

Article

GestroNet: A Framework of Saliency Estimation and Optimal Deep Learning Features Based Gastrointestinal Diseases Detection and Classification

Muhammad Attique Khan ^{1,*} , Naveera Sahar ², Wazir Zada Khan ², Majed Alhaisoni ³, Usman Tariq ⁴ , Muhammad H. Zayyan ⁵ , Ye Jin Kim ⁶ and Byoungchol Chang ^{7,*}

¹ Department of Computer Science, HITEC University, Taxila 47080, Pakistan

² Department of Computer Science, University of Wah, Wah Cantt, Rawalpindi 47040, Pakistan

³ Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

⁴ Department of Management Information Systems, CoBA, Prince Sattam Bin Abdulaziz University, Al-Khraj 16278, Saudi Arabia

⁵ Computer Science Department, Faculty of Computers and Information Sciences, Mansoura University, Mansoura 35516, Egypt

⁶ Department of Computer Science, Hanyang University, Seoul 04763, Korea

⁷ Center for Computational Social Science, Hanyang University, Seoul 04763, Korea

* Correspondence: attique.khan@hitecuni.edu.pk (M.A.K.); bcchang@hanyang.ac.kr (B.C.)



Citation: Khan, M.A.; Sahar, N.; Khan, W.Z.; Alhaisoni, M.; Tariq, U.; Zayyan, M.H.; Kim, Y.J.; Chang, B. GestroNet: A Framework of Saliency Estimation and Optimal Deep Learning Features Based Gastrointestinal Diseases Detection and Classification. *Diagnostics* **2022**, *12*, 2718. <https://doi.org/10.3390/diagnostics12112718>

Academic Editor: Jae-Ho Han

Received: 2 October 2022

Accepted: 4 November 2022

Published: 7 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: In the last few years, artificial intelligence has shown a lot of promise in the medical domain for the diagnosis and classification of human infections. Several computerized techniques based on artificial intelligence (AI) have been introduced in the literature for gastrointestinal (GIT) diseases such as ulcer, bleeding, polyp, and a few others. Manual diagnosis of these infections is time consuming, expensive, and always requires an expert. As a result, computerized methods that can assist doctors as a second opinion in clinics are widely required. The key challenges of a computerized technique are accurate infected region segmentation because each infected region has a change of shape and location. Moreover, the inaccurate segmentation affects the accurate feature extraction that later impacts the classification accuracy. In this paper, we proposed an automated framework for GIT disease segmentation and classification based on deep saliency maps and Bayesian optimal deep learning feature selection. The proposed framework is made up of a few key steps, from preprocessing to classification. Original images are improved in the preprocessing step by employing a proposed contrast enhancement technique. In the following step, we proposed a deep saliency map for segmenting infected regions. The segmented regions are then used to train a pre-trained fine-tuned model called MobileNet-V2 using transfer learning. The fine-tuned model's hyperparameters were initialized using Bayesian optimization (BO). The average pooling layer is then used to extract features. However, several redundant features are discovered during the analysis phase and must be removed. As a result, we proposed a hybrid whale optimization algorithm for selecting the best features. Finally, the selected features are classified using an extreme learning machine classifier. The experiment was carried out on three datasets: Kvasir 1, Kvasir 2, and CUI Wah. The proposed framework achieved accuracy of 98.20, 98.02, and 99.61% on these three datasets, respectively. When compared to other methods, the proposed framework shows an improvement in accuracy.

Keywords: stomach diseases; contrast enhancement; saliency estimation; Bayesian optimization; features optimization

1. Introduction

Cancer is a deadly disease that is currently the leading cause of death worldwide [1]. Cancer is caused primarily by genetics, but it is also influenced by environmental factors. Environmental factors such as a person's eating habits and community behaviors are the

leading causes of 50% of cancer cases. This disease typically develops after 20–30 years of exposure to harmful cancer-causing agents [2]. Modern medical technologies are being used to detect cancer in its early stages, with radical resection being used to treat approximately 50% of cases. The stomach is a vital muscular part of the body that aids in food digestion. It is positioned to the left of the upper abdomen. The food is removed from the esophagus via a muscular valve known as the gastroesophageal sphincter [3]. The stomach is responsible for three major functions. The first and most important task is to store the food that we consume. Furthermore, by releasing gastric juices, it breaks down, and digests the food consumed. Finally, the digested food is transferred to the small intestine. Stomach infection, also known as gastric cancer, is one of the fourth leading causes of cancer deaths worldwide, with an average survival ratio of less than 12 months for advanced cancer stages. It is a polygenic disease in which multiple factors, both genetic and environmental, play an important role in its development. Every year, an estimated 1 million new cases are identified worldwide [4]. Gastric cancer cannot be avoided; if the warning signs are ignored or not treated in the early stages, it can develop into a tumor. Most treatments for it include chemotherapy, surgery, and radiation, among other things. Men have a higher risk of developing gastric cancer compared to women. Aside from environmental and genetic factors, *Helicobacter pylori* (*H. pylori*) are a type of bacteria that enters the body and lives in the digestive region, resulting in ulcers and gastritis over time. Gastritis is an inflammation of the stomach lining caused by the same bacterium that causes stomach ulcers. Gastritis symptoms include hiccups, heartburn, and gagging [5].

The most common gastric disease is colorectal cancer, which is the third most common cancer and affects both men and women equally [6]. Colorectal cancer, also known as bowel cancer, is characterized by three infections: bleeding, ulcer, and polyp. According to statistics, approximately 1.6 million people suffer from aching bowel infections, and approximately 200,000 new cases of colorectal cancer are diagnosed each year [7]. These gastrointestinal infections (GIT) can be cured if detected and treated early. The diagnosis of GIT, particularly at an early stage with improved accuracy, has emerged as the focus of current research [8].

Endoscopy is an effective method for identifying gastric cancer and is one of several diagnostic measures that could be used to detect it. The accuracy of endoscopy biopsy removal is approximately 98%; however, this procedure is time consuming, costly, requires trained medical specialists, and if not performed correctly, can result in multiple complications in the patient [9]. Another treatment for gastric cancer diagnosis is gastroscopy and laparoscopy, in which a camera is inserted into a patient's esophagus and stomach and an analysis is made using double contrast imaging of the stomach. The majority of gastrointestinal infections can lead to colorectal cancer, which manifests as short bowel syndrome and hemorrhoids. Wireless capsule endoscopy (WEC) is a painless method of identifying infections such as ulcers and polyps in the gastrointestinal tract (GIT) areas of patients with limited access without the need for surgery [10]. In this procedure, the patient is asked to swallow a camera in the shape of a capsule with dimensions of $11 \times 30 \text{ mm}^2$ and no exterior wiring, which captures images and sends them to a data recorder via an RF transmitter as it travels to the gastrointestinal tract (GIT). The video frames captured have a resolution of 255×255 pixels and are compressed in jpeg format. This entire process takes an average of 120 min to complete, whereas in normal circumstances it takes approximately 2 h [11].

An average of 60,000 images of a single patient are manually analyzed, which is sometimes impossible for even a qualified doctor to carry out [12]. Though analysis of all image frames is not required, in order to obtain accurate results, the doctor evaluates all captured frames, which lead to a major disagreement. To address this, specialists have used a variety of computer aided diagnosis (CAD) techniques to help identify gastrointestinal tract (GIT) infections using wireless capsule endoscopy (WCE) images. However, selecting image frames containing abnormalities is a difficult task due to the similarity of signs containing texture, shape, and color [13], which makes accurately categorizing the nature

of infection difficult. To bridge this gap in computer aided diagnosis (CAD) systems, various researchers used computational pathology techniques and algorithms from image processing (IP), artificial intelligence (AI), machine learning (ML), and deep learning (DL), with promising results [14]. Deep learning applications have yielded promising results in the classification of cancer, cell segmentation, and predicting the outcome of several gastrointestinal infections in recent years (GI) [15].

A significant amount of work has been done in the field of medical imaging to assist general practitioners in the accurate identification and classification of human diseases such as breast tumor, lung cancer, brain tumor [16], and infections linked to the stomach such as gastrointestinal tract (GIT) using wireless capsule endoscopy (WCE) images [17]. The stomach is an important organ in the human body. Gastritis, ulcers, polyps, and bleeding are all examples of harmful stomach diseases. Khan et al. [18] presented a deep learning architecture for the detection and classification of gastrointestinal tract (GIT) anomalies. The entire procedure is evaluated using two datasets: Private and KVASIR. When compared to the existing techniques, the proposed technique proves to be more effective. The accuracy achieved for the private dataset is 99.8%, while the accuracy for the KVASIR dataset is 86.4%. However, the study also addresses the disadvantage of varying texture features for disease classification. A feature optimized DL technique for the classification of gastrointestinal infections using wireless capsule endoscopy (WCE) images is proposed in the study [19]. The proposed technique is evaluated using two publicly available datasets. The assessed results' accuracy is 99.5%, which is more effective when compared to existing methods and distinct optimal feature sets. However, the study also mentions the applied fusion method's shortcoming of being time consuming, which can be overcome in future studies by building a CNN model from the start. Khan et al. [20] presented a computerized automated system for classifying gastrointestinal tract (GIT) infections from wireless capsule endoscopy (WCE) images using a robust deep CNN feature selection method. Infected areas are segmented using the CFbLHS method before CNN features are computed. Furthermore, only the best features were chosen for the final classification. Private datasets are used to conduct the experiment, resulting in a maximum accuracy of 99.5% over a computational time period of 21.15 s. The shortcomings of manual procedures for identifying gastric infections are overcome by the use of various high-tech practices that help physicians detect gastric abnormalities using WCE images. The study [21] envisions a fully computerized deep learning feature fusion centered architecture for the classification of numerous gastric anomalies. For evaluating the results, a database for wireless capsule endoscopy (WCE) images is created, resulting in a maximum accuracy of 99.46% when compared to the existing techniques. The overall results show that the preprocessing phase of the CNN model is effective in the learning procedure, and that the fusion of optimal features improves the accuracy. However, unrelated features and redundancy were still observed. A case study was conducted to assess the pathologist's deep learning competence in diagnosing gastric infections [22]. In this study, 16 professional pathologists inferred a total of 110 whole slide images (WSI) containing 50 malignant and 60 benign tumors with and without deep learning assistance. This case study concluded that deep learning-based assistance aided in achieving maximum area curvature under ROC-AUC, higher sensitivity, and a normal analysis time span when compared to unassisted deep learning. As a result, it was determined that deep learning-based assistance was effective, accurate, and efficient in the detection of gastric tumors. Majid et al. [23] presented an automated technique for identifying and classifying stomach infections using endoscopic images. The technique is divided into four phases: feature extraction, feature fusion, feature selection, and classification. The proposed method is evaluated using a database comprised of four datasets: CVC-ClinicDB, KVASIR, and ETIS-LaribPolypDB. The evaluation results show that the features selection method performs well by refining the overall computational time period, with a maximum accuracy of 96.5%. Authors in [24] presented a technique for assessing and classifying gastric ailments using wireless capsule endoscopy (WCE) images that cover the four stages. The first phase employs HSI color modification, which is

followed by the infection segmentation phase via the saliency approach. For image fusion in the third phase, a probability method is used. Finally, traditional features are extracted and classified by machine learning methods. Al-Adhaileh et al. [25] presented a method for the detection of gastrointestinal contagions. Three deep learning based models have been employed such as AlexNet, GoogleNet, and ResNet-50 that can aid medical practitioners to concentrate on the areas which have been overlooked during diagnosis. KVASIR dataset is used for the evaluation process and achieved an accuracy of 97%.

Preprocessing of original images, segmentation of cancer region, feature extraction, and classification are all important steps in a CAD system. Preprocessing is an important step in medical image processing because it allows important information to be visualized more effectively. This step improves the segmentation process, which has a later impact on the extraction of accurate features. However, segmentation of the infected stomach region is difficult due to the change in shape and the presence of the infected region on the boundary region. Incorrect segmentation of infected regions not only reduces segmentation accuracy but also raises the error rate in the classification phase due to the extraction of irrelevant features. Feature extraction is an important step in accurate classification, but several redundant features are sometimes extracted. The main reason is that deep models are trained using raw images or inaccurately segmented images. Furthermore, recent studies show that deep models were trained on static hyperparameters, which can be improved by using a dynamic initialization approach. To address the issue of redundant features, researchers developed several feature selection methods, including genetic algorithm-based selection, ant colony-based selection, entropy-based selection, and others. The best features are selected using these techniques; however, based on the analysis of these methods, it is also discovered that some important features are also removed during the selection process. In this paper, we proposed deep saliency estimation and an optimal deep features-based framework for classifying stomach infections. The following are our major contributions:

- We proposed a hybrid sequential fusion approach for contrast enhancement. The purpose of this approach is improving the contrast of infected region in the image that further helps in better segmentation.
- A deep saliency-based infected region segmentation and localization technique is proposed.
- A fine-tuned MobileNet-V2 model is trained on localized images and hyperparameters are optimized using Bayesian Optimization. Usually, the hyperparameters were initialized in a static way.
- A hybrid whale optimization algorithm is proposed for the selection of best features.

2. Materials and Methods

A new automated framework is proposed in this work for GIT diseases detection and classification using saliency estimation and Bayesian optimization deep learning features. The proposed framework consists of several steps that include preprocessing to classification. Original images are improved in the preprocessing step that further segmented through saliency map estimation and mathematical formulation. The segmented regions are then used to train a pre-trained fine-tuned model called MobileNet-V2 using transfer learning. The fine-tuned model's hyperparameters were initialized using Bayesian optimization (BO) and extract features from average pooling layer. After that, we proposed a hybrid whale optimization algorithm for the selection of best features. Finally, the selected features are classified using an extreme learning machine classifier. Figure 1 shows the overall framework of GIT diseases segmentation and classification. The detail of each step is provided in the below subsections.

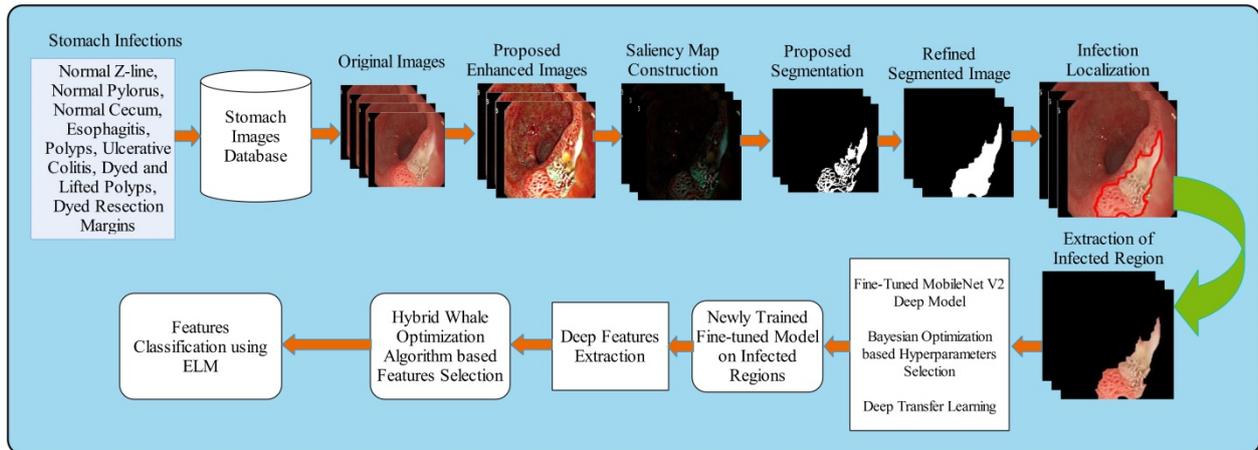


Figure 1. Proposed saliency estimation and deep learning framework for GIT diseases classification.

2.1. Proposed Contrast Enhancement

Contrast enhancement is an important step in the domain of image processing based on some important properties such as low contrast improvement and noise estimation. In this work, we proposed a hybrid approach for contrast enhancement that based on some fusion of sequential steps. Consider, we have a database denoted by Δ having several images of different classes. Let $\phi(i, j)$ is an input image of dimension $256 \times 256 \times 3$ and $\tilde{\phi}$ is a resultant image. For this, initially a CNN-based denoising network is employed and the bubble regions in the image is removed. This process is defined as follows:

$$I_1 = \frac{c_i}{\sum_{i=1}^3 (\tilde{C})}, \tilde{C} \in (c_1, c_2, c_3) \tag{1}$$

$$f_1 = \phi(c_1, \tilde{\phi}) \tag{2}$$

$$f_2 = \phi(c_2, \tilde{\phi}) \tag{3}$$

$$f_3 = \phi(c_3, \tilde{\phi}) \tag{4}$$

$$f_4 = \text{cat}(3, f_1, f_2, f_3) \tag{5}$$

where $\tilde{\phi}$ is pre-trained network [26], f_4 is a denoise image, and c_i denotes the extracted three channels, respectively. In the next step, top-bottom hat filtering is applied on f_4 to improve the local and global information as follows:

$$\phi_{top} = \text{Top}(f_4, s) \tag{6}$$

$$\phi_{bot} = \text{Bot}(f_4, s) \tag{7}$$

$$\phi_{fused} = (\phi_{top} + \phi_{bot}) - \phi(i, j) \tag{8}$$

This resultant image has some brightness effects that are resolved through haze removal existing approach [26]. The output can be written as follows:

$$\phi_{hz} = \text{HZ}(\phi_{fused}) \tag{9}$$

For highlighting the important information in the image, we performed multiplication operation that was finally fused with ϕ_{hz} image for final output.

$$\phi_{ml} = \phi_{hz} * \phi_{nw} \tag{10}$$

$$\phi_{nw} = \phi_{fused} * \phi_{top} \tag{11}$$

$$\tilde{\phi} = \sum(\phi_{ml} + \phi_{hz}) \tag{12}$$

The resultant $\tilde{\phi}$ is visually shown in Figure 2.

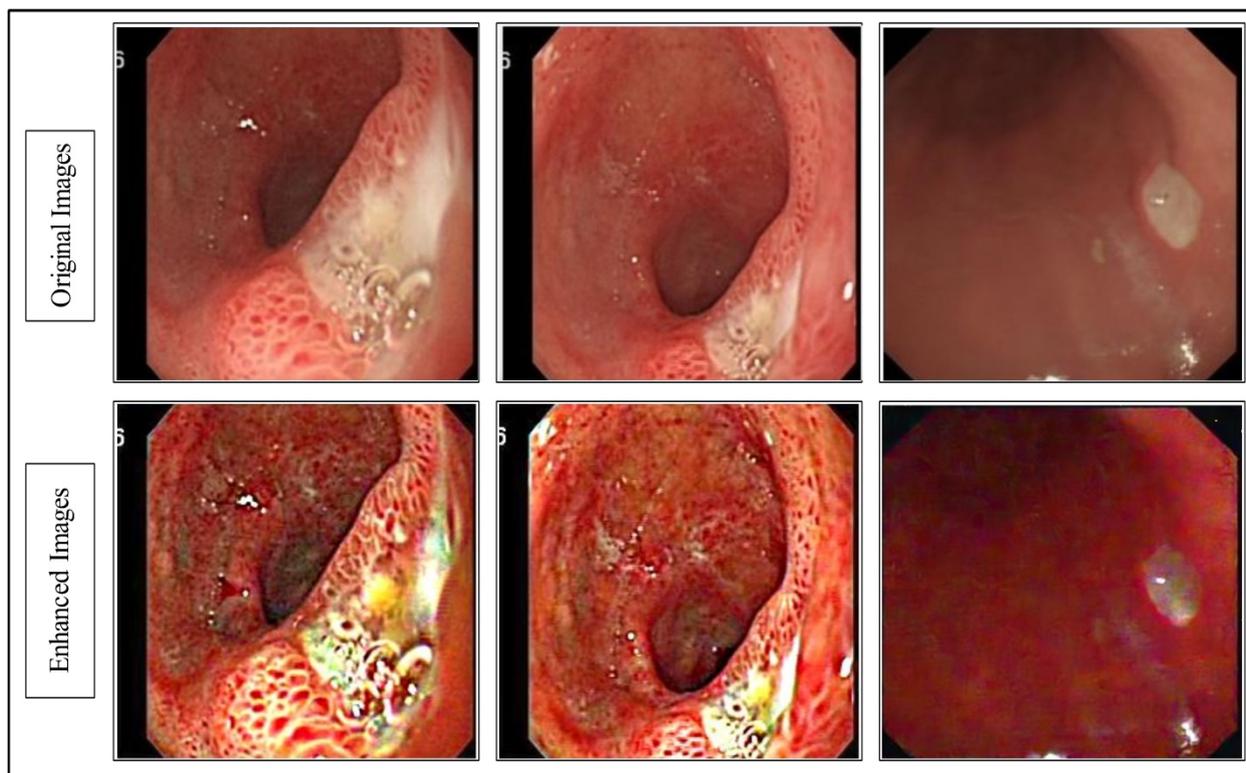


Figure 2. Proposed hybrid contrast enhancement image effects.

2.2. Proposed Saliency Map Based Segmentation

Saliency-based segmentation of an object is a new domain of research in the imaging and medical domain. In medical domain, the saliency-based segmentation of infected region is a new challenge and can perform better than the traditional techniques such as thresholding, clustering etc. In this work, we proposed a deep saliency based segmentation method called deep saliency map of infected region. The proposed technique works in four serial steps. In the first step, we design a 14 layered CNN architecture and trained on enhanced images. In the second step, weights of second convolutional layer were visualized and merged into a single image. In the third step, thresholding-based convert image into a binary form that refined in the last step by employing some mathematical operations such as closing and filling.

The newly designed 14 layered CNN architecture includes three convolutional layers having filter size 3×3 and stride 2×2 , 2 max pooling layers having filter size 2×2 and stride 2×2 , 3 batch normalization layers, 3 activation layers (ReLU), one average pooling layer, one fully connected layer, and last one is Softmax layer (can be seen in Figure 3). This designed architecture is trained on enhanced images, whereas the learning rate is 0.05, epochs are 100, mini batch size is 32, momentum value is 0.6, dropout factor is 0.5, and stochastic gradient descent (SGD) is employed as optimizer. After the training, we visualized the weights of second convolutional layer and merge all those sub-images which have clear pattern into a single image.

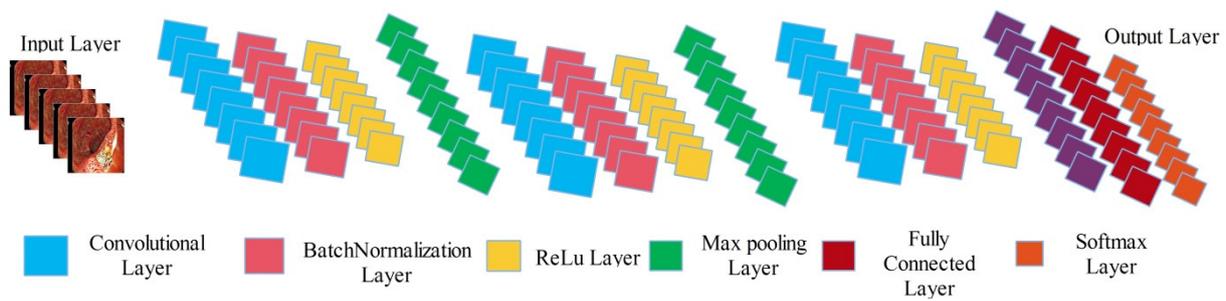


Figure 3. Proposed 14-layered CNN architecture for saliency map construction.

The resultant image is further refined by Equation (13) to improve the visual map of infected region as follows:

$$\tilde{O}_{map} = \left[O_{map} + \sum_{i=1} c_1(\tilde{\phi}) \right] \quad (13)$$

Here, \tilde{O}_{map} denotes the refined infection saliency map image, as illustrated in Figure 4, O_{map} denotes the original saliency mapped image, $\tilde{\phi}$ is proposed enhanced image. After that, the resultant refined saliency mapped image is converted into a binary form by employing the following equation:

$$\phi_{binary} = \begin{cases} 1 & \text{if } \tilde{O}_{map} \geq t \\ 0 & \text{Otherwise} \end{cases}, \text{ where } t = \text{Avg Value} \quad (14)$$

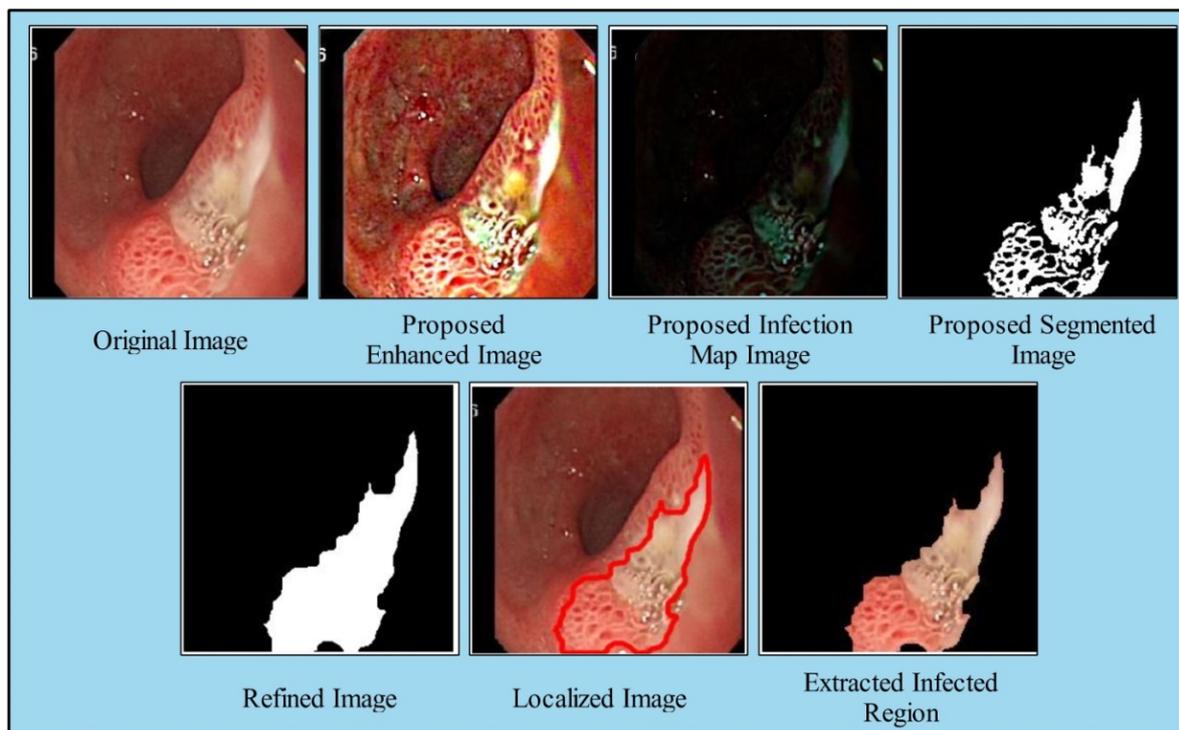


Figure 4. Proposed saliency based infected region detection.

The proposed binary image ϕ_{binary} as shown in Figure 4 is further refined by mathematical operations such as filling and closing. The effects after the closing and filling operations are shown in Figure 4. The final refined binary image is further localized through active contour approach and infected regions are later utilized for the training of a deep model. A few sample localized images have been illustrated in Figure 5.

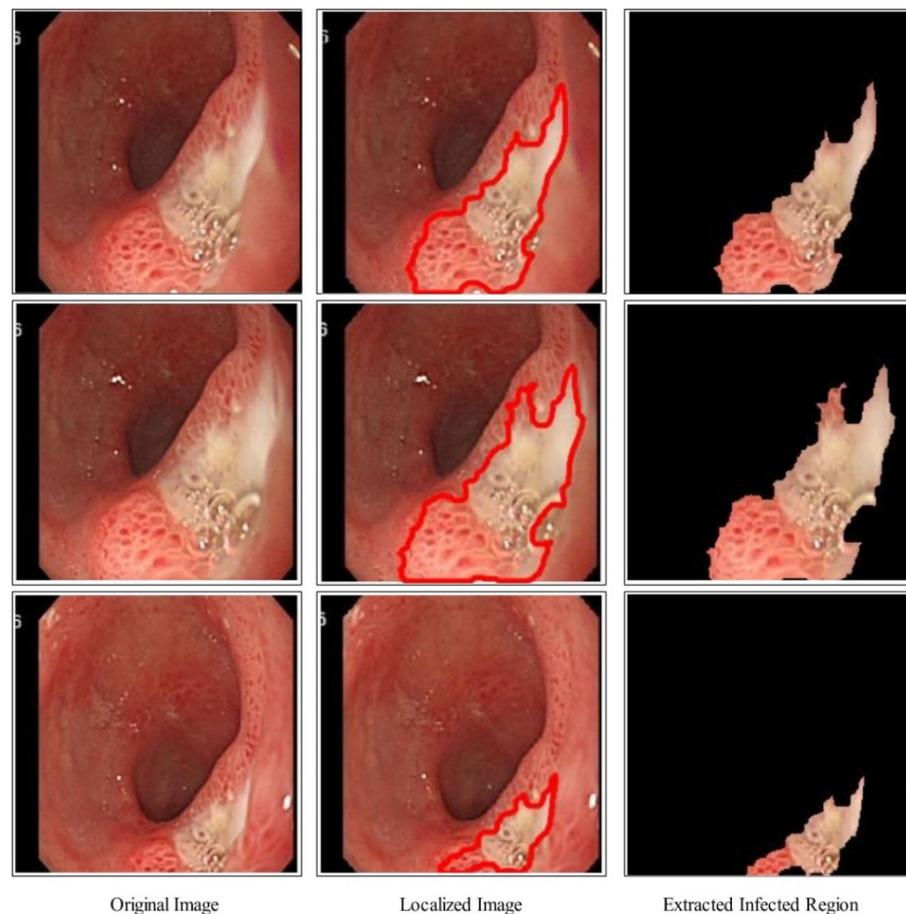


Figure 5. Proposed ulcer localization using deep saliency map.

2.3. Deep Learning Features

In this work, we extracted deep learning features for the infected region classification. For deep features extraction, a pre-trained MobileNet-V2 [27] deep model is employed. This model contains around 154 layers including convolutional layers, pooling layers, and fully connected. This model is specifically designed for the classification and general feature generation of images. This model uses 3.4 M constraints that are fewer than other generally preferred models of convolutional neural network (CNN) for example AlexNet uses 61 M, ResNet50 uses 23 M constraints, VGGNets uses 138 M, and GoogleNet uses 7 M constraints. Originally, the output layer of this model consists of 1000 object classes. We fine-tuned this model and replaced the last layers with new layers according to the output classes of proposed framework such as classes in selected datasets. The training is performed through deep transfer learning, as shown in Figure 6. In this figure, it is noted that the knowledge is transferred to the fine-tuned model. In the training process, usually static hyperparameters have been initialized but in this work, we employed Bayesian optimization for the initialization of hyperparameters. The following hyperparameters are initialized through BO-learning rate (0.0001–1), momentum (0.6–0.8), and L2-Regularization ($1e^{-10}$, $1e^{-2}$). After that, a newly trained model is obtained that includes GIT disease classes.

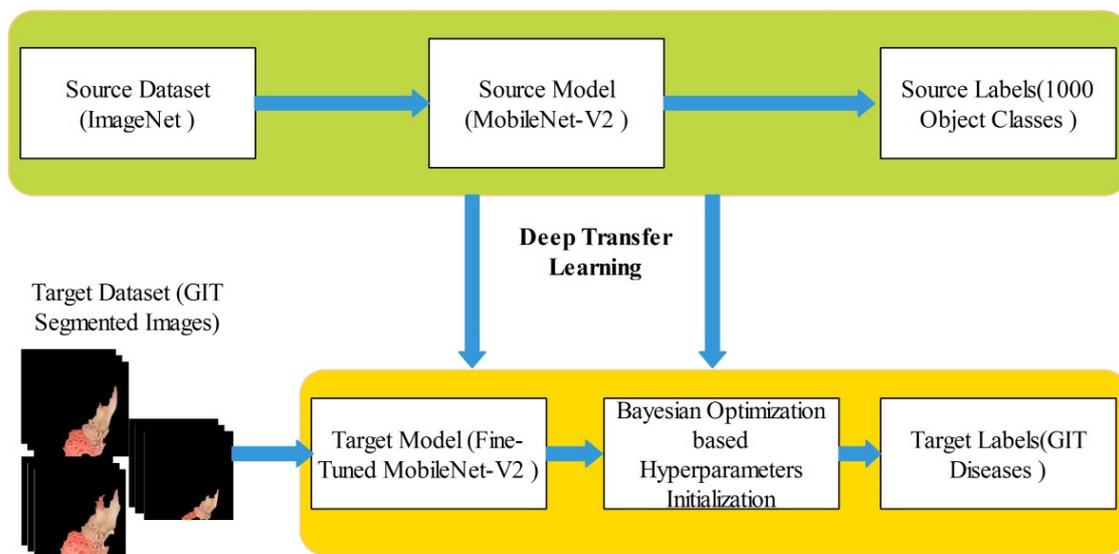


Figure 6. Deep TL and BO-based training of fine-tuned model on segmented GIT disease images.

2.4. Bayesian Optimization

For the purpose to optimize the expensive noisy tasks of black box, Bayesian optimization [28] is employed in this work. A substantial improvement with notional outcomes has been stated by the current revolution in Bayesian optimization. The strategy of BO relies over the assembly of heuristic model on which several objective tasks are disseminated to the objective of concern from the input space.

$$\mathcal{D} = \{(a_x, b_x)\}_{x=1}^{\mathcal{N}} \tag{15}$$

where \mathcal{N} is the over-all sum of annotations of the input objective sets. An acquisition function is applied to the variance and mean which is an interchange between exploitation and exploration. In order to decide the next input for assessment, a proxy optimization is executed by continuing the Bayesian optimization. Functions used in BO are disseminated using GPs due to flexibility, ambiguity, and systematic properties. Hence to overcome minimization complications, BO is utilized and defined as follows:

$$y^* = \underset{y \in \mathcal{X}}{\operatorname{argmin}} g(y) \tag{16}$$

In the above stated equation, \mathcal{X} denotes the dense subset of $\mathbb{R}^{\mathcal{K}}$. For the meta-parameters of substitute model, let borderline analytical variance of heuristic model be $\sigma^2(y, \Theta) = \Sigma(y, y; \Theta)$ and $\mu(y; \mathcal{D}, \Theta)$ that represents the analytical mean and it is described as follows:

$$\gamma(y) = \frac{g(y_{BEST}) - \mu(y; \mathcal{D}, \Theta)}{\sigma(y; \mathcal{D}, \Theta)} \tag{17}$$

where $g(y_{BEST})$ represents the minimum perceived value. The estimated enhanced benchmark is shown as:

$$\alpha_{FI}(y; \mathcal{D}, \Theta) = \sigma(y; \mathcal{D}, \Theta) \cdot [\gamma(y)\Phi(\gamma(y)) + \mathcal{M}(\gamma(y); 0, 1)] \tag{18}$$

Here, the symbol Φ represents the cumulative function and $\mathcal{M}(\cdot; 0, 1)$ signifies the density of normal standard. After the training of this model on infected cropped regions, we obtained a newly trained model that is later utilized for the features extraction. The features are extracted from the average pooling layer of dimension $N \times 1280$. As illustrated in Figure 1, the extracted features are optimized through a hybrid whale optimization algorithm.

2.5. Proposed Feature Selection Algorithm

Feature selection is a vital preprocessing phase in machine learning. The collection of huge quantity of data and information may result into noise and irrelevant data which in turn impacts the accuracy of the system. Feature selection is a significant approach that selects the best features from the original feature matrix. The purpose of this step is to ignore the redundant information and minimize the computational time. In this work, we proposed a hybrid optimization algorithm based on Whale optimization [29] and Harris Hawks optimization [30]. The mean deviation formulation is applied after both algorithms to remove redundancy among features. Mathematically, the Whale optimization algorithm is defined as follows:

$$\vec{X} = \vec{U} \cdot \vec{Y}(v) - \vec{Y}(v) \tag{19}$$

$$\vec{Y}(v+1) = \vec{Y}(v) - \vec{W} \cdot \vec{X} \tag{20}$$

In the above stated equations, \vec{Y}^* implies the top most attained result, whereas v signifies the number of repetitions. The symbols \vec{W} and \vec{U} are constants and defined as follows:

$$\vec{W} = 2 \cdot \vec{c} \cdot \vec{d} - \vec{c} \tag{21}$$

$$\vec{U} = 2 \cdot \vec{d} \tag{22}$$

where \vec{d} implies random vector among [0, 1], \vec{c} is also a random vector which is meant to regulate the overall conjunction method and it declines linearly from 2 to 0 during the repetitions. The value of \vec{c} can be calculated using the below equation.

$$\vec{c} = 2 - v \frac{2}{V} \tag{23}$$

where v implies the current repetition and V implies the large number of repetitions. The transition course between exploration and exploitation is shown by vector \vec{W} . The exploration agents will keep on exploring the space when the absolute value of vector \vec{W} is greater than one. Moreover, when absolute value of vector \vec{W} is less than one i.e., $|\vec{W}| < 1$, it will result in exploitation of the solution. The two major approaches included in this algorithm are known as encircling technique and spiral shaped technique. The encircling process can be attained by minimizing the value of W , whereas the second process update the distance between the current search agent which is attained at point \vec{Y}^* and the exploration agent. Mathematically, it is formulated as follows:

$$Y(v+1) = X.e^{al} \cdot \cos(2\pi l) + Y^*(v) \tag{24}$$

A probability value of 0.5 is set for the purpose of signifying the explorative behavior for further execution. This procedure is stated below as:

$$\vec{Y}(v+1) = \begin{cases} X^l \cdot e^{al} \cdot \cos(2\pi l) + Y^*(v), & P \geq 0.5 \\ \vec{Y}^*(v) - \vec{W} \cdot \vec{X}, & P \leq 0.5 \end{cases} \tag{25}$$

In Equation (25), the probability value of 0.5 is selected for the selection of final features but after the analysis of this formulation on different probability values, we observed that the static value is not a good choice; therefore, we modified this equation by employing a median value of selected features instead of 0.5. Hence, the above equation can be written as:

$$\vec{Y}(v+1) = \begin{cases} X^l \cdot e^{al} \cdot \cos(2\pi l) + Y^*(v), & P \geq MD \\ \vec{Y}^*(v) - \vec{W} \cdot \vec{X}, & P \leq MD \end{cases} \tag{26}$$

$$MD = \begin{cases} X^* \left[\frac{n+1}{2} \right] & \text{if } n \text{ is odd} \\ \frac{X^* \left[\frac{n}{2} \right] + X^* \left[\frac{n}{2} + 1 \right]}{2} & \text{if } n \text{ is even} \end{cases} \quad (27)$$

where MD denotes median value and X^* denotes the current iteration features utilized for the final selection through Equation (25). Through Equation (25), the dimension of selected features is $N \times 728$. These features are passed to Harris Hawks optimization for one extra step refinement. Harris Hawks approach is replicated by the system in two states. During first state, this approach settles at several random localities nearby their family or cluster. In the second state, Harris Hawks can live around other supporters of household or cluster, $Q \geq MD$ is for the first scenario and for later state $Q < MD$.

$$Y(v + 1) = \begin{cases} Y_{RAND} - d_1 |Y_{rand}(v) - 2d_2 Y(v)|, & Q \geq Mean \\ (Y_{RABBIT}(v) - Y_M(v) - d_3(lb + d_4(ub - lb))), & Q < Mean \end{cases} \quad (28)$$

In the above stated question, $Y(v + 1)$ represents the location of succeeding hawk where as $Y_{rand}(v)$ represents the current point of the selected hawk from prevailing population. The mean value-based selection is performed instead of 0.5 static values. More information can be found here [30]. The ELM classifier is selected as a fitness function and fitness value is computed based on the error rate. This process was continued for initialized iterations such as 200 in this work. At the end, we obtained a final selected feature vector of dimension $N \times 624$ that was finally classified using extreme learning machine (ELM) classifier.

3. Results

The proposed framework experimental process is conducted in this section in the form of numerical values and plots. Three publicly available datasets have been employed for the experimental process such as Kvasir 1 [31], Kvasir 2 [32], and CUI Wah [33]. The Kvasir datasets consists of eight different classes as illustrated in Figure 7. In the Kvasir V1, each class includes 500 images (total 4000 images), whereas the Kvasir V2 dataset includes 1000 images in each class (total 8000 images). The 50:50 approach was opted for training and testing of deep models. The cross validation opted for value 10. Several classifiers have been utilized for the comparison of ELM classifier accuracy such as fine tree, quadratic SVM (Q-SVM), weighted KNN (W-KNN), and bi-layered neural network (Bi-Layer NN). The performance of each classifier is analyzed through accuracy and computational time. The entire framework is simulated on MATLAB2022a using a desktop computer with 16 GB of Ram and 8 GB graphics card.

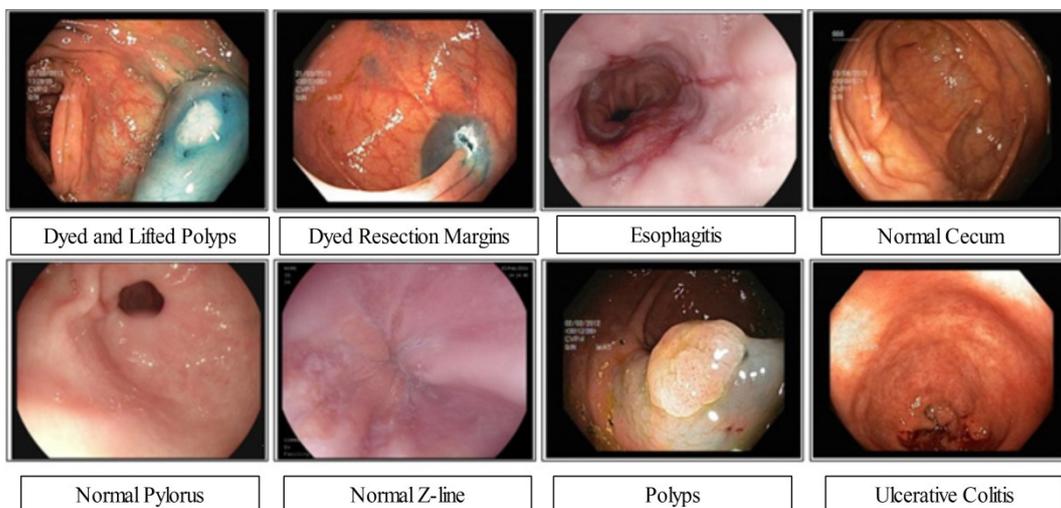


Figure 7. Samples GIT disease images collected from Kvasir-V2 dataset [32].

3.1. CUI WCE Dataset Results

The classification results of proposed framework on CUI WCE dataset has been presented in Table 1. This table presented the results in four different experiments. In the first experiment (Org-MobV2), the fine-tuned MobileNet-V2 deep model is trained on original dataset and extract features that are later utilized for the classification. In the second experiment (Enh-MobV2), enhanced WCE images have been utilized and passed to fine-tuned model for training that were later utilized for features extraction and classification. In the third experiment (Seg-MobV2), localized infected images have been fed to fine-tune MobileNet-V2 for training that were later utilized for features extraction and classification. In the last experiment (Proposed), the entire proposed framework is utilized and classification was performed.

Table 1. Proposed framework classification results on CUI WCE dataset.

Classifier	Org-MobV2	Enh-MobV2	Seg-MobV2	Proposed	Accuracy (%)	Time (s)
ELM	✓				95.24	116.5424
		✓			96.94	110.2010
			✓		97.39	103.1152
				✓	99.61	69.5442
Fine Tree	✓				90.56	131.5032
		✓			91.04	122.5629
			✓		92.39	116.0424
				✓	94.84	74.5006
Q-SVM	✓				92.10	147.0302
		✓			94.56	141.5624
			✓		94.90	136.9206
				✓	96.36	89.1432
W-KNN	✓				91.04	134.1142
		✓			92.50	129.5260
			✓		94.14	121.1124
				✓	94.80	80.5142
Bi-Layer NN	✓				92.52	119.4504
		✓			94.06	113.1492
			✓		95.84	106.5824
				✓	98.16	71.0062
XGBOOST	✓				92.26	126.0047
		✓			92.95	110.8072
			✓		93.60	106.4921
				✓	94.00	70.6075

In Table 1, the maximum obtained accuracy of first experiment called Org-Mobv2 is 95.24% on ELM classifier whereas the minimum computational time is 116.5424 s. Moreover, the lowest accuracy of this experiment is 90.56% for fine-tree classifier. The best accuracy of the second experiment (Enh-MobV2) is 96.94% on ELM, whereas the lowest accuracy is 91.06% on fine-tree. The minimum computational time of this experiment is 110.2010 s, whereas the highest noted time is 141.5624 s. In the third experiment (Seg-MobV2), the maximum obtained accuracy is 97.39% that improved compared to the first two experiments. Moreover, it is also noted that computational time is little decreased after this

experiment. For the proposed framework, the maximum obtained accuracy is 99.61% that is significantly improved compared to the first three experiments. Moreover, the computational time is significantly reduced after this step due to the use of optimization algorithm. In addition, Figure 8 illustrated a confusion matrix of ELM classifier that can be utilized for the verification of proposed accuracy. Hence, overall proposed framework obtained improved accuracy and consumes less time compared to the previous experiments.

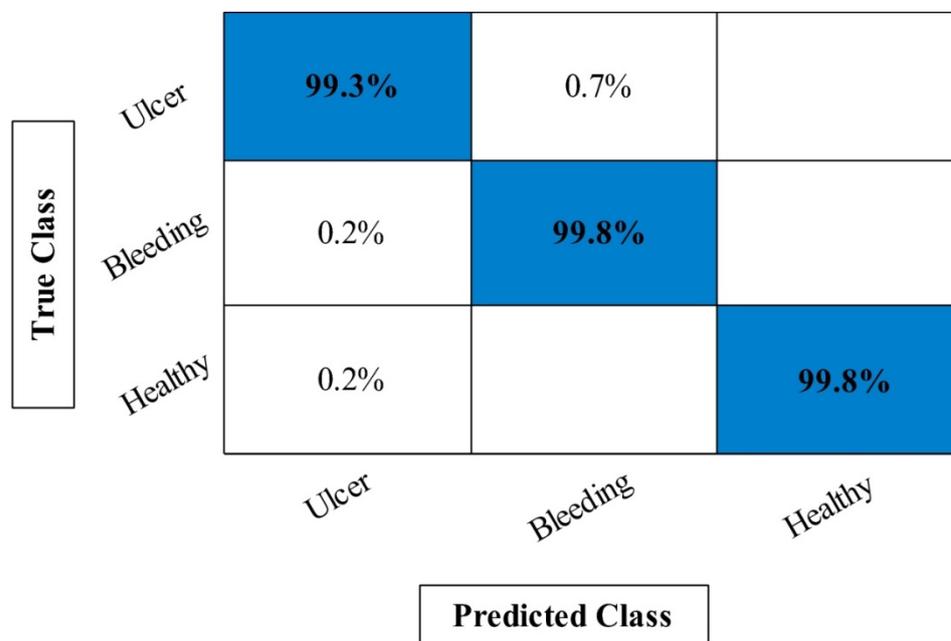


Figure 8. Confusion matrix of proposed framework for ELM classifier on CUI WCE dataset.

3.2. KVASIR V1 Dataset Results

The classification results of Kvasir V1 dataset has been presented in Table 2. Similar to results of CUI WCE dataset as presented in Section 3.1, the four experiments have been performed on Kvasir V1 dataset. In the first experiment, the maximum achieved accuracy is 95.24% whereas the minimum noted computational time is 92.1124 s. In the second experiment, the best obtained accuracy is 95.80% on ELM classifier. Compared to the first experiment, the current accuracy is little improved and also the computational time is reduced from 92.1124 s to 90.3645 s. The best obtained accuracy of third experiment is 96.14% on ELM classifier that is improved compared to the first two experiments. In addition, the computational time is reduced from 92.1123 s to 84.1046 s. The proposed framework obtained maximum accuracy of 98.20%, which is improved compared to the previous three experiments. Moreover, the lowest accuracy of this experiment is 93.02% W-KNN classifier. Moreover, the computational time of proposed framework is significantly reduced. Figure 9 shows the confusion matrix of ELM classifier for the proposed framework that can be utilized for the verification of the obtained accuracy of 98.20%. Based on the results, we can determine that the proposed framework performed better in both accuracy and computational time.

Table 2. Proposed framework classification results on Kvasir V1 dataset.

Classifier	Org-MobV2	Enh-MobV2	Seg-MobV2	Proposed	Accuracy (%)	Time (s)
ELM	✓				95.24	92.1124
		✓			95.80	90.3645
			✓		96.14	84.1046
				✓	98.20	52.5046

Table 2. Cont.

Classifier	Org-MobV2	Enh-MobV2	Seg-MobV2	Proposed	Accuracy (%)	Time (s)
Fine Tree	✓				93.04	98.2446
		✓			93.64	95.3604
			✓		94.03	87.2942
Q-SVM				✓	96.46	63.1142
	✓				91.62	97.3046
		✓			92.14	94.2946
			✓		93.42	87.5042
W-KNN				✓	95.14	71.0246
	✓				90.54	99.6214
		✓			91.32	97.5429
			✓		91.98	90.1120
Bi-Layer NN				✓	93.02	61.1129
	✓				94.34	94.1126
		✓			94.96	91.6624
			✓		95.70	86.2404
XGBOOST				✓	97.10	55.5246
	✓				93.58	104.5093
		✓			94.02	101.9226
			✓		95.16	93.5521
			✓	95.90	69.4050	

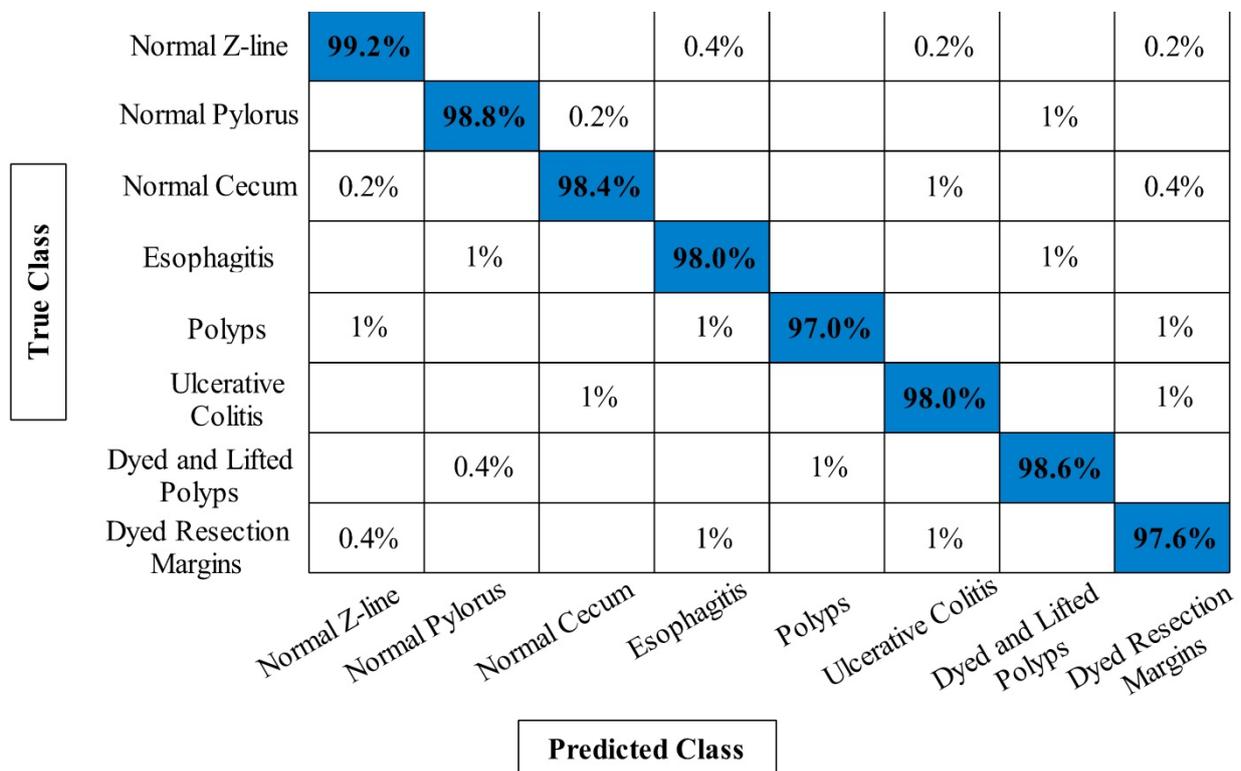


Figure 9. Confusion matrix of proposed framework for ELM classifier on Kvasir V1 dataset.

3.3. KVASIR V2 Classification Results

The classification results of proposed framework on Kvasir V2 dataset have been presented in Table 3. Similar to Tables 1 and 2, four experiments have been performed for this dataset to analyze the entire framework. For experiment 1, 94.20% accuracy is obtained on ELM classifier that is better compared to the other classifiers listed in this table. The lowest obtained accuracy of this experiment is 92.14% on fine-tree classifier. In addition, the computational time is also noted for each classifier and minimum time is 191.5246 s. In the second experiment, 95.18% best accuracy is obtained on ELM that is improved compared to the first experiment. This shows the importance of contrast enhancement step. In the third experiment, 96.76% accuracy is achieved that is better compared to the first two experiments. Moreover, this experiment consumes less time i.e., 173.1142 s, compared to the other listed experiments. Finally, proposed framework is employed and obtained an accuracy of 98.02% that is significantly improved compared to the other experiments. Moreover, the computational time of this experiment is far better than the other experiments on all classifiers. Figure 10 illustrates the confusion matrix of Kvasir V2 dataset on ELM classifier that can be utilized for the verification of the proposed framework accuracy.

Table 3. Proposed framework classification results on Kvasir V2 dataset.

Classifier	Org-MobV2	Enh-MobV2	Seg-MobV2	Proposed	Accuracy (%)	Time (s)
ELM	✓				94.20	191.5246
		✓			95.18	186.5509
			✓		96.76	173.1142
				✓	98.02	102.5026
Fine Tree	✓				92.14	205.0426
		✓			93.24	201.0020
			✓		93.60	191.5462
				✓	95.46	120.2500
Q-SVM	✓				92.92	226.2042
		✓			93.60	216.1120
			✓		94.10	204.0526
				✓	96.40	140.0329
W-KNN	✓				92.62	207.1246
		✓			92.94	203.0204
			✓		93.56	195.5509
				✓	95.84	120.5426
Bi-Layer NN	✓				93.04	195.0694
		✓			94.16	191.1124
			✓		94.86	185.0329
				✓	96.94	110.0046
XGBOOST	✓				93.00	220.0945
		✓			93.84	211.2572
			✓		94.10	182.9443
				✓	94.90	136.0790

True Class	Normal Z-line	98.6%		1%			0.4%		
	Normal Pylorus		98.5%	1%		0.5%			
	Normal Cecum		0.4%	98%		1%	0.6%		
	Esophagitis		1%		97.5%	0.5%	1%		
	Polyps		0.4%		0.8%	97.8%	1%		
	Ulcerative Colitis			0.8%			98.2%	1%	
	Dyed and Lifted Polyps	0.5%			1%		0.4%	98.1%	
	Dyed Resection Margins		1%		1%			0.5%	97.5%
		Normal Z-line	Normal Pylorus	Normal Cecum	Esophagitis	Polyps	Ulcerative Colitis	Dyed and Lifted Polyps	Dyed Resection Margins
		Predicted Class							

Figure 10. Confusion matrix of proposed framework for ELM classifier on Kvasir V2 dataset.

3.4. Discussion and Comparison

A brief discussion of proposed framework is also presented in terms of some visual facts and quantitative values. Figure 5 shows the qualitative results of infected lesion localization using proposed saliency based segmentation. From this figure, it is clearly illustrated that the proposed framework correctly segmented both larger and smaller infected regions. The quantitative results of all selected datasets have been presented under Tables 1–3 and confusion matrices in Figures 8–10. From these figures, it is observed that the accuracy is improved by employing the proposed algorithm compared to the other mentioned methods such as Org-MobV2, Enh-MobV2, and Seg-MobV2. We also performed a *t*-test [34,35] and presented a hypothesis in which we assume that the accuracy will not degrade after employing the proposed algorithm ($h = 1$) and accuracy will degrade after employing the proposed algorithm ($h = 0$). We consider only ELM classifier values for all three selected datasets. For CUI WCE dataset, the performance results for ELM classifier were highly significant such as $p = 1.74 \times 10^{(-6)}$ for *t*-test. For Kvasir V1 dataset, the *p* value is computed by *t*-test and shows that the performance of ELM classifier is significant such as $p = 6.63 \times 10^{(-7)}$. Similarly, the *t*-test is performed for Kvasir V2 dataset and obtained *p* value of $p = 1.50 \times 10^{(-6)}$ that shows the significance performance of ELM classifier for proposed framework.

Moreover, a comparison is also conducted of proposed framework with some other deep neural nets such as VGG16, VGG19, AlexNet, and ResNet50. A few quantitative facts are presented in Table 4. Based on this table, it clearly proves the better performance of proposed framework. Finally, we compare the proposed framework accuracy with some recent techniques such as Khan et al. [36]. In this technique, the authors used a deep learning-based framework and achieved accuracies of 99.42, 97.85, and 97.2%, for CUI WCE dataset, Kvasir V1, and Kvasir V2, respectively. The proposed framework achieved accuracies of 99.61, 98.20, and 98.02%, respectively.

Table 4. Comparison of proposed framework with other neural nets using selected datasets.

Method	Accuracy (%)		
	CUI WCE Dataset	Kvasir V1 Dataset	Kvasir V2 Dataset
VGG16 Model embedded in Figure 1 instead of MobileNetV2	95.80	94.28	95.11
VGG19 Model embedded in Figure 1 instead of MobileNetV2	96.24	95.02	95.76
AlexNet Model embedded in Figure 1 instead of MobileNetV2	95.16	94.66	94.90
ResNet50 Model embedded in Figure 1 instead of MobileNetV2	97.88	95.89	96.16
Khan et al. [36], 2022	99.42	97.85	97.85
Proposed Framework	99.61	98.20	98.02

4. Conclusions

This article proposes a new deep saliency estimation and Bayesian Optimization learning-based framework for detecting and classifying GIT diseases. The experiment was carried out on three publicly available datasets and yielded accuracies of 99.61, 98.20, and 98.02%, which were better compared to the previous method. The classification accuracy was improved by the contrast enhancement step. This step also increases the likelihood of correctly locating the infected region in the image. MobileNet-V2, a pre-trained deep model, is chosen and trained using Bayesian optimization and deep transfer learning. The benefit of this step was improved hyperparameter initialization, which was used for fine-tuning model training. In addition, we proposed a hybrid optimization algorithm for selecting the best features. This algorithm selects the best features to improve classification accuracy while decreasing computational time. The contrast enhancement and hyperparameter optimization were the work's strengths. Furthermore, feature optimization reduced irrelevant information. As a future work, the contrast enhanced images will be passed to CNN models such as UNET and MASK RCNN for infected lesion segmentation. Moreover, weights of CNN models will be optimized through evolutionary optimization techniques.

Author Contributions: Conceptualization, M.A.K., N.S., W.Z.K., M.A., U.T. and B.C.; data curation, M.A.K., N.S., U.T. and M.H.Z.; formal analysis, M.A.K., N.S., W.Z.K., M.H.Z., Y.J.K. and B.C.; funding acquisition, Y.J.K. and B.C.; investigation, N.S., W.Z.K., M.A., U.T. and Y.J.K.; methodology, M.A.K., N.S., M.H.Z. and Y.J.K.; project administration, W.Z.K., M.A., U.T. and B.C.; resources, M.A. and U.T.; software, M.A.K., M.H.Z., Y.J.K. and B.C.; supervision, W.Z.K.; visualization, M.H.Z. and Y.J.K.; writing—original draft, M.A.K. and Y.J.K.; writing—review and editing, M.A., U.T. and B.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by “Human Resources Program in Energy Technology” of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resources from the Ministry of Trade, Industry & Energy, Republic of Korea. (No. 20204010600090).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this work are publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jackson, K.J.; Emmons, K.R.; Nickitas, D.M. Role of Primary Care in Detection of Subsequent Primary Cancers. *J. Nurse Pract.* **2022**, *18*, 478–482. [[CrossRef](#)]
2. Bhardwaj, P.; Bhandari, G.; Kumar, Y.; Gupta, S. An Investigational Approach for the Prediction of Gastric Cancer Using Artificial Intelligence Techniques: A Systematic Review. *Arch. Comput. Methods Eng.* **2022**, *29*, 4379–4400. [[CrossRef](#)]

3. Ayyaz, M.S.; Lali, M.I.U.; Hussain, M.; Rauf, H.T.; Alouffi, B.; Alyami, H.; Wasti, S. Hybrid deep learning model for endoscopic lesion detection and classification using endoscopy videos. *Diagnostics* **2021**, *12*, 43. [[CrossRef](#)] [[PubMed](#)]
4. Polaka, I.; Bhandari, M.P.; Mezmale, L.; Anarkulova, L.; Veliks, V.; Sivins, A.; Lescinska, A.M.; Tolmanis, I.; Vilkoite, I.; Ivanovs, I. Modular Point-of-Care Breath Analyzer and Shape Taxonomy-Based Machine Learning for Gastric Cancer Detection. *Diagnostics* **2022**, *12*, 491. [[CrossRef](#)] [[PubMed](#)]
5. Nautiyal, H.; Kazmi, I.; Kaleem, M.; Afzal, M.; Ahmad, M.M.; Zafar, A.; Kaur, R. Mechanism of action of drugs used in gastrointestinal diseases. In *How Synthetic Drugs Work*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 391–419.
6. Sinicrope, F.A. Increasing Incidence of Early-Onset Colorectal Cancer. *N. Engl. J. Med.* **2022**, *386*, 1547–1558. [[CrossRef](#)]
7. Zhao, Y.; Hu, B.; Wang, Y.; Yin, X.; Jiang, Y.; Zhu, X. Identification of gastric cancer with convolutional neural networks: A systematic review. *Multimed. Tools Appl.* **2022**, *81*, 11717–11736. [[CrossRef](#)]
8. Deb, A.; Perisetti, A.; Goyal, H.; Aloysius, M.M.; Sachdeva, S.; Dahiya, D.; Sharma, N.; Thosani, N. Gastrointestinal Endoscopy-Associated Infections: Update on an Emerging Issue. *Dig. Dis. Sci.* **2022**, *67*, 1718–1732. [[CrossRef](#)]
9. Gholami, E.; Tabbakh, S.R.K. Increasing the accuracy in the diagnosis of stomach cancer based on color and lint features of tongue. *Biomed. Signal Process. Control* **2021**, *69*, 102782. [[CrossRef](#)]
10. Alam, M.W.; Vedaiei, S.S.; Wahid, K.A. A fluorescence-based wireless capsule endoscopy system for detecting colorectal cancer. *Cancers* **2020**, *12*, 890. [[CrossRef](#)]
11. Kim, H.J.; Gong, E.J.; Bang, C.S.; Lee, J.J.; Suk, K.T.; Baik, G.H. Computer-Aided Diagnosis of Gastrointestinal Protruded Lesions Using Wireless Capsule Endoscopy: A Systematic Review and Diagnostic Test Accuracy Meta-Analysis. *J. Pers. Med.* **2022**, *12*, 644. [[CrossRef](#)]
12. Amin, M.S.; Shah, J.H.; Yasmin, M.; Ansari, G.J.; Khan, M.A.; Tariq, U.; Kim, Y.J.; Chang, B. A Two Stream Fusion Assisted Deep Learning Framework for Stomach Diseases Classification. *CMC-Comput. Mater. Contin.* **2022**, *73*, 4423–4439. [[CrossRef](#)]
13. Iizuka, O.; Kanavati, F.; Kato, K.; Rambeau, M.; Arihiro, K.; Tsuneki, M. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci. Rep.* **2020**, *10*, 1–11.
14. Son, G.; Eo, T.; An, J.; Oh, D.J.; Shin, Y.; Rha, H.; Kim, Y.J.; Lim, Y.J.; Hwang, D. Small Bowel Detection for Wireless Capsule Endoscopy Using Convolutional Neural Networks with Temporal Filtering. *Diagnostics* **2022**, *12*, 1858. [[CrossRef](#)] [[PubMed](#)]
15. Prabhu, S.; Prasad, K.; Robels-Kelly, A.; Lu, X. AI-based carcinoma detection and classification using histopathological images: A systematic review. *Comput. Biol. Med.* **2022**, *142*, 105209. [[CrossRef](#)] [[PubMed](#)]
16. Javed, K.; Khan, S.A.; Saba, T.; Habib, U.; Khan, J.A.; Abbasi, A.A. Human action recognition using fusion of multiview and deep features: An application to video surveillance. *Multimed. Tools Appl.* **2020**, *3*, 1–27.
17. Lahoura, V.; Singh, H.; Aggarwal, A.; Sharma, B.; Mohammed, M.A.; Damaševičius, R.; Kadry, S.; Cengiz, K. Cloud computing-based framework for breast cancer diagnosis using extreme learning machine. *Diagnostics* **2021**, *11*, 241. [[CrossRef](#)]
18. Naz, J.; Alhaisoni, M.; Song, O.-Y.; Tariq, U.; Kadry, S. Segmentation and classification of stomach abnormalities using deep learning. *CMC-Comput. Mater. Contin.* **2021**, *69*, 607–625. [[CrossRef](#)]
19. Majid, A.; Hussain, N.; Alhaisoni, M.; Zhang, Y.-D.; Kadry, S.; Nam, Y. Multiclass stomach diseases classification using deep learning features optimization. *Comput. Mater. Contin.* **2021**, *69*, 1–15.
20. Sharif, M.; Akram, T.; Yasmin, M.; Nayak, R.S. Stomach deformities recognition using rank-based deep features selection. *J. Med. Syst.* **2019**, *43*, 329.
21. Sarfraz, M.S.; Alhaisoni, M.; Albeshier, A.A.; Wang, S.; Ashraf, I. StomachNet: Optimal deep learning features fusion for stomach abnormalities classification. *IEEE Access* **2020**, *8*, 197969–197981.
22. Ba, W.; Wang, S.; Shang, M.; Zhang, Z.; Wu, H.; Yu, C.; Xing, R.; Wang, W.; Wang, L.; Liu, C. Assessment of deep learning assistance for the pathological diagnosis of gastric cancer. *Mod. Pathol.* **2022**, *35*, 1262–1268. [[CrossRef](#)] [[PubMed](#)]
23. Majid, A.; Yasmin, M.; Rehman, A.; Yousafzai, A.; Tariq, U. Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection. *Microsc. Res. Technol.* **2020**, *83*, 562–576. [[CrossRef](#)] [[PubMed](#)]
24. Rashid, M.; Sharif, M.; Javed, K.; Akram, T. Classification of gastrointestinal diseases of stomach from WCE using improved saliency-based method and discriminant features selection. *Multimed. Tools Appl.* **2019**, *78*, 27743–27770.
25. Hmoud Al-Adhaileh, M.; Mohammed Senan, E.; Alsaade, W.; Aldhyani, T.H.; Alsharif, N.; Abdullah Alqarni, A.; Uddin, M.I.; Alzahrani, M.Y.; Alzain, E.D.; Jadhav, M.E. Deep Learning Algorithms for Detection and Classification of Gastrointestinal Diseases. *Complexity* **2021**, *2021*, 6170416. [[CrossRef](#)]
26. Park, D.; Park, H.; Han, D.K.; Ko, H. Single image dehazing with image entropy and information fidelity. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4037–4041.
27. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
28. Wu, J.; Chen, X.-Y.; Zhang, H.; Xiong, L.-D.; Lei, H.; Deng, S.-H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40.
29. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [[CrossRef](#)]
30. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [[CrossRef](#)]

31. Escobar, J.; Sanchez, K.; Hinojosa, C.; Arguello, H.; Castillo, S. Accurate deep learning-based gastrointestinal disease classification via transfer learning strategy. In Proceedings of the 2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), Popayan, Colombia, 15–17 September 2021; pp. 1–5.
32. Wang, W.; Yang, X.; Li, X.; Tang, J. Convolutional-capsule network for gastrointestinal endoscopy image classification. *Int. J. Intell. Syst.* **2022**, *37*, 5796–5815. [[CrossRef](#)]
33. Liaqat, A.; Shah, J.H.; Sharif, M.; Yasmin, M.; Fernandes, S.L. Automated ulcer and bleeding classification from WCE images using multiple features fusion and selection. *J. Mech. Med. Biol.* **2018**, *18*, 1850038. [[CrossRef](#)]
34. Calvo, B.; Santafé Rodrigo, G. scmamp: Statistical comparison of multiple algorithms in multiple problems. *R J.* **2016**, *8*, 1–11. [[CrossRef](#)]
35. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
36. Muhammad, K.; Wang, S.-H.; Alsubai, S.; Binbusayyis, A.; Alqahtani, A.; Majumdar, A.; Thinnukool, O. Gastrointestinal diseases recognition: A framework of deep neural network and improved moth-crow optimization with dcca fusion. *Hum.-Cent. Comput. Inf. Sci.* **2022**, *12*, 25.