
		✓			86.45	86.34	13.66
			✓		89.36	89.40	10.60
				✓	91.67	91.62	8.38

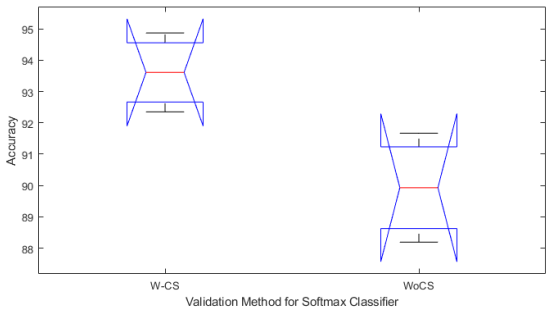

FIGURE 9. Statisticl analysis of Softmax classifier.

We also conducted a comparison of the proposed method with other neural nets such as Alexnet, Vgg16, etc. results are given in Table 12. This table, the results are described for both ELM and Softmax classifiers. The main reason of this experiment is the importance of the choice of deep neural network like ResNet101 for this work. In this table, it is noted that Vgg16 achieve an accuracy of 95.38%, Vgg19 achieves 96.80%, AlexNet achieve 94.26%, GoogleNet achieve 93.47%, and ResNet50 achieve 97.90%, respectively. For the Softmax classifier, the maximum attained accuracy is 94.87%. Our proposed framework achieved an accuracy of 99.46% using ELM and 94.87% using Softmax classi-

TABLE 12. Comparison of proposed classification accuracy with other neural nets using ELM and Softmax classifiers.

Method	Different CNN Network based Comparison						Measure
	Vgg16	Vgg19	AlexNet	GoogleNet	ResNet50	Proposed Framework	Accuracy (%)
ELM	❖						95.38
		❖					96.80
			❖				94.26
				❖			93.47
					❖		97.90
							99.46
Softmax	❖						92.87
		❖					94.75
			❖				92.26
				❖			90.66
					❖		91.04
						❖	**94.87**

fier. This difference in accuracy also showing the choice of ELM as a classifier for stomach diseases classification. All other neural nets are replace with ResNet101 in the proposed framework (Figure 1) to obtain these results.

Comparison with a few relevant published techniques is presented in Table 13. In this table, we added the details of the total number of images and accuracy results. In [14], authors presented a deep learning based approach and later they applied genetic algorithm for the selection of important features. In the experimental process, they utilized a total of 9,889 WCE images and achieved classification accuracy of 96.50%. Authors in [13] presented a CNN-based approach and utilized 6,000 WCE images for experimental process. On the selected images, they achieved an accuracy of 99.46%. In our proposed method, we utilized in total three stomach infection classes and one healthy class. Each class consists of 4,000 WCE images and achieved an accuracy of 99.46%. This accuracy shows that the proposed method gives significant performance as compared to the-state-of-the-art techniques.

TABLE 13. Comparison with exiting techniques.

Ref	Year	Total Images	Accuracy (%)
[14]	2020	9,889	96.50%
[13]	2020	6,000	98.40%
Proposed	2020	16,000	99.46%

V. CRITICAL ANALYSIS

A detailed critical analysis is conducted in this section to analyze the performance of the implemented contrast stretching approach. Table 14 shows the results of various contrast

TABLE 14. Comparison of contrast enhancement approach with few other techniques.

Method	Measures			
	Accuracy (%)	Sensitivity (%)	Error (%)	Time (sec)
BPHE	97.39	97.23	2.77	71.3664
Top-hat	96.50	96.34	3.66	50.0982
CLAHE	95.16	95.04	4.60	52.1006
Weiner Filtering	94.87	94.56	5.44	63.9240
Proposed	99.46	99.34	0.66	86.6447

enhancement approaches such as brightness preserving HE, CLAHE, top-hat filtering, Weiner filter, and the proposed approach. The MATLAB codes of these algorithms are available online except the proposed method and we embedded these methods in the place of proposed contrast enhancement approach. Using these techniques, we conducted experiments and based on the results it is concluded that the performance of the proposed approach is better in terms of accuracy. However, in terms of the computational time, other methods perform better as compared to the proposed approach. Based on these results, it can be deduced that the proposed method is most suitable for this work. To the strength of the these results, we also conducted a statistical analysis and perform execution in 200 times. The results are tabulated in Table 15. In this table, it is described that the CI of proposed approach is $98.425 \pm 1.434(\pm 1.46\%)$ for confidence level $(95\%, 1.96\sigma_{\bar{x}})$. The σ is also calculated for each technique and values are 0.76, 1.39, 1.1, 2.10, and 1.035, respectively. From this table, it is show that the proposed results are significant better as compared to other listed techniques.

TABLE 15. Statistical analysis of proposed method for different contrast enhancement filters using ELM classifiers.

Method	Min (%)	Avg (%)	Max (%)	σ	CI
BPHE	95.86	96.62	97.39	0.76	96.625 $\pm$ 1.06($\pm$1.10%)
Top-Hat	93.72	95.11	96.50	1.39	95.11 $\pm$ 1.926($\pm$2.66%)
CLAHE	92.96	94.06	95.16	1.1	94.05 $\pm$ 1.525($\pm$1.62%)
Weinner	90.66	92.76	94.87	2.10	92.765 $\pm$ 2.917($\pm$3.14%)
Proposed	**97.39**	**98.425**	**99.46**	**1.03**	**98.425 $\pm$ 1.434($\pm$1.46%)**

In addition, we compared the performance of selected optimization algorithms with conventional techniques such as Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA). We implemented these methods in the proposed framework and computed the results. The results are computed with two different classifiers- Softmax and ELM. Results are tabulated in Table 16 Table 17. In these tables,

TABLE 16. Comparison of metaheuristic techniques results with conventional techniques using ELM classifier.

Method	Min (%)	Avg (%)	Max (%)	σ	CI
PCA	81.96	83.52	85.08	1.56	83.52 $\pm$ 2.162($\pm$2.59%)
LDA	79.42	81.3	83.18	1.88	81.3 $\pm$ 2.606($\pm$3.20%)
DE	87.66	89.64	91.62	1.98	89.64 $\pm$ 2.744($\pm$3.06%)
ESCA	**89.84**	**91.16**	**92.47**	**1.32**	**91.16 $\pm$ 1.836($\pm$2.01%)**

TABLE 17. Comparison of metaheuristic techniques results with conventional techniques using Softmax classifier.

Method	Min (%)	Avg (%)	Max (%)	σ	CI
PCA	78.44	80.5	82.56	2.06	80.5 $\pm$ 2.855($\pm$3.55%)
LDA	76.18	77.9	79.62	1.72	77.9 $\pm$ 2.384($\pm$3.06%)
DE	83.46	86.46	89.46	3.0	86.46 $\pm$ 4.158($\pm$4.81%)
ESCA	**86.84**	**88.44**	**90.04**	**1.6**	**88.44 $\pm$ 2.217($\pm$2.51%)**

it is shown that the metaheuristic techniques are significantly performed well on metaheuristic techniques. Also, the ELM classifier gives better results and it is significantly proved by both tables (16 and 17).

Further, we testd our proposed scheme without using contrast enhancement and features fusion. In this strategy, we selected Softmax as a classifier and series model for feature extraction. We extractd features from FC layer 7 and performed optimization using ECSA, DE, and PCA. Results are given in Table 18. In comparison with the results of the proposed method presented in Table 9, the results of this strategy are too far and not significant. Furthermore, this comparison also shows that the fusion process has a great impact in improving the accuracy rate. In the last, we also test our method on one single dataset named "The Kvasir Dataset". The results are tabulated in Table 19. From this table, it is also show that the proposed method work significantly better for single dataset.

TABLE 18. Comparison of classification results without using contrast stretching and features fusion.

Model	Optimization Technique			Measures		
	ECSA	DE	PCA	Sen (%)	Accuracy (%)	Time (sec)
VGG16	✓			89.46	90.10	56.8853
		✓		88.62	88.70	54.6730
			✓	85.14	85.24	46.0078
VGG19	✓			89.94	90.20	59.6824
		✓		89.16	89.32	55.0773
			✓	86.22	86.40	47.3221
AlexNet	✓			88.04	88.10	53.6483
		✓		87.86	87.90	50.7710
			✓	84.12	84.34	41.6355

TABLE 19. Testing of proposed method on single dataset named.

Method	Min (%)	Avg (%)	Max (%)	σ	CI
ELM	**95.24**	**98.42**	**98.86**	**1.03**	**98.425 $\pm$ 1.434($\pm$1.46%)**
Fine Tree	88.44	90.84	93.24	2.4	90.84 $\pm$ 3.326($\pm$3.66%)
Naïve Bayes	86.91	89.535	92.16	2.62	89.535 $\pm$ 3.638($\pm$4.06%)
Cubic SVM	90.52	92.90	95.28	2.38	92.9 $\pm$ 3.299($\pm$3.55%)
Weighted KNN	91.04	93.03	95.02	1.99	93.03 $\pm$ 2.758($\pm$2.96%)
Cubic KNN	89.63	90.92	92.21	1.29	90.92 $\pm$ 1.788($\pm$1.97%)
E-Btd Tree	87.80	89.98	92.16	2.18	89.98 $\pm$ 3.021($\pm$3.36%)
E-Bg Tree	93.22	94.65	96.08	1.43	94.65 $\pm$ 1.982($\pm$2.09%)

VI. CONCLUSION

Diagnosis of stomach abnormalities using computerized techniques is showing a great help for accurate identification from WCE images. The manual procedure of stomach abnormalities is not a good choice as it is time consuming and expensive. Fully automated optimal deep features fusion based architecture is proposed in this work for multiple stomach abnormalities classification. A database of WCE images is prepared which consists of four different categories of stomach abnormalities to perform evaluation. The proposed method achieved an accuracy of 99.46% which is the highest as compared to the existing techniques. Based on the results presented in this work, it is observed that the preprocessing step is useful in the learning process of a CNN model. Through learned CNN model, few redundant and irrelevant features were still perceived. Therefore, it is essential to select the most optimal solutions. Also, we perceived that the fusion of optimal features improved the accuracy and also had an impact on the system prediction time. In the future studies, we will consider more clinical data and train a CNN model from the scratch. If the model predicts accurate results, it can be helpful for the researchers working in this field of research.

REFERENCES

[1] K. Muhammad, S. Khan, J. D. Ser, and V. H. C. de Albuquerque, "Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 30, 2020, doi:10.1109/TNNLS.2020.2995800.
[2] F. Chen, K. Muhammad, and S.-H. Wang, "Three-dimensional reconstruction of CT image features based on multi-threaded deep learning calculation," *Pattern Recognit. Lett.*, vol. 136, pp. 309–315, Aug. 2020.
[3] I. Mehmood, M. Sajjad, K. Muhammad, S. I. A. Shah, A. K. Sangaiah, M. Shoaib, and S. W. Baik, "An efficient computerized decision support system for the analysis and 3D visualization of brain tumor," *Multimedia Tools Appl.*, vol. 78, no. 10, pp. 12723–12748, May 2019.

[4] A. Rehman, M. A. Khan, T. Saba, Z. Mehmood, U. Tariq, and N. Ayesha, "Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture," *Microsc. Res. Technique*, to be published.

[5] M. A. Khan, M. Qasim, H. M. J. Lodhi, M. Nazir, K. Javed, S. Rubab, A. Din, and U. Habib, "Automated design for recognition of blood cells diseases from hematopathology using classical features selection and ELM," *Microsc. Res. Technique*, to be published.

[6] S. G. Little, K. A. Akin-Little, E. G. Waldon, and P. Garzaro, "Intestinal disorders," Tech. Rep., 2006.

[7] K. Muhammad, S. Khan, N. Kumar, J. Del Ser, and S. Mirjalili, "Vision-based personalized wireless capsule endoscopy for smart healthcare: Taxonomy, literature review, opportunities and challenges," *Future Gener. Comput. Syst.*, vol. 113, pp. 266–280, Dec. 2020.

[8] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA, A Cancer J. for Clinicians*, vol. 66, no. 1, pp. 7–30, Jan. 2016.

[9] Y. Fu, W. Zhang, M. Mandal, and M. Q.-H. Meng, "Computer-aided bleeding detection in WCE video," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 2, pp. 636–642, Mar. 2014.

[10] G. Iddan, G. Meron, and A. Glukhovsky, "Wireless capsule endoscopy," *Nature*, vol. 405, p. 417, May 2000.

[11] Z. Fireman, "Diagnosing small bowel Crohn's disease with wireless capsule endoscopy," *Gut*, vol. 52, no. 3, pp. 390–392, Mar. 2003.

[12] A. Liaqat, M. A. Khan, M. Sharif, M. Mittal, T. Saba, K. S. Manic, and F. N. H. Al Attar, "Gastric tract infections detection and classification from wireless capsule endoscopy using computer vision techniques: A review," *Current Med. Imag. Formerly Current Med. Imag. Rev.*, vol. 16, Apr. 2020.

[13] 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.

[14] 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," *Microsc. Res. Technique*, vol. 83, no. 5, pp. 562–576, May 2020.

[15] M. A. Khan, M. Sharif, T. Akram, M. Yasmin, and R. S. Nayak, "Stomach deformities recognition using rank-based deep features selection," *J. Med. Syst.*, vol. 43, no. 12, p. 329, Dec. 2019.

[16] M. A. Khan, M. A. Khan, F. Ahmed, M. Mittal, L. M. Goyal, D. Jude Hemanth, and S. C. Satapathy, "Gastrointestinal diseases segmentation and classification based on duo-deep architectures," *Pattern Recognit. Lett.*, vol. 131, pp. 193–204, Mar. 2020.

[17] M. Sharif, M. A. Khan, M. Rashid, M. Yasmin, F. Afza, and U. J. Tanik, "Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images," *J. Experim. Theor. Artif. Intell.*, pp. 1–23, Feb. 2019.

[18] S. A. Khan, M. A. Khan, O.-Y. Song, and M. Nazir, "Medical imaging fusion techniques: A survey benchmark analysis, open challenges and recommendations," *J. Med. Imag. Health Informat.*, vol. 10, no. 11, pp. 2523–2531, Nov. 2020.

[19] A. Liaqat, M. A. Khan, J. H. Shah, M. Sharif, M. Yasmin, and S. L. Fernandes, "Automated ulcer and bleeding classification from WCE images using multiple features fusion and selection," *J. Mech. Med. Biol.*, vol. 18, no. 4, Jun. 2018, Art. no. 1850038.

[20] T. Akram, M. A. Khan, M. Sharif, and M. Yasmin, "Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features," *J. Ambient Intell. Humanized Comput.*, pp. 1–20, Sep. 2018.

[21] M. A. Khan, I. Ashraf, M. Alhaisoni, R. Damaševičius, 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, p. 565, Aug. 2020.

[22] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Hershey, PA, USA: IGI Global, 2010, pp. 242–264.

[23] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[27] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[28] J.-Y. Yeh, T.-H. Wu, and W.-J. Tsai, "Bleeding and ulcer detection using wireless capsule endoscopy images," *J. Softw. Eng. Appl.*, vol. 7, p. 422, May 2014.

[29] M. A. Khan, K. Javed, S. A. Khan, T. Saba, U. Habib, J. A. Khan, and A. A. Abbasi, "Human action recognition using fusion of multiview and deep features: An application to video surveillance," *Multimedia Tools Appl.*, pp. 1–27, Mar. 2020.

[30] H. Alaskar, A. Hussain, N. Al-Aseem, P. Liatsis, and D. Al-Jumeily, "Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images," *Sensors*, vol. 19, no. 6, p. 1265, Mar. 2019.

[31] 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 Appl.*, vol. 78, no. 19, pp. 27743–27770, Oct. 2019.

[32] A. K. Kundu, S. A. Fattah, and K. A. Wahid, "Least square saliency transformation of capsule endoscopy images for PDF model based multiple gastrointestinal disease classification," *IEEE Access*, vol. 8, pp. 58509–58521, 2020.

[33] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," in *Proc. Int. Conf. Multimedia Modeling*, 2020, pp. 451–462.

[34] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. Saliency maps from physicians," *Computerized Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015.

[35] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, no. 2, pp. 283–293, Mar. 2014.

[36] J. Bernal, J. Sánchez, and F. Vilariño, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognit.*, vol. 45, no. 9, pp. 3166–3182, Sep. 2012.

[37] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, Feb. 1997.

[38] C. J. Kuo and H. Wu, "Gaussian probability bi-histogram equalization for enhancement of the pathological features in medical images," *Int. J. Imag. Syst. Technol.*, vol. 29, no. 2, pp. 132–145, Jun. 2019.

[39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[40] N. Naheed, M. Shaheen, S. A. Khan, M. Alawairdhi, and M. A. Khan, "Importance of features selection, attributes selection, challenges and future directions for medical imaging data: A review," *Comput. Model. Eng. Sci.*, vol. 125, no. 1, pp. 314–344, 2020.

[41] K. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. Berlin, Germany: Springer, 2006.

[42] S. Ouadfel and M. Abd Elaziz, "Enhanced crow search algorithm for feature selection," *Expert Syst. Appl.*, vol. 159, Nov. 2020, Art. no. 113572.

[43] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.