Journal of Experimental & Theoretical Artificial Intelligence

ISSN: 0952-813X (Print) 1362-3079 (Online) Journal homepage: https://www.tandfonline.com/loi/teta20

Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images

Muhammad Sharif, Muhammad Attique Khan, Muhammad Rashid, Mussarat Yasmin, Farhat Afza & Urcun John Tanik

To cite this article: Muhammad Sharif, Muhammad Attique Khan, Muhammad Rashid, Mussarat Yasmin, Farhat Afza & Urcun John Tanik (2019): Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images, Journal of Experimental & Theoretical Artificial Intelligence, DOI: 10.1080/0952813X.2019.1572657

To link to this article: https://doi.org/10.1080/0952813X.2019.1572657

Published online: 01 Feb 2019.

JOURNAL OF EXPERIMENTAL & THEORETICAL ARTIFICIAL INTELLIGENCE
https://doi.org/10.1080/0952813X.2019.1572657

Taylor & Francis
Taylor & Francis Group

ARTICLE

Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images

Muhammad Sharif[a], Muhammad Attique Khan[b], Muhammad Rashid[a], Mussarat Yasmin[a], Farhat Afza[a] and Urcun John Tanik[c]

[a]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt, Pakistan; [b]Department of Computer Science and Engineering, HITEC University, Taxila, Pakistan; [c]Department of Computer Science and Information Systems, Texas A&M University-Commerce, USA

ABSTRACT

Gastrointestinal tract (GIT) infections such as ulcers, bleeding, polyps, Crohn's disease and cancer are quite familiar today worldwide. Wireless capsule endoscopy (WCE) is an efficient means of investigation of GIT diseases. However, still several challenges exist in this domain, such as lesion shape, colour, texture, size and irregularity. To deal with these problems, several computer-based methods are introduced in computer vision domain but they used only hand-crafted features which produced wrong predictions several times. In this research, a new technique is applied based on the fusion of deep convolutional (CNN) and geometric features. Initially, disease regions are extracted from given WCE images using a new approach named contrast-enhanced colour features . Geometric features are extracted from segmented disease region. Thereafter, unique VGG16 and VGG19 deep CNN features fusion are performed based on Euclidean Fisher Vector . The unique features are fused with geometric features which are later fed to conditional entropy approach for best features selection. The selected features are finally classified by K-Nearest Neighbour . A privately collected database which consists of 5500 WCE images is utilised for the evaluation of the proposed method and achieved best classification accuracy of 99.42% and precision rate of 99.51%. The classification accuracy proves the authenticity of the proposed approach.

ARTICLE HISTORY
Received 13 May 2018
Accepted 27 December 2018

KEYWORDS
Contrast stretching; ulcer segmentation; deep CNN features; features fusion; features selection

Introduction

Gastrointestinal tract (GIT) infections such as bleeding, ulcers polyps, Crohn's disease and cancer are quite familiar nowadays. In America, 135,430 new GI diseases occurred since 2017 (Siegel et al., 2017). A survey of adult population is conducted worldwide to diagnose GIT infections since 2017. Results show that 1.1 M deaths occurred due to lung cancer, 765,000 because of stomach cancer, 525,000 deaths are due to colon cancer and 385,000 deaths happened as a result of breast cancer. Likewise, 18% of adults belong to Brazil, 11% of China, 20% of EU-5, 12% of Russia and 21% of the US who are diagnosed with GI diseases. (http://www.who.int/whr/1998/media_centre/50facts/en/; https://www.statista.com/statistics/418515/adult-prevalence-of-gastrointestinal-conditions-by-country/). From all GI diseases, ulcer and bleeding are very common tumours which hit small and large bowls. About

CONTACT Mussarat Yasmin mussaratabdullah@gmail.com; MuhammadAttique Khan attique.khan440@gmail.com; Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan

1.6 million people are currently suffering bowl infections in America and about 200,000 new cases are happening every year since 2011. These diseases can be rectified if detected and diagnosed at an early stage (Fu, Zhang, Mandal, & Meng, 2014; Liaqat et al., 2018). Recently, wireless capsule endoscopy (WCE) technology has been introduced for the diagnosis of GIT diseases. WCE wirelessly captures colour images of GIT in an average duration of 8 h. These images are conveyed through a wireless device which is further utilised in clinics by physicians to make the symptomatic decisions (Yuan, Wang, Li, & Meng, 2015). But this technology has several challenges because one WCE video consists of about 57,000 frames which make the system more expensive and time-consuming. Also, examining a large number of video frames is difficult for physicians for the investigation of infection regions.

To deal with above challenges, computer-aided diagnostic (CAD) systems are introduced by several researchers (Liaqat et al., 2018) which consist of two general steps – infection detection and infection classification (Liu, Gu, Xie, Xiong, & Qin, 2012; Yuan, Li, & Meng, 2017). Most of the existing approaches focus on detection of one particular disease, such as ulcer (Yuan et al., 2015), bleeding (Yuan, Li, & Meng, 2016a) and polyps (Bernal et al., 2015). However, the characteristics of infected regions are shown with different texture, colour, shape, size and irregularities that make the diagnostics process difficult and intense (Berens & Fisher, 2008).

A CAD system generally consists of four principal steps, including frame stretching, infection extraction, infection representation and classification. Frame preprocessing is an essential step for any CAD system due to the complex background, noise and similarity between healthy and diseased regions. In the segmentation step, infection region is extracted by image processing methods, such as thresholding, saliency approaches and many more. However, this step is difficult due to above-listed challenges, such as irregularities, similarity in colour, texture and shape. Feature extraction is a most important step in any CAD system used for the representation of an image. Several types of features are extracted for any CAD system, such as texture, colour, shape, geometric and point (Akram, Khan, Sharif, & Yasmin, 2018; Sharif, Tanvir, Munir, Khan, & Yasmin, 2018). These extracted features are finally utilised by classifiers, such assupport vector machine (SVM), neural network (NN), decision trees and ensemble methods (Dey, Ashour, Shi, & Sherratt, 2017; Li et al., 2017; Saba et al., 2016).

Motivation and contributions

Recently, deep learning pays much attention to the domain of machine learning based on its several applications, such as video surveillance, agriculture and biometrics (Asadi-Aghbolaghi et al., 2017; Kamilaris & Prenafeta-Boldú, 2018; Sundararajan & Woodard, 2018). Besides the success of deep learning in other machine learning applications, it also gives best solutions for accurate classification and good accuracy for medical imaging, such as skin cancer, breast tumour and brain tumour (Hu et al., 2018). In deep learning, CNN shows much attention for classification and gives better accuracy against large number of medical datasets. CNN is used for automatic feature extraction consisting of both local and highly deep patterns utilised for classification. However, shape-based features such as histogram and geometric are used for shape information of infection in the given images. Therefore, inspired with CNN applications and performance of shape features, a new method is proposed based on geometric and CNN features. The proposed approach deals with complex classification problems, such as colour variations and texture information.

Two core steps are performed for this automated system such as lesion region is extracted by new contrast-enhanced colorcolour features (CECF) approach and CNN feature extraction. The proposed CECF approach gets along with two sub-steps: i) improve the local contrast of infection area by hybrid contrast stretching technique which incorporates top-bottom hat filtering and median filter; ii) Hue, Saturation, Variation (HSV) transformation is performed to extract colour features which are further utilised for similarity measures between HSV colour image pixels. Then, a weight value is assigned to set a threshold function for lesion segmentation. In the second step, deep CNN and geometric features are extracted. VGG16 and VGG19CNN models are utilised for

deep features. Later, max-pooling is performed and Euclidean Fisher Vector (EFV) method is used for features fusion. Finally, conditional entropy is performed for best features selection which is later classified by K-Nearest Neighbour (KNN) classifier for classification. The major contributions are given below:

(1) A hybrid contrast stretching method is proposed in the preprocessing step which controls global contrast of input image and improves the local contrast. The local contrast makes the lesion area distinct as compared to the healthy region. Two famous contrast stretching filters as top-bottom hat are implemented and their outputs are combined in a new image which is later improved by median filtering.
(2) HSV colour transformation is performed and colour features are extracted by employing statistical parameters. The extracted features are combined in one matrix and mean deviation (MD) is calculated. A weight value is obtained through MD which is later utilised for cluster initialisation of extracted features. The features which are higher than the weight value are assigned as distinct cluster while others are assigned as separate cluster. Finally, a conditional probability-based threshold value is obtained which is utilised in threshold function for final segmentation.
(3) Geometric and CNN features are computed from enhanced RGB and segmented images. The geometric features are taken from segmented lesion region and DCNN features are calculated from enhanced RGB images. Then, 3×3 max-pooling operation is performed to select the higher value features.
(4) DCNN features are fused by EFV method to remove the redundant information. After that, geometric features are simply concatenated with EFV unique DCNN feature vector. Finally, conditional entropy is implemented on fused features to selects the best features.
(5) A new database is designed consisting of 5500 WCE images of 1500 ulcer, 2000 bleeding and 2000 healthy.

Related work

Lesion segmentation and feature extraction have gained much attention in the area of CV and image processing based on their several applications, such as surveillance (Khan, Akram, Sharif, Javed, et al., 2018; Khan et al., 2017; Shah, Chen, Sharif, Yasmin, & Fernandes, 2017; Sharif et al., 2017), medical imaging (Ali, Yasmin, Sharif, & Rehmani, 2018; Amin, Sharif, Yasmin, & Fernandes, 2018; Bokhari, Syedia, Sharif, Yasmin, & Fernandes, 2018; Naqi, Sharif, Yasmin, & Fernandes, 2018; Nasir et al., 2018), agriculture (Sharif et al., 2018), biometrics (Sharif, Khan, Faisal, Yasmin, & Fernandes, 2018), scene classification (Ansari, Shah, Yasmin, Sharif, & Fernandes, 2018) and few more (Amin, Sharif, Yasmin, & Fernandes, 2017). In the domain of medical imaging, WCE is a new imaging technology which travels through the human body to investigate stomach infection. However, large number of video frames makes the system more complex and difficult for physicians to investigate the diagnosis. To resolve this kind of issues, in the literature, several computer-based methods are introduced for diagnosis of GIT diseases. Barbosa et al. (Barbosa, Ramos, Correia, & Lima, 2009) presented colour curvelet covariance statistical texture features for automatic lesion detection from WCE images. Said et al. (Charfi & El Ansari, 2018) introduced a new approach for colon abnormalities detection from WCE images using local binary patterns (LBP), variance, and discrete wavelet transform features. The extracted features are calculated with their texture information which is later classified by SVM and NN. Mohsen et al. (Hajabdollahi et al., 2018) presented a simple and effective technique for the segmentation of bleeding area from WCE frames. Hue, saturation, intensity and LAB colour transformations are utilised for feature extraction and classification by multi perceptron learning approach hence showed improved performance. Yuan and Meng (2017) described an auto encoder-based deep learning approach named SSAEIM. The presented approach is formed by nearest neighbour graph and describes the essential structure of given image. The images in similar category extract

similar features and images in the distinct class are kept away. The presented method depicted improved performance and achieved an accuracy of 98.0% for polyps' detection. Farah et al. (Deeba, Islam, Bui, & Wahid, 2018) presented an automatic approach for bleeding region detection based on the fusion of two SVM classifiers. RGB and HSV colour spaces are used and their first-order statistical histogram features are extracted. After that, nested cross-validation is performed to optimise the classifier for better accuracy. Yuan et al. (2017) designed a new CAD system for the classification of GIT diseases. In the presented method, colour scale-invariant features are extracted and separated from all four types of GIT diseases, such as polyps, bleeding, ulcer and healthy using K-Means clustering approach. For feature coding, a linear coding algorithm is introduced named as saliency and adaptive locality-constrained linear coding which encodes the piece of features based on adaptive coding. The experimental results reveal that the presented method showed improved performance and achieved 88.61% accuracy. Ghosh, Fattah, Wahid, Zhu, and Ahmad (2018) introduced an automatic approach for bleeding detection by cluster-based statistical features. The statistical features are extracted from spatial blocks based on their colour pixels information. In the introduced approach, two clusters are obtained and their cluster-based features are extracted. After that, global and differential features are taken used to detect the bleeding region. Yuan, Li, and Meng (2016b) described BoF approach for polyps' classification from WCE images. Different texture features from neighbourhoods of key points are extracted and all of them are combined for classification. The experimental results reveal that introduced method achieved 93.2% accuracy which is good as contrast to existing approaches. Yuan et al. (2015) introduced a saliency technique for ulcer. To segment important regions from an input image, multilevel super pixels segmentation is performed. The colour and texture features are extracted from each super pixel region and combined them to create a saliency map. The created saliency map is applied on input images to recognise the ulcer regions. Moreover, a linear coding scheme named LLC is also introduced which gives improved performance and achieved an accuracy of 92.65%.

The above studies mostly focused on texture, colour and point features for the classification and recognition of GIT infection; however, recent advances in deep learning may affect the recognition accuracy. Moreover, it can be noticed that most of the researchers just focused on frames classification, such as ulcer, bleeding, and so on; however, they never focused on infection segmentation. Moreover, as features selection is a challenging task in medical imaging for recognition of infections, additional and irrelevant features degrade overall recognition accuracy. Therefore to deal with all of these problems, in this article, a new computer-aided method is proposed which is based on CNN and colour-based segmentation.

Proposed method

The proposed method involves two principal steps – i) geometric feature extraction from a segmented lesion, ii) deep CNN feature extraction and fusion for classification. To control the global contrast of input image, preprocessing is done at the very first step and then CIELAB colour features-based segmentation is performed. In the second step, deep CNN features are computed from enhanced RGB images and fused their information which is further refined by EFV and best features selection approach. The detailed proposed design of automated system is shown in Figure 1.

Contrast-enhanced colour features (CHCF)

Colour features are much important in the area of CV to predict an object based on its colour information. It gains much attention due to various CV applications. In the area of medical imaging, colour features are mostly used for classification, such as skin cancer (melanoma, benign), breast tumour and brain tumour. In view of the importance of colour features, a new CHCF method is proposed in this article which is the combination of contrast stretching and CIELSB colour features.

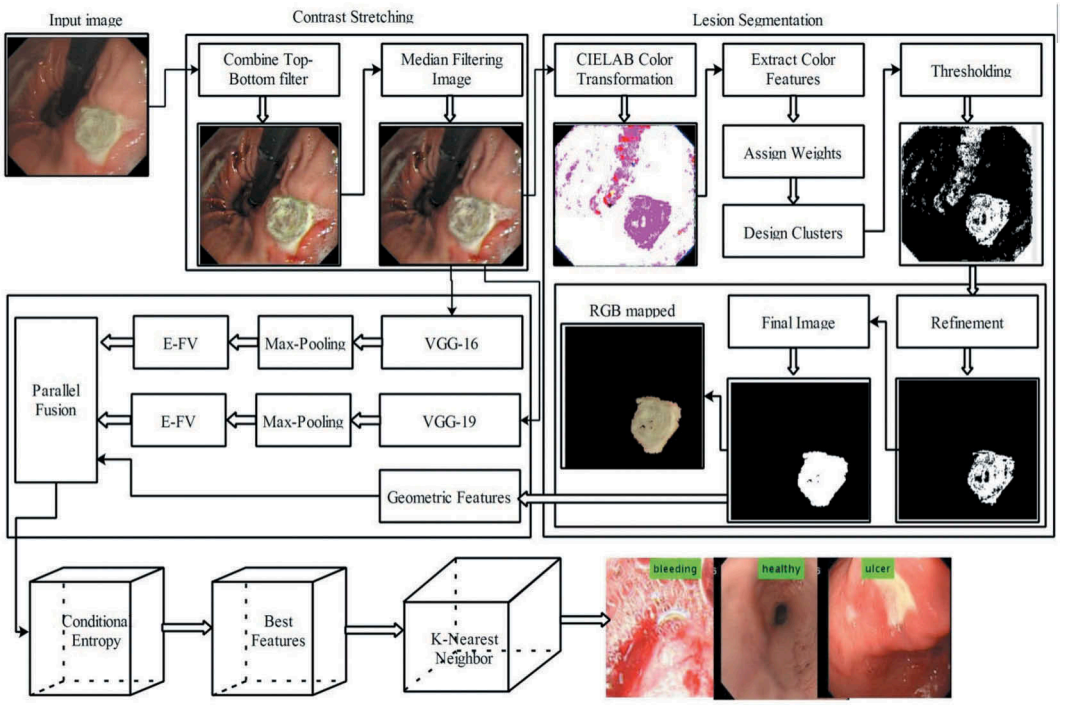

Figure 1. System architecture of proposed GIT infections detection and classification.

Contrast elongating

Contrast elongating is performed to increase lesion contrast in the given image. It is one of the very first steps in image processing to improve the contrast of given image. The input image is invented and the resultant image is more suitable as related to the original image (Negi & Bhandari, 2014). This process is very helpful in the domain of medical imaging due to several challenges, such as noisy effects during the acquisition, boundary enhancement, automatic lesion detection and automatic contrast compensation for segmentation. The major advantage of contrast stretching is to remove extra information which is not essential for the next step. Also, global contrast of input image is controlled and lesion contrast is improved which may be helpful for disease segmentation. The formulation of implemented contrast stretching approach is given below.

Let Δ denotes the selected database which is prepared from WCE video. WCE videos are collected from POF Hospital, Pakistan and each video consists of thousands of frames. To process these videos, initially videos are transformed into frames and then proposed method is applied. Let $J(i,j) \in \Delta$, where $J(x,y)$ is an extracted frame of size 512×512 as shown in Figure 2.

In contrast stretching phase, top-hat and bottom-hat filters are performed. The top-hat filter improves the contrast of foreground regions whereas bottom-hat filter controls the background contrast. The formulation of top and bottom hat filters is given in Equations (1) and (2) below.

$$\xi_t(i,j) = F_T(J(i,j),\beta) \tag{1}$$

$$\xi_b(i,j) = F_B(J(i,j),\beta) \tag{2}$$

where $\xi_t(i,j)$ signifies top-hat filtered image, $\xi_b(i,j)$ implies bottom-hat filtered image, $F_T()$ and $F_B()$ denote the top and bottom hat filtered functions and β denotes the stretching parameter initialised to 12. After that, information of both top and bottom hat filters is combined in one image which makes the image better in terms of visual information and easy background. Further,

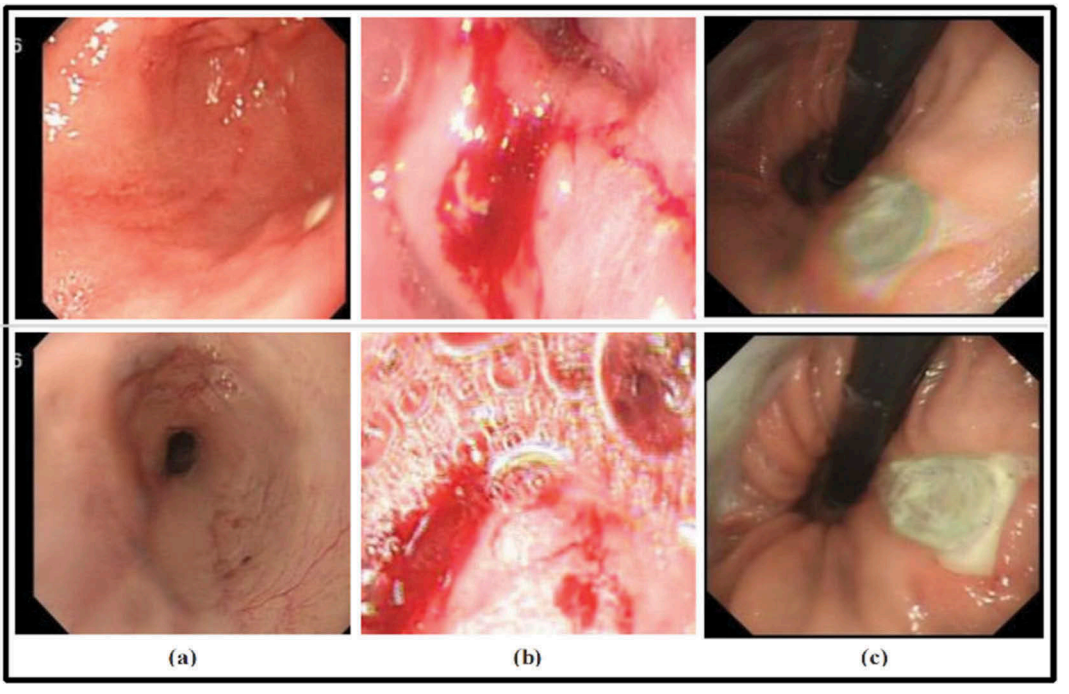

Figure 2. Sample images from WCE videos. (a) Healthy samples, (b) bleeding samples and (c) ulcer samples.

3D median filter is applied on combined information to remove the salt and pepper noise (Sun, Zhao, & Yuan, 2018). The formulation of this process is defined as in Equations (3)–(5).

$$\xi_A(i,j) = \sum(\xi_t(i,j), J(i,j)) \tag{3}$$

$$\xi_C(i,j) = \sum(\xi_A(i,j), \xi_b(i,j)) \tag{4}$$

$$\xi_M(i,j) = F_{MD}(\xi_C(i,j), \alpha) \tag{5}$$

where $\xi_A(i,j)$ shows addition of top-hat and original image pixels information, $\xi_C(i,j)$ denotes combined pixels information image, $\xi_M(i,j)$ represents 3D median filtered image, $F_{MD}()$ is median filtered function and α denotes the filter size which is $3 \times 3 \times 3$. The contrast stretching effects are shown in Figure 3 in which (b) and (c) show the top and bottom hat filtered images, respectively, (d) denotes the combined top-bottom filtered information and (e) shows the median filter effects after combined filtered image.

Lesion detection

Lesion detection is an important step in medical imaging to perceive infection region from the given image. Several methods are proposed in the literature for lesion segmentation in medical imaging, such as skin cancer, brain tumour, breast tumour and many more (Dhane et al., 2016, 2017). Detection of GIT diseases from WCE images is a difficult task due to several issues, such as irregularity of infection region, shape of a lesion, colour similarity and size of disease region. To resolve these issues, CIELAB colour features-based approach is proposed here. In this approach, CIELAB colour transformation is performed on the median filtered image to extract colour features as mean (μ), variance, SD, skewness and kurtosis. The extracted colour features are combined in matrix and MD is calculated. Through MD, a weight value is obtained which is used to make the cluster of extracted features. The features which

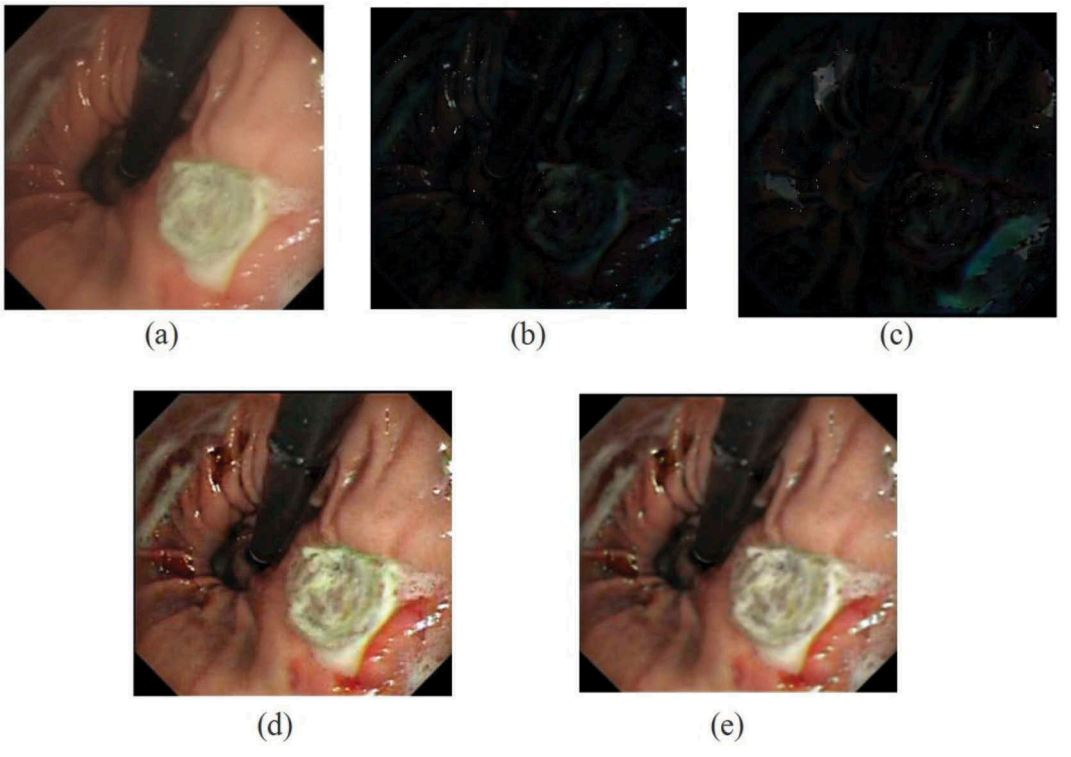

Figure 3. Proposed contrast stretching effects on input image: (a) original RGB image, (b) top-hat filtering, (c) bottom-hat filtering, (d) combined top-bottom hat image and (e) median filtered image.

are higher than the weight value are assigned as distinct cluster while others are assigned as separate cluster hence two clusters are obtained through the weight value. Finally, a conditional probability-based threshold value is obtained which is utilised in threshold function. The above formulation is performed as follows.

$\xi_M(i,j)$ is a median filtered RGB image of size $512 \times 512 \times 3$, where three channels are red, green and blue and are extracted by Equation (6) below.

$$\xi_R(i,j) = \frac{r}{\sum_{k=1}^{3} K_k}, \cdot\xi_G(i,j) = \frac{g}{\sum_{k=1}^{3} K_k}, \cdot\xi_B(i,j) = \frac{b}{\sum_{k=1}^{3} K_k} \tag{6}$$

Then, CIELAB transformation is done using extracted $\xi_R(i,j)$, $\xi_G(i,j)$ and $\xi_B(i,j)$ channels as given in Equations (7)–(10).

$$L = \begin{cases} 116\left(\frac{Y}{Y'}\right)^{\frac{1}{3}} - 16 & if\ \frac{Y}{Y'} > 0.0088 \\ 903.03\ \left(\frac{Y}{Y'}\right) & if\ \frac{Y}{Y'} > 0.0088 \end{cases} \tag{7}$$

$$a = 500 \times \left(f\left(\frac{X}{X'}\right) - f\left(\frac{Y}{Y'}\right)\right) \tag{8}$$

$$b = 200 \times \left(f\left(\frac{Y}{Y'}\right) - f\left(\frac{Z}{Z'}\right)\right) \tag{9}$$

$$f(I) = \begin{cases} I^{\frac{1}{3}} & if\ I > 0.0088 \\ 7.787 \times t + \frac{16}{116} & if\ t \leq 0.0088 \end{cases} \tag{10}$$

where X implies linear combination, Y denotes the luminance channel and Z is a blue stimulation, whereas $X = \frac{Y}{y}x$, $Y = \frac{Xy}{y}$, and $Z = \frac{Y}{y}(1 - x - y)$. The effects of CIELAB transformation are shown in Figure 4.

Using above CIELAB colour transformed image, colour features are extracted and these parameters provide a vector of size 1 × 1027. Mathematically, these features are computed from Equations (11)–(16).

$$Mean = \bar{f} = \frac{f_1,\ f_2,\ f_3, \ldots\ f_n}{N} = \sum_{i=1}^{N} \frac{f_i}{N} \tag{11}$$

$$variance = \frac{\sum_{i=1}^{N} \left(f_i - \bar{f}\right)^2}{N - 1} \tag{12}$$

$$SD = \sqrt{variance} \tag{13}$$

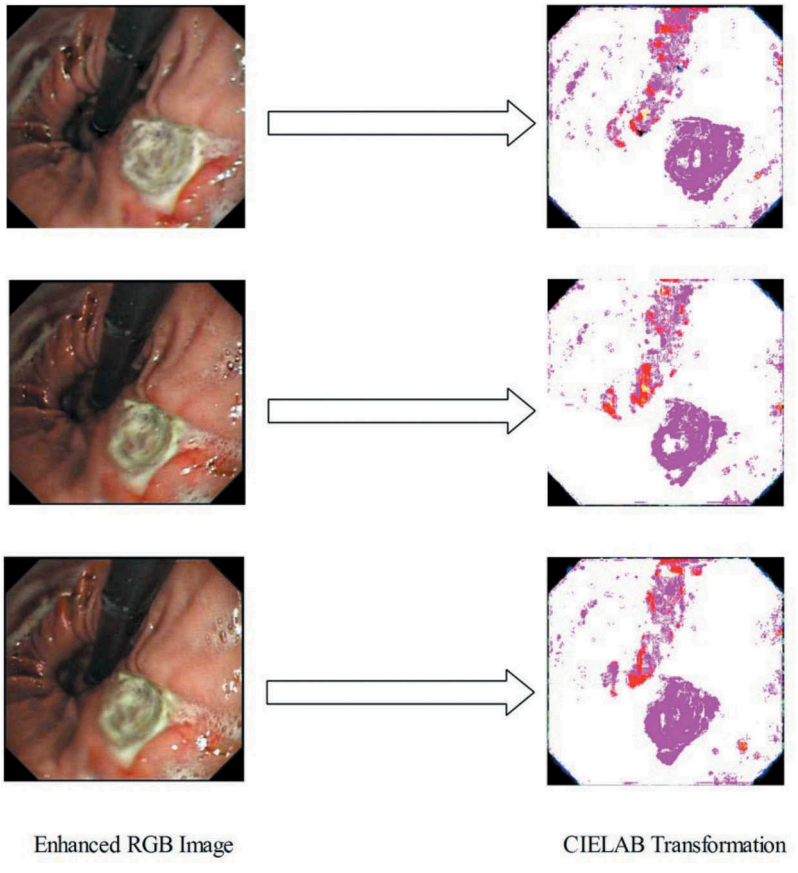

Figure 4. CIELAB colour transformation for colour feature extraction. The enhanced RGB image denotes the median filtered images represented as $\xi_M(i,j)$ containing three channels as red (R), green (G) and blue (B), respectively.

$$Kurtosis = \frac{\sum_{i=1}^{N} \frac{(f_i - \bar{f})}{N}}{SD^4} \tag{14}$$

$$Skewness = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{N} \left(\frac{(f_i - \bar{f})}{SD} \right)^3 \tag{15}$$

$$\xi_{fused}(i) = \sum_{j=1}^{M} [f_j] \tag{16}$$

where $\xi_{fused}(i)$ denotes the fused colour feature vector of size 1×1027, j shows the extracted colour feature parameters and M represents total number of colour parameters. After this, MD of colour feature vector is computed as in Equation (17) given below.

$$\xi_{MD} = \sum_{i=1}^{N} \frac{|f_i - \bar{f}|}{N} \tag{17}$$

The calculated MD value is used as a weight value to make clusters of colour feature vector. The features which are higher or equal than the weight value are assigned $C1$ cluster whereas others are assigned cluster $C2$. Then, a conditional probability is calculated from both clusters to select a lower probability value which is utilised in a threshold function to select the lesion and healthy regions. The calculation of threshold function is explained below.

Suppose $\xi_{LP}(f)$ denotes the lowest probability colour feature calculated from conditional probability as given in Equations (18)–(20).

$$P_c = P(f_i | f_j) = \frac{P(f_i \cap f_j)}{P(f_j)} \tag{18}$$

$$P(f_i \cap f_j) = P(f_i \cap f_j) P(f_j) \tag{19}$$

$$TH(i,j) = \begin{cases} 255 & if\ f_{ij} \geq \xi_{LP}(f) \\ 100 & if\ f_{ij} < \xi_{LP}(f) \end{cases} \tag{20}$$

where P_c denotes the conditional probability variable, $P(f_i \cap f_j)$ represents conditional probability between i and j, $TH(i,j)$ is a threshold function for binary image. In the above threshold function, f_i and f_j show the extracted features. The value 255 denotes white pixels if the value of f_{ij} is greater than the lowest probability colour feature $\xi_{LP}(f)$ otherwise 100 value denotes black pixels. The white pixels are infection areas whereas black pixels are considered as healthy regions which are removed in the segmentation step. The selection values such as 255 and 100 are selected by heuristic approach. Through this approach, selection process is iterated 500 times to produce good values on selected database but these values may be not suitable for other datasets. In the iteration process, a selection variable k is initialised which returns two best values after 500 iterations for healthy and infection regions. The effects of threshold function are shown in Figure 5. Finally, thresholding results are refined by performing two morphological operations, such as dilation and fillings. The effect of morphological operations is given in Figure 6.

Thereafter, some geometric features are calculated from final segmented image to capture the lesion information in terms of shape, size, texture, area and width. For this purpose, six geometric features are calculated, including area, solidity, major axis length, minor axis length, orientation, eccentricity, height and width. These features are combined in one matrix and are later fused in deep CNN features extracted in the next section.

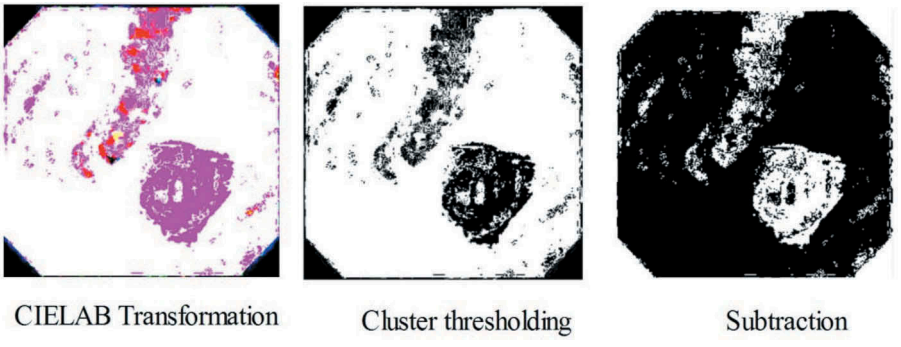

Figure 5. Proposed probability-based colour features clustering effects.

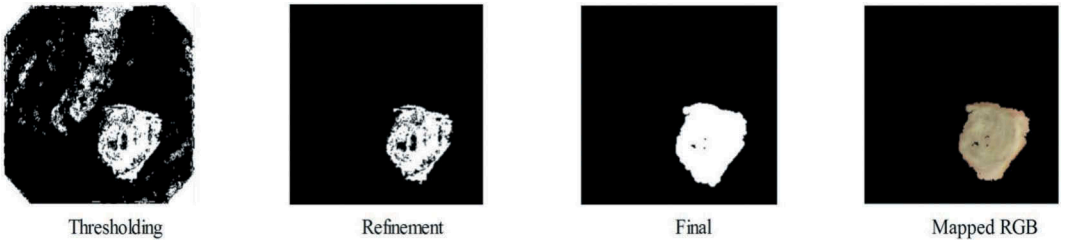

Figure 6. Results after morphological operations.

Deep CNN feature extraction

Deep CNN has confirmed exciting performance in various CV domains, such as face recognition, action recognition, plants diseases recognition and medical imaging (Khan, Akram, Sharif, Awais, et al., 2018; Rashid et al.). These listed tasks noted that CNN is not useful only for small data but equally shows improved performance on large data (Chen, Patel, & Chellappa, 2016; Martis, Gurupur, Lin, Islam, & Fernandes, 2018; Raza et al., 2018). The major advantage of CNN is that when one model is generated, it can be used for other tasks by fine tuning. Recently, several deep CN models are introduced, such as VGG (Simonyan & Zisserman, 2014), AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), ResNet (He, Zhang, Ren, & Sun, 2016), GoogleNet (Szegedy et al., 2015) and Yolo (Redmon, Divvala, Girshick, & Farhadi, 2016). These models are used in machine learning for deep feature extraction. In general, a simple CNN model consists of a convolutional layer, pooling layer, Relu layer, normalisation layer, FC layer and one SoftMax function for classification. The convolutional layer follows simple or local features while FC layer exploits as a decision-maker of features. Therefore, feature extraction through CNN is highly succeeded, more efficient and useful as compared to hand-crafted features (Li & Yu, 2016). In this article, two pre-trained models like VGG16 and VGG19 are utilised for deep CNN feature extraction.

VGG16 is pre-trained model proposed by Simonyan and Zisserman (2014) which is trained on ImageNet database consisting of more than a million samples of 100 different object classes. The size of input layers of this model is $224 \times 224 \times 3$. Complete architecture is having one input layer, five convolutional layers (used only 3×3 size), seven ReLu layers, five pooling layers (used only 2×2 size), three fully connected layers, two drop layers, one prob layer and one SoftMax classification function.

VGG19 is a pre-trained model (Suman et al., 2017) used in ImageNet Large Scale Visual Recognition Challenge (Fu et al., 2014). This model is trained on millions of images of 1000 different object classes. The rich features are extracted for learning of this model. It contains a total of 47

layers, including input layer, 16 convolutional layers, 18 Relu layers, 5 pooling layers, 3 fully connected layers, 2 drop layers and 1 SoftMax function for classification.

In this work, VGG16 and VGG19 deep learning models are trained on WCE enhanced images. In the training process, initially selected deep learning models are loaded and then a path is set for enhanced median filtered images. Further, preprocessed images are utilised for training the models. Afterwards, a ratio of 50% is selected for the training process. Input and output layers are defined for performing activation. For VGG16 and VGG19, input layer is X0 and output layer is FC6 (VGG16) and FC7 (VGG19). After the activation function, two vectors are obtained for both models of size $M \times 4096$ and $N \times 4096$, where M and N denote number of training samples. Finally, the extracted features are reshaped based on their training samples and number of extracted features for both models of size $M \times 4096$ and $N \times 4096$. Both vectors are fused in one matrix and selection process is performed for best features selection. The selected features are fed to a classifier for saving a trained model. Thereafter, the trained model is utilised for the testing process.

In the testing phase, first activation is performed on output layer FC6 (VGG16 model) and deep CNN features of size 1×4096 are obtained. A 3×3 max-pooling operation is applied on extracted vector. There is an available option of mean/average pooling in which average of window's components are taken. Max features are primarily designed to extract extreme features such as edges and boundaries whereas in case of mean pooling, there is a high probability of losing these extreme features by embedding the smoothness factor, i.e. average. As this research extracts extreme features, hence not in a position to lose any of the details, therefore considering mean features will not work accordingly.

The max-pooling returns an output vector of size 1×455 which is saved in a new matrix as shown in Figure 7. From VGG19 model, activation function is applied on output layer FC7 to extract deep features of dimension 1×4096. Similarly to VGG16, a 3×3 max-pooling operation is performed on extracted matrix and an output vector of size 1×455 is obtained. The utilised architecture for deep CNN feature extraction from VGG19 model is shown in Figure 8.

After deep CNN feature extraction, a new technique is proposed, namely EFV for unique features fusion. In EFV approach, fisher vector-based similar features are sorted into one separate cluster and other features are sorted into another cluster. Thereafter, those features are removed which appear more than one time in both clusters. These features are removed by Euclidean distance (ED) approach. ED is calculated between similar features of both DCNN vectors based on their neighbourhood feature. EFV approach gives unique feature vector which is later fused parallel in one matrix. Later on, conditional entropy approach is applied on fused vector for best features selection. The working of proposed EFV approach is given below.

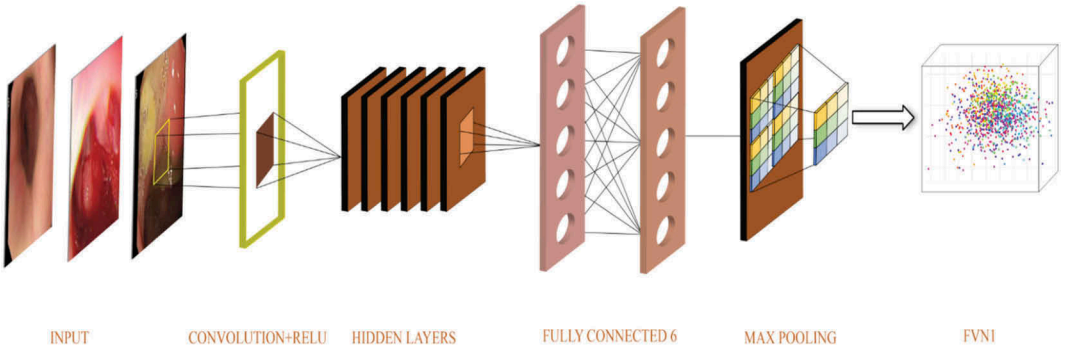

Figure 7. Utilised architecture of VGG16 for deep CNN feature extraction.

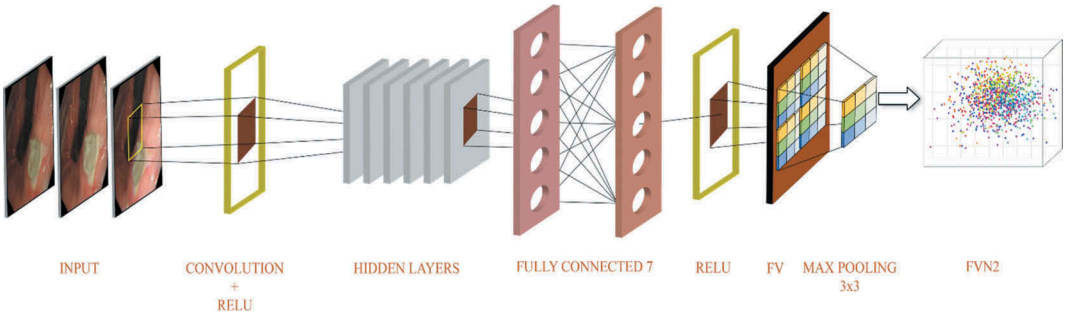

Figure 8. Employed architecture of VGG19 for feature extraction.

Let $FV1$ denotes the VGG16 deep CNN features of length 1×455, $FV2$ denotes VGG19 deep CNN features of dimension 1×455 and $Fv(i)$ implies the output fused unique feature vector. Let $FV(i) = \{f_t,\ t = 1,\ 2, 3 \ldots N\}$ be a N number of extracted features, $f_t \in \aleph$. Let u_λ be a PDF which models the generative process of elements in N where $\lambda = [\lambda_1, \lambda_2, \ldots \lambda_M]^{'} \in \mathbb{R}^M$ denotes M parameters of u_λ. Then, the score function is defined as in Equation (21).

$$K^{FK}(Q) = U_\lambda{}^{FV'} F_\lambda{}^{-1} U_\lambda{}^{Q} \tag{21}$$

where F_λ denotes the semi-definite. The above expression is further modified in Equation (22).

$$= G_\lambda{}^{FV'}.\ G_\lambda{}^{Q}$$

$$= G_\lambda{}^{FV} = L_\lambda U_\lambda{}^{FV} = L_\lambda \nabla \log u_\lambda(FV) \tag{22}$$

Through above expression, similar features are sorted in initial indexes which are further modified by minimum distance-based distinctive formula. From the distinctive formula, similar features are removed by a threshold function given in Equation (23) below.

$$T^F(i) = \begin{cases} \psi_s\ (i) & if\ FV1_i(\lambda) \neq FV2_j(\lambda) \\ \psi_N(j) & OtherWise \end{cases} \tag{23}$$

The above formulation selects features into distinct parameters as $\psi_s(i)$ and $\psi_N\ (j)$, where $\psi_s(i)$ shows distinct fused features merged by a parallel approach and $\psi_N\ (j)$ denotes similar features of both vectors. Thereafter, ED is performed on $\psi_N(j)$ and redundant features are removed to select only one feature of the same value. Later on, both distinct feature vectors are combined into one matrix by serial approach. After that, geometric features obtained in 'Lesion detection' section are simply merged in fused deep CNN vector. A conditional entropy-based approach is applied on fused vector and best features are selected. In the literature, conditional entropy is mostly used to control the uncertainty and randomness in CV and machine learning areas and not for features selection. Instead, Shannon entropy is used for features selection (Khan, Gani, Wahab, & Singh, 2018).

The detailed architecture of deep CNN and geometric features fusion and selection is shown in Figure 9.

The major aim of best features selection is to remove irrelevant features and improve the recognition accuracy. Moreover, another advantage of selection approach is to make the classification time better which is necessary for any computer-based system. The process of conditional entropy-based selection is described below.

Let there be a fused feature vector ψ_{fused} of features $f_i,\ f_j$, where $f_i \in X$ and $f_j \in Y$. Then the function of entropy-based best features selection is defined through Equations (24)–(26).

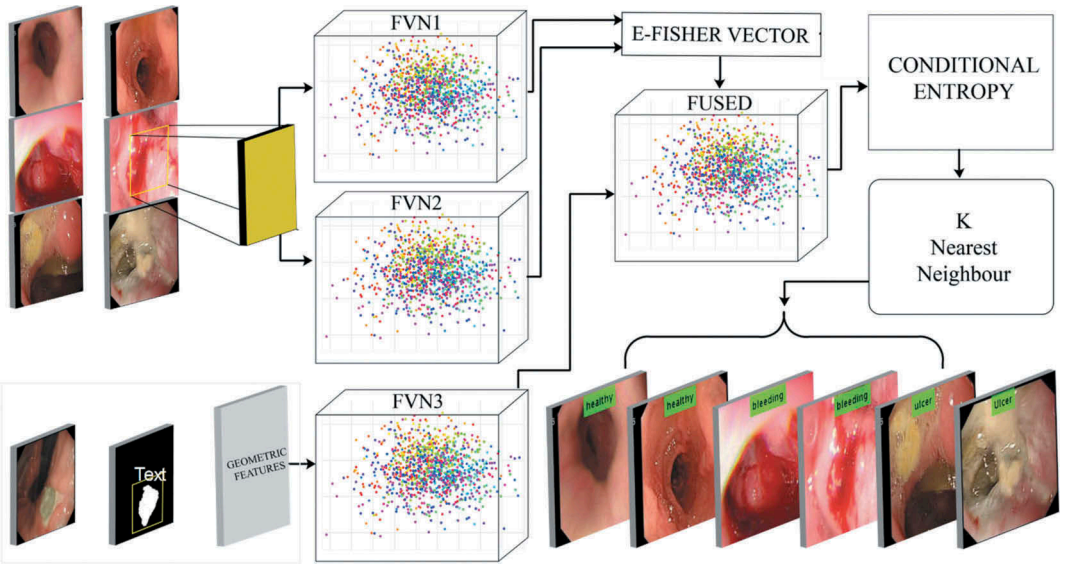

Figure 9. Proposed features fusion and selection approach.

$$S(FV_i) = -\sum_{i\in X} p(f_i) \sum_{j\in Y} p(f_j/f_i) \log p(f_j/f_i) \tag{24}$$

$$= \sum_{i\in X,\ j\in Y} p(f_j, f_i) \log\ p(f_j/f_i) \tag{25}$$

$$S(FV_i) = \sum_{i\in X,\ j\in Y} p(f_j, f_i) \log\ p\left(\frac{p(f_i)}{p(f_i,\ f_j)}\right) \tag{26}$$

The features that are selected and discarded through entropy are plotted in Figure 10.

Figure 10 explains that the features in the range of below 0.4 are discarded after employing entropy and features higher than 0.4 are selected for final classification. The selected features are finally fed to Fine KNN (FKNN) (Larose, 2005; Ramteke & Monali, 2012) for classification. The labelled results using FKNN are shown in Figure 11.

Results and discussion

The proposed technique is determined on a privately collected database which consists of total 10 videos obtained from POF Hospital, Pakistan. These videos are converted into frames and 4500 frames are selected consisting of 2000 bleeding, 2000 healthy and 1500 ulcer disease. For experimental results, deep CNN features are calculated from enhanced RGB images and fused with geometric features from segmented lesion images. Thereafter, best features are selected which are later classified by KNN. The performance of KNN is compared with six other classification algorithms like quadratic SVM (QSVM), cubic SVM (CSVM), medium Gaussian SVM (MGSVM), linear discriminant analysis (LDA), weighted KNN (WKNN) and sub-space KNN (SKNN). Seven performance measures, including accuracy, sensitivity, specificity, false positive rate (FPR), area under the curve (AUC), precision rate and classification time are implemented for analysis process. All experiments are performed on MATLAB 2018a platform using the desktop of 8 GB RAM and 8 GB GPU.

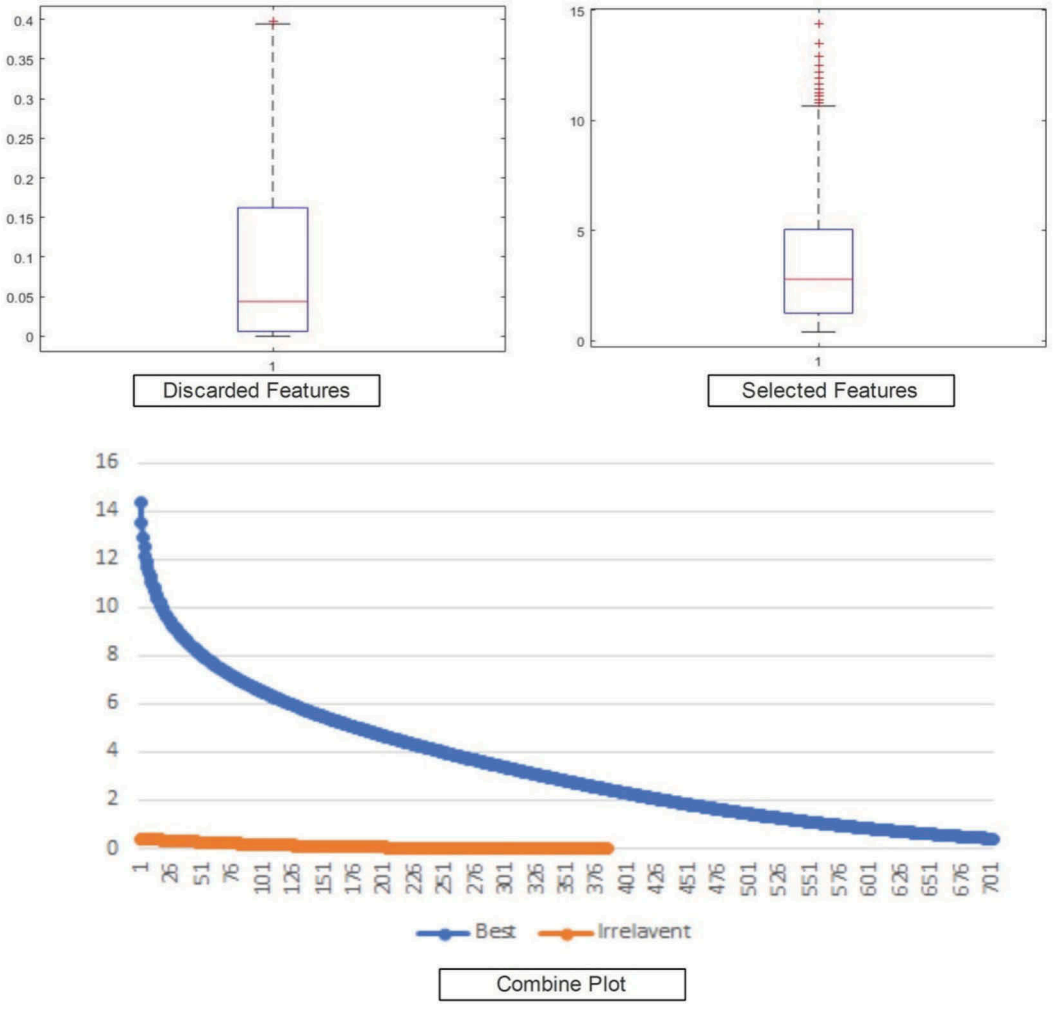

Figure 10. Range of entropy-based best features selection and abandoned. The box plots explain the minimum, average and highest values of extracted features whereas below combined plot explains selected and abandoned features.

Results

In this section, the detailed results of proposed method are presented in terms of accuracy, sensitivity, specificity, FPR, AUC and precision rate which are calculated by confusion matrices. Confusion matrix also coined as error matrix is primarily designed to illustrate interclass (actual class and predicted class) differences and errors. In this article, the possible number of classes are three, including ulcer, bleeding and healthy and three generated confusion matrices (Figures 12–14) are elaborating three different strategies; i) classification is performed on fused deep CNN features extracted from VGG16 and VGG19, ii) classification is performed on the fusion of geometric and deep CNN features, and iii) proposed features selection method is employed for classification performance. Therefore, each matrix is only representing class differences but by implementing multiple strategies. Moreover, the computational time of each classifier is also computed which is helpful for researcher in real-time scenario.

Deep CNN features are extracted in the first phase by utilising VGG16 and VGG19 models as shown in Figure 9 and their pattern information is fused to make the system efficient. The features are fused based on their uniqueness performed by EFV approach. 50:50 ratio is proposed for system evaluation and 10-

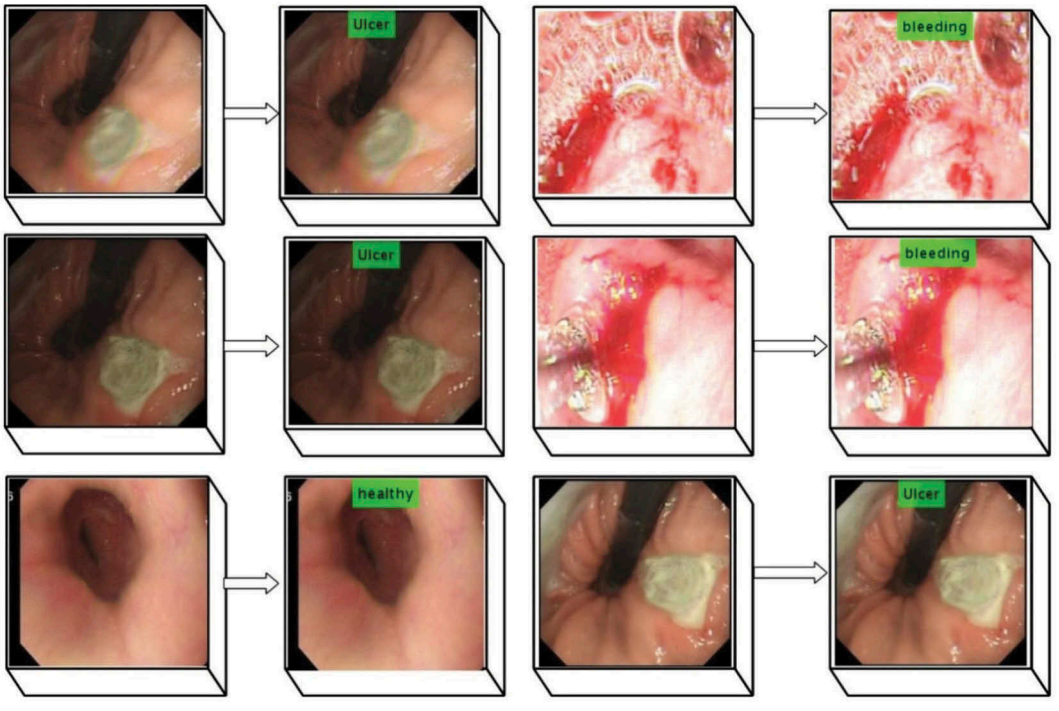

Figure 11. Proposed labelled results on private dataset.

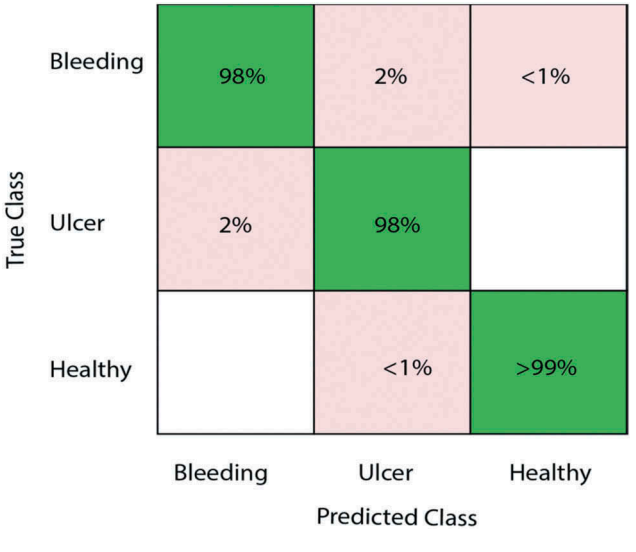

Figure 12. Confusion matrix for deep CNN features fusion on MGSVM.

fold cross-validation is performed on all experiments for classification. Based on this approach, the best accuracy is 98.6% achieved on MGSVM as presented in Table 1. Other parameter for MGSVM such as FP rate is 0.007, sensitivity rate is 0.987, specificity is 98.67% and precision rate is 98.41%. The classification time of selected classification methods is also computed and achieved best time is 131.2 s on KNN

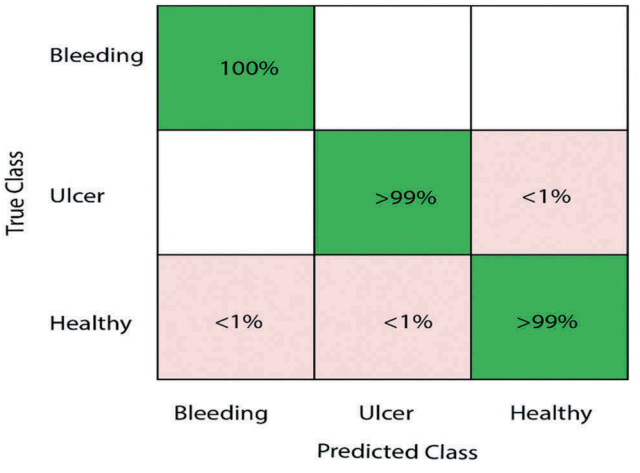

Figure 13. Confusion matrix for proposed unique deep CNN and geometric features fusion.

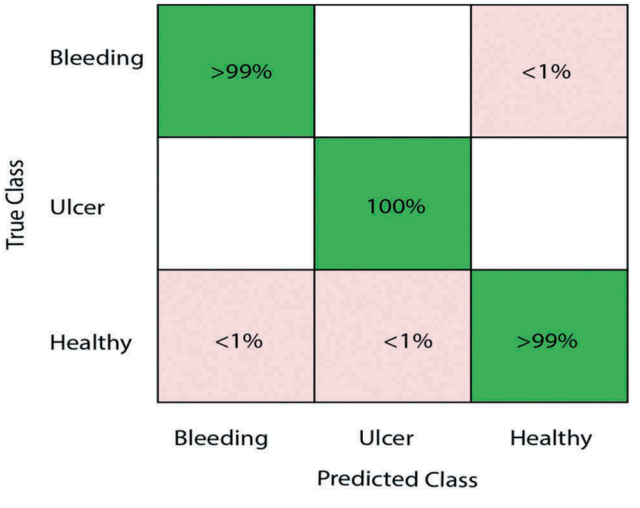

Figure 14. Confusion matrix for proposed features selection approach.

classifier. The best classification results of fused deep CNN on MGSVM are also confirmed by their confusion matrix in Figure 12.

In the second experiment, geometric features are further combined with unique deep CNN fused matrix produced from EFV approach. The aim of fused geometric features with deep CNN is to get efficient classification performance.

Similarly, as in experiment 1, 50:50 ratio is used for training and testing in the second experiment as well. The maximum classification accuracy of 99.32% is achieved on KNN classifier. The other performance parameters for KNN, such as AUC are 1.0, sensitivity rate 1.0, precision rate 99.34% and specificity 100%. However, from this experiment, it is noticed that the accuracy time after geometric features fusion is enlarged up to 0.93% as associated to only unique deep CNN features fusion. Hence, the classification time of this experiment is 136.83 s as given in Table 2. The classification accuracy of the proposed method is also confirmed by Figure 13. From Table 2, the second highest accuracy is 99.23% achieved on QSVM but the second best execution time is 141.5 s on CSVM.

Table 1. Classification performances on fusion of deep CNN features using 10-fold cross-validation.

Sr. No	Classifier	Accuracy (%) (Avg, Rng)	Time (sec)	CL	FP rate	Sensitivity (%)	Specificity (%)		AUC	PPV (%) (Avg, Rng)
1	KNN	(98.1, 98.4)	131.28	B	0.007	0.99	99	98.67	1.00	(98.12, 98.34)
				U			99		1.00	
				H			98		1.00	
2	QSVM	(98.0, 98.4)	194.66	B	0.006	0.99	99	98.67	1.00	(98.04, 98.34)
				U			99		1.00	
				H			98		1.00	
3	CSVM	(97.9, 98.1)	236.18	B	0.006	0.99	99	98.67	1.00	(97.86, 98.40)
				U			99		1.00	
				H			98		1.00	
4	**MGSVM**	(**98.3**, **98.6**)	158.49	B	0.007	0.99	99	98.67	0.98	(**98.23**, **98.41**)
				U			99		0.98	
				H			98		1.00	
5	LDA	(97.2, 98.1)	135.58	B	0.012	0.99	99	98.67	0.98	(97.21, 97.71)
				U			99		0.98	
				H			98		1.00	
6	WKNN	(97.2, 97.6)	133.54	B	0.013	0.98	97.5	97.67	1.00	(97.07, 97.67)
				U			98		0.99	
				H			97.5		1	
7	SKNN	(98.1, 98.4)	765.9	B	0.007	0.99	99	98.67	0.98	(98.19, 98.41)
				U			99		0.98	
				H			98		1	

Table 2. Classification accuracy for fusion of geometric and unique deep CNN features.

Sr. No	Classifier	Accuracy (%) (Avg, Rng)	Time (s)	CL	FP rate	Sensitivity (%)	Specificity (%)		AUC	PPV (%) (Avg, Rng)
1	**KNN**	(**99.13**, **99.32**)	**136.83**	B	0.00	1.00	100	100	1.00	(**99.21**, **99.34**)
				U			100		1.00	
				H			100		1.00	
2	QSVM	(99.01, 99.23)	170.93	B	0.00	1.00	100	100	1.00	(98.92, 99.04)
				U			100		1.00	
				H			100		1.00	
3	CSVM	(98.97, 99.18)	141.5	B	0.00	1.00	100	100	1.00	(98.98, 99.14)
				U			100		1.00	
				H			100		1.00	
4	MGSVM	(98.99, 99.19)	153.92	B	0.00	1.00	100	100	1.00	(99.06, 99.16)
				U			100		1.00	
				H			100		1.00	
5	LDA	(99.06, 99.12)	338.2	B	0.00	1.00	100	100	1.00	(99.08, 99.14)
				U			100		1.00	
				H			100		1.00	
6	WKNN	(99.11, 99.21)	609.3	B	0.00	0.99	99.5	99.67	1.00	(99.05, 99.14)
				U			99.5		1.00	
				H			100		1.00	
7	SKNN	(98.99, 99.20)	150.7	B	0.003	0.99	99	99.34	1.00	(98.12, 98.67)
				U			99		1.00	
				H			100		1.00	

In the third experiment, features selection approach is performed named as conditional entropy to select best features from fused deep CNN and geometric features. The conditional entropy is employed on fused vector to obtain a new vector which is unique and more informative. Similar to experiment 1 and experiment 2, 50:50 approach is opted. Thereafter, features are fused and best features are selected using conditional entropy approach and reached highest accuracy of 99.42% for proposed method as given in Table 3. The other performance parameters for proposed method such as sensitivity are 1.0, specificity 100% and precision rate 99.51%. The classification accuracy of proposed method is also confirmed by confusion matrix in Figure 14. In addition, positive predictive value and negative predictive value for proposed features selection method are also computed as plotted in Figure 15. Moreover, the classification time of proposed method is also

Table 3. Classification performance on proposed features selection approach.

Sr. No	Classifier	Accuracy (%) (Avg, Rng)	Time (s)	CL	FP rate	Sensitivity (%)	Specificity (%)		AUC	PPV (%) (Avg, Rng)
1	**KNN**	(**99.34**, **99.42**)	**12.59**	B	0.00	1.00	100	100	1	(**99.42**, **99.51**)
				U			100		1	
				H			100		1	
2	QSVM	(99.13, 99.21)	38.05	B	0.00	1.00	100	100	1	(99.26, 99.33)
				U			100		1	
				H			100		1	
3	CSVM	(99.09, 99.23)	41.82	B	0.00	1.00	100	100	1	(99.26, 99.33)
				U			100		1	
				H			100		1	
4	MGSVM	(99.01, 99.18)	26.50	B	0.00	1.00	100	100	1	(99.25, 99.33)
				U			100		1	
				H			100		1	
5	LDA	(99.16, 99.30)	36.41	B	0.00	1.00	100	100	1	(98.98, 99.00)
				U			100		1	
				H			100		1	
6	WKNN	(99.04, 99.16)	57.2	B	0.00	0.99	99.5	99.67	1	(98.98, 99.00)
				U			99.5		1	
				H			100		1	
7	SKNN	(99.01, 99.06)	31.7	B	0.00	1.00	100	100	1	(98.98, 99.00)
				U			100		1	
				H			100		1	

calculated as shown in Figure 16 (Proposed features selection). The best classification time achieved after applying features selection method is 12.59 s for proposed method. The second best execution time is 26.50% which is obtained on MGSVM.

Performance of a classifier clearly depends on the effectiveness or accuracy of extracted features. Additionally, algorithm's design also plays a vital role in the classification accuracy because no classifier is designed to work optimal for all cases. One algorithm, let KNN, may work best for one set of features but may not be efficient enough for other set of features. Moreover, it has been observed in the literature (Akram, Laurent, Naqvi, Alex, & Muhammad, 2018) that KNN performed scientifically well when training and testing data is not of considerable size, whilst, SVM performed much better when dealing with colossal amount of data. Therefore, in the presented case and considering the above hypothesis, it has been verified by implementation and experimentation that KNN performs significantly improved as compared to SVM.

From all the above experiments, it is clear that the conditional entropy performs efficiently as compared to the fusion of VGG16 and VGG19 deep features. Also, the fusion of unique deep CNN and geometric features performs well; however, it is noticed that although fusion process improves the classification accuracy but also increases the classification time which is not good for any real-time system. This problem is successfully dealt with conditional entropy approach which selects best features for classification. Through this approach, the classification accuracy is significantly improved and also reduces the classification time. The comparison of all three experiments in terms of classification time is shown in Figure 16 where it is clearly observed that features selection approach significantly reduces the classification time.

Furthermore, the proposed method is also evaluated on different training/testing ratios, such as 50:50, 70:30 and 80:20. The accuracy on selected classifiers is plotted in Figure 17 in which best classification accuracy is achieved on KNN as 99.42% for selection ratio 50:50 whereas on 70:30 it is 98.96% and 96.52% on 80:20. It is also noticed that best classification accuracy for 70:30 approach is 99.16% achieved on LDA and for 80:20 ratio, best accuracy is 97.61% on SKNN. The results plotted in Figure 17 clearly show that 50:50 approach is outperforming on selected classifiers.

Additionally, a comparison of proposed selection method is conducted in Table 4 with a few hybrid features fusion procedures, such as CNN, the fusion of CNN+Geometric features, CNN+LBP features, CNN+HOG features, CNN+SFTA features, CNN+GLCM features, CNN+SIFT features and

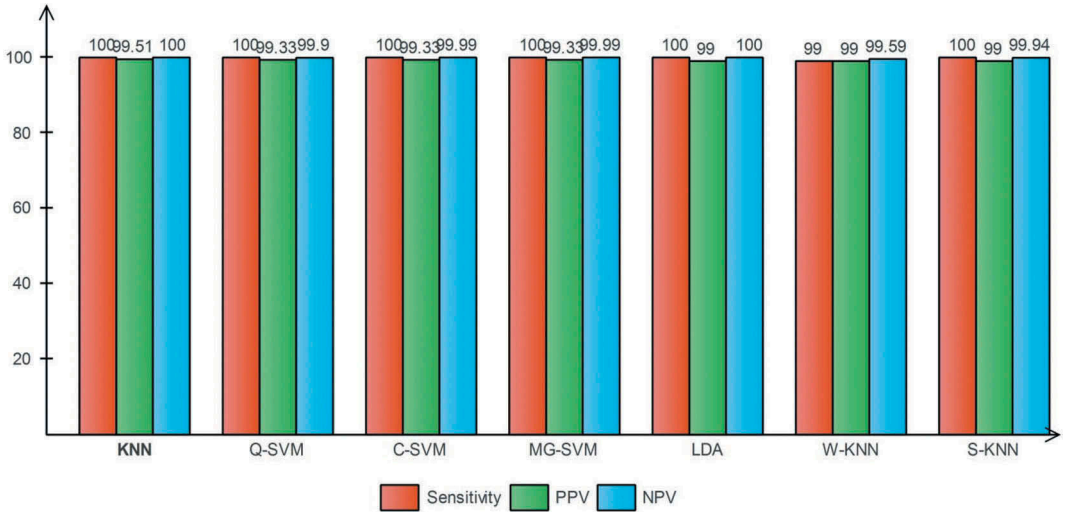

Figure 15. Computation of sensitivity, PPV and NPV for proposed features selection approach.

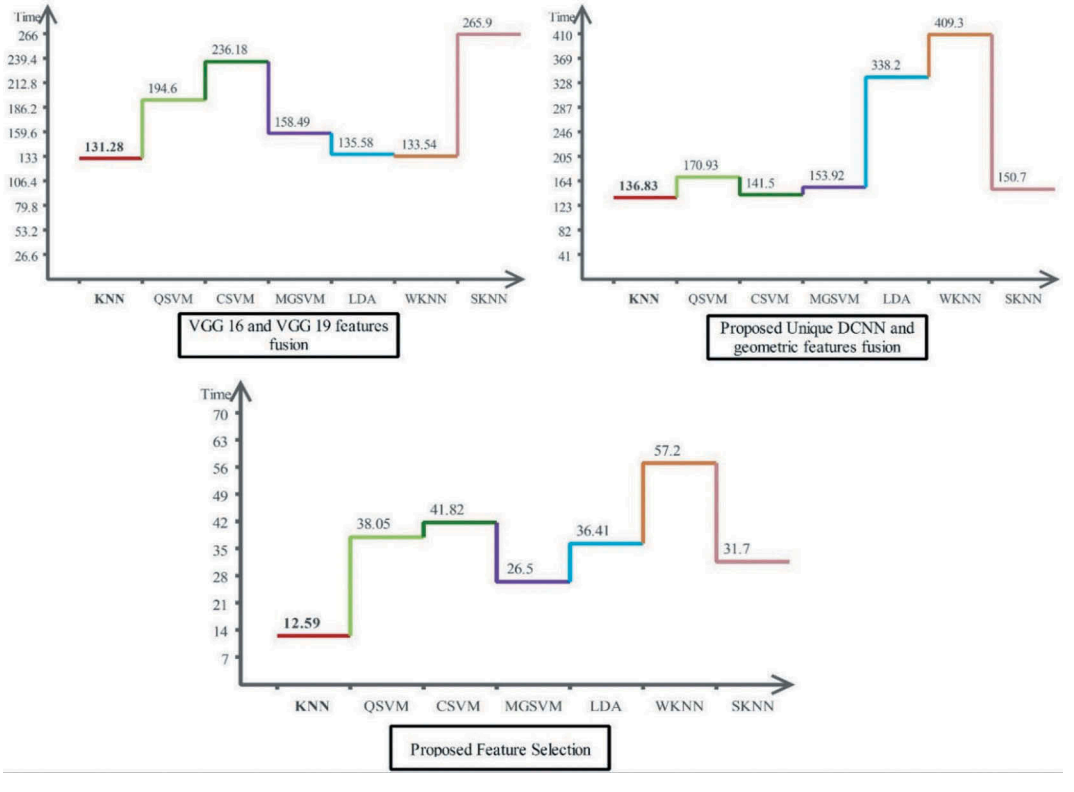

Figure 16. Comparison of proposed methods in terms of classification time.

CNN+Colour features. In Table 4, the achieved classification results on KNN classifier are 98.40%, 99.32%, 93.92%, 90.21%, 94.26%, 91.76%, 89.29% and 95.67%. The texture features like LBP and SFTA reported FN rate of 6.08% and 5.74% whereas CNN+Colour features reported 4.33%. The results show that CNN+Geometric features combination outperformed and reported negative rate

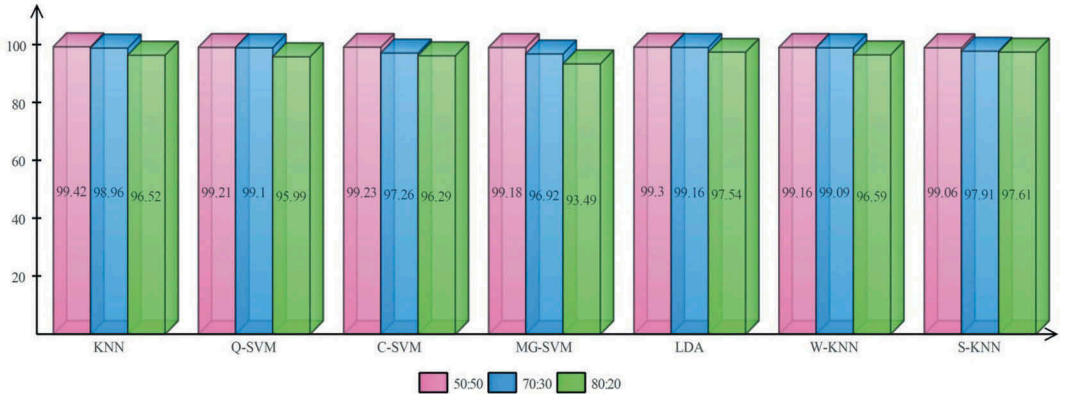

Figure 17. Comparison of different training and testing ratios on proposed selected approach.

Table 4. Features level comparison with proposed approach for KNN classifier.

	Measures		
Method	Accuracy (%)	FNR (%)	Time (s)
CNN	98.40	1.60	131.28
CNN+Geometric	99.32	0.68	136.83
CNN+LBP	93.92	6.08	156.91
CNN+HOG	90.21	9.79	259.41
CNN+SFTA	94.26	5.74	141.21
CNN+GLCM	91.76	8.24	156.92
CNN+SIFT	89.29	10.71	436.54
CNN+Colour	95.67	4.33	149.24
Proposed	**99.51**	**0.49**	**31.70**

is 0.68% along with classification time of 136.83 s. Moreover, the selection process directly affects classification performance and classification time. The KNN performance on proposed section method is significantly good. The achieved KNN accuracy is 99.51% with FN rate as 0.49 and classification time as 31.7 s which shows that CNN performed well.

In the last, the performance of proposed approach is compared with some existing approaches, such as Suman et al. (2017), (2017), Yuan et al. (2017) and Maghsoudi and Alizadeh (2018). In Suman et al. (2017), 96.22% sensitivity rate and 97.89% accuracy for GIT diseases classification are reported. In Suman et al. (2017) and Yuan et al. (2017), classification accuracy is mentioned as 97.67% and 88.61%, respectively. In 2018, Maghsoudi and Alizadeh (2018) achieved improved performance and obtained sensitivity rate of 99.76%. To compare these performances, proposed method achieved classification accuracy as 99.54%, sensitivity rate 100% and precision rate 99.51%. Moreover, the classification time of proposed method is calculated and best time is 12.59 s which is pointedly good as compared to state-of-the-art existing approaches as presented in Table 5.

Table 5. Comparison of classification accuracy with existing approaches.

	Performance parameters			
Method	Sensitivity (%)	Precision (%)	Accuracy (%)	Time (s)
Yuan et al. (2017)	-	-	88.61	-
Suman et al. (2017)	96.22	-	97.89	-
Suman et al. (2017)	-	-	97.67	-
Maghsoudi and Alizadeh (2018)	99.76	-	-	-
Proposed	**100.0**	**99.51**	**99.54**	**12.59**

Conclusion

A new method is proposed for GIT diseases recognition from WCE images in this research which incorporates two principle steps. In the first step, CECF approach is proposed for segmentation of lesion region in the diseased images and their geometric features are extracted. Then, in the second step, unique deep CNN features are extracted from EFV approach and best features are selected based on conditional entropy approach. The proposed approach accomplishes significantly well for complex WCE images, such as low contrast, high texture and similarity in a lesion and healthy region. Moreover, the proposed EFV features fusion method achieved improved classification accuracy on individual features. It is also concluded that the proposed selection approach not only improves the classification performance but also decreases the classification time.

In future, a new deep CNN model can be designed for GIT diseases classification. Moreover, increase in number of images and computational cost of any algorithm can also be considered.

Acknowledgement

The WCE images which are utilised in this article for system evaluation are provided by ERCP lab of Endoscopy Department POF Hospital, Wah Cantt, Pakistan and approved by Faculty of Medicine, Department of Clinical, Wah Medical College, Wah Cantt, Pakistan. We also like to thank Dr Muzammil Jamil who verified these images.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Akram, T., Khan, M. A., Sharif, M., & Yasmin, M. (2018). Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features. *Journal of Ambient Intelligence and Humanized Computing*, 1–20.

Akram, T., Laurent, B., Naqvi, S. R., Alex, M. M., & Muhammad, N. (2018). A deep heterogeneous feature fusion approach for automatic land-use classification. *Information Sciences*, *467*, 199–218.

Ali, H., Yasmin, M., Sharif, M., & Rehmani, M. H. (2018). Computer assisted gastric abnormalities detection using hybrid texture descriptors for chromoendoscopy images. *Computer Methods and Programs in Biomedicine*, *157*, 39–47.

Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. (2017). A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognition Letters*.

Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. (2018). Big data analysis for brain tumor detection: Deep convolutional neural networks. *Future Generation Computer Systems*, *87*, 290–297.

Ansari, G. J., Shah, J. H., Yasmin, M., Sharif, M., & Fernandes, S. L. (2018). A novel machine learning approach for scene text extraction. *Future Generation Computer Systems*, *87*, 328–340.

Asadi-Aghbolaghi, M., Clapes, A., Bellantonio, M., Escalante, H. J., Ponce-López, V., Baró, X., ... & Escalera, S. (2017, May). A survey on deep learning based approaches for action and gesture recognition in image sequences. In Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference IEEE, Washington DC, USA (pp. 476-483).

Barbosa, D., Correia, J. H., Ramos, J., & Lima, C. S. (2009). Automatic detection of small bowel tumors in capsule endoscopy based on color curvelet covariance statistical texture descriptors. In IEEE EMBS Annual International Conference–EMBC 2009 (pp. 6683). IEEE.

Berens, J., & Fisher, M. (2008). Wireless capsule endoscopy color video segmentation. *IEEE Transactions on Medical Imaging*, *27*(12), 1769–1781.

Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., & Vilariño, F. (2015). WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, *43*, 99–111.

Bokhari, F., Syedia, T., Sharif, M., Yasmin, M., & Fernandes, S. L. (2018). Fundus image segmentation and feature extraction for the detection of glaucoma: A new approach. *Current Medical Imaging Reviews*, *14*(1), 77–87.

Charfi, S., & El Ansari, M. (2018). Computer-aided diagnosis system for colon abnormalities detection in wireless capsule endoscopy images. *Multimedia Tools and Applications*, *77*(3), 4047–4064.

Chen, J.-C., Patel, V. M., & Chellappa, R. (2016). *Unconstrained face verification using deep cnn features*. Applications of Computer Vision (WACV), 2016 IEEE Winter Conference, Lake Placid, NY, USA (pp. 1-9). IEEE.

Deeba, F., Islam, M., Bui, F. M., & Wahid, K. A. (2018). Performance assessment of a bleeding detection algorithm for endoscopic video based on classifier fusion method and exhaustive feature selection. *Biomedical Signal Processing and Control*, *40*, 415–424.

Dey, N., Ashour, A. S., Shi, F., & Sherratt, R. S. (2017). Wireless capsule gastrointestinal endoscopy: Direction-of-arrival estimation based localization survey. *IEEE Reviews in Biomedical Engineering*, *10*, 2–11.

Dhane, D. M., Krishna, V., Achar, A., Bar, C., Sanyal, K., & Chakraborty, C. (2016). Spectral clustering for unsupervised segmentation of lower extremity wound beds using optical images. *Journal of Medical Systems*, *40*(9), 207.

Dhane, D. M., Maity, M., Mungle, T., Bar, C., Achar, A., Kolekar, M., & Chakraborty, C. (2017). Fuzzy spectral clustering for automated delineation of chronic wound region using digital images. *Computers in Biology and Medicine*, *89*, 551–560.

Fu, Y., Zhang, W., Mandal, M., & Meng, M. Q.-H. (2014). Computer-aided bleeding detection in WCE video. *IEEE Journal of Biomedical and Health Informatics*, *18*(2), 636–642.

Ghosh, T., Fattah, S. A., Wahid, K. A., Zhu, W.-P., & Ahmad, M. O. (2018). Cluster based statistical feature extraction method for automatic bleeding detection in wireless capsule endoscopy video. *Computers in Biology and Medicine*, *94*, 41–54.

Hajabdollahi, M., Esfandiarpoor, R., Soroushmehr, S., Karimi, N., Samavi, S., & Najarian, K. (2018). Segmentation of bleeding regions in wireless capsule endoscopy images an approach for inside capsule video summarization. *arXiv preprint arXiv:1802.07788*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition. Retrieved from https://www.statista.com/statistics/418515/adult-prevalence-of-gastrointestinal-conditions-by-country/

Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., & Sun, Q. (2018). Deep learning for image-based cancer detection and diagnosis – A survey. *Pattern Recognition*, *83*, 134–149.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90.

Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., & Saba, T. (2018). CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. *Computers and Electronics in Agriculture*, *155*, 220–236.

Khan, M. A., Akram, T., Sharif, M., Javed, M. Y., Muhammad, N., & Yasmin, M. (2018). An implementation of optimized framework for action classification using multilayers neural network on selected fused features. *Pattern Analysis and Applications*, 1–21.

Khan, M. A., Sharif, M., Javed, M. Y., Akram, T., Yasmin, M., & Saba, T. (2017). License number plate recognition system using entropy-based features selection approach with SVM. *IET Image Processing*, *12*(2), 200–209.

Khan, S., Gani, A., Wahab, A. W. A., & Singh, P. K. (2018). Feature selection of denial-of-service attacks using entropy and granular computing. *Arabian Journal for Science and Engineering*, *43*(2), 499–508.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. Paper presented at the Advances in neural information processing systems.

Larose, D. T. (2005). k-nearest neighbor algorithm. *Discovering Knowledge in Data: An Introduction to Data Mining*, 90–106.

Li, G., & Yu, Y. (2016). Visual saliency detection based on multiscale deep CNN features. *IEEE Transactions on Image Processing*, *25*(11), 5012–5024.

Li, Z., Dey, N., Ashour, A. S., Cao, L., Wang, Y., Wang, D., … Shi, F. (2017). Convolutional neural network based clustering and manifold learning method for diabetic plantar pressure imaging dataset. *Journal of Medical Imaging and Health Informatics*, *7*(3), 639–652.

Liaqat, A., Khan, M. A., Shah, J. H., Sharif, M., Yasmin, M., & Fernandes, S. L. (2018). Automated ulcer and bleeding classification from wce images using multiple features fusion and selection. *Journal of Mechanics in Medicine and Biology*, *18*, 1850038.

Liu, X., Gu, J., Xie, Y., Xiong, J., & Qin, W. (2012). *A new approach to detecting ulcer and bleeding in wireless capsule endoscopy images*. Paper presented at the Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference in Hong Kong, China.

Maghsoudi, O. H., & Alizadeh, M. (2018). Feature based framework to detect diseases, tumor, and bleeding in wireless capsule endoscopy. *arXiv preprint arXiv:1802.02232*.

Martis, R. J., Gurupur, V. P., Lin, H., Islam, A., & Fernandes, S. L. (2018). *Recent advances in big data analytics, internet of things and machine learning*. Future Generation Computer Systems. 88, 696-698.

Naqi, S., Sharif, M., Yasmin, M., & Fernandes, S. L. (2018). Lung nodule detection using polygon approximation and hybrid features from CT images. *Current Medical Imaging Reviews*, *14*(1), 108–117.

Nasir, M., Attique Khan, M., Sharif, M., Lali, I. U., Saba, T., & Iqbal, T. (2018). An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. *Microscopy Research and Technique*, *81*, 528–543.

Negi, S. S., & Bhandari, Y. S. (2014). *A hybrid approach to image enhancement using contrast stretching on image sharpening and the analysis of various cases arising using histogram*. Paper presented at the Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India. 2014.

Ramteke, R., & Monali, Y. K. (2012). Automatic medical image classification and abnormality detection using K-nearest neighbour. *International Journal of Advanced Computer Research*, *2*(4), 190–196.

Rashid, M., Khan, M. A., Sharif, M., Raza, M., Sarfraz, M. M., & Afza, F. (2018). Object detection and classification: A joint selection and fusion strategy of deep convolutional neural network and SIFT point features. *Multimedia Tools and Applications*, 1–27.

Raza, M., Sharif, M., Yasmin, M., Khan, M. A., Saba, T., & Fernandes, S. L. (2018). Appearance based pedestrians' gender recognition by employing stacked auto encoders in deep learning. *Future Generation Computer Systems*, *88*, 28–39.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Saba, L., Dey, N., Ashour, A. S., Samanta, S., Nath, S. S., Chakraborty, S., … Suri, J. S. (2016). Automated stratification of liver disease in ultrasound: An online accurate feature classification paradigm. *Computer Methods and Programs in Biomedicine*, *130*, 118–134.

Shah, J. H., Chen, Z., Sharif, M., Yasmin, M., & Fernandes, S. L. (2017). A novel biomechanics-based approach for person re-identification by generating dense color sift salience features. *Journal of Mechanics in Medicine and Biology*, *17* (07), 1740011.

Sharif, M., Khan, M. A., Akram, T., Javed, M. Y., Saba, T., & Rehman, A. (2017). A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropy-based features selection. *EURASIP Journal on Image and Video Processing*, *2017*(1), 89.

Sharif, M., Khan, M. A., Faisal, M., Yasmin, M., & Fernandes, S. L. (2018). A framework for offline signature verification system: Best features selection approach. *Pattern Recognition Letters*.

Sharif, M., Khan, M. A., Iqbal, Z., Azam, M. F., Lali, M. I. U., & Javed, M. Y. (2018). Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Computers and Electronics in Agriculture*, *150*, 220–234.

Sharif, M., Tanvir, U., Munir, E. U., Khan, M. A., & Yasmin, M. (2018). Brain tumor segmentation and classification by improved binomial thresholding and multi-features selection. *Journal of Ambient Intelligence and Humanized Computing*, 1–20.

Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G., Barzi, A., & Jemal, A. (2017). Colorectal cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, *67*(3), 177–193.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint: arXiv:1409.1556*.

Suman, S., Hussin, F. A., Malik, A. S., Ho, S. H., Hilmi, I., Leow, A. H.-R., & Goh, K.-L. (2017). Feature selection and classification of ulcerated lesions using statistical analysis for WCE images. *Applied Sciences*, *7*(10), 1097.

Suman, S., Malik, A. S., Riegler, M., Ho, S. H., Hilmi, I., & Goh, K. L. (2017). *Detection and classification of bleeding region in WCE images using color feature*. Paper presented at the Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, New York, NY, USA.

Sun, Y., Zhao, Y., & Yuan, L. (2018). Quantifying nano-pore heterogeneity and anisotropy in gas shale by synchrotron radiation nano-CT. *Microporous and Mesoporous Materials*, *258*, 8–16.

Sundararajan, K., & Woodard, D. L. (2018). Deep learning for biometrics: A survey. *ACM Computing Surveys (CSUR)*, *51* (3), 65.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). *Going deeper with convolutions*. CVPR.

Yuan, Y., Li, B., & Meng, M. Q.-H. (2016a). Bleeding frame and region detection in the wireless capsule endoscopy video. *IEEE Journal of Biomedical and Health Informatics*, *20*(2), 624–630.

Yuan, Y., Li, B., & Meng, M. Q.-H. (2016b). Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images. *IEEE Transactions on Automation Science and Engineering*, *13*(2), 529–535.

Yuan, Y., Li, B., & Meng, M. Q.-H. (2017). WCE abnormality detection based on saliency and adaptive locality-constrained linear coding. *IEEE Transactions on Automation Science and Engineering*, *14*(1), 149–159.

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

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