Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Review article

A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging

Tariq Rahim, Muhammad Arslan Usman, Soo Young Shin*

Kumoh National Institute of Technology, Gumi, Gyeongbuk 39177, Republic of Korea

ARTICLE INFO

ABSTRACT

Wireless capsule endoscopy (WCE) is a process in which a patient swallows a camera-embedded pill-shaped device that passes through the gastrointestinal (GI) tract, captures and transmits images to an external receiver. WCE devices are considered as a replacement of conventional endoscopy methods which are usually painful and distressful for the patients. WCE devices produce over 60,000 images typically during their course of operation inside the GI tract. These images need to be examined by expert physicians who attempt to identify frames that contain inflammation/disease. It can be hectic for a physician to go through such a large number of frames, hence computer-aided detection methods are considered an efficient alternative.

Various anomalies can take place in the GI tract of a human being but the most important and common ones and the aim of this survey are ulcers, polyps, and tumors. In this paper, we have presented a survey of contemporary computer-aided detection methods that take WCE images as input and classify those images in a diseased/abnormal or disease-free/normal image. We have considered methods that detect tumors, polyps and ulcers, as these three diseases lie in the same category. Furthermore, general abnormalities and bleeding inside the GI tract may be the symptoms of these diseases; so an attempt is also made to enlighten the research work done for abnormalities and bleeding detection inside WCE images. Several studies have been included with in-depth detail of their methodologies, findings, and conclusions. Also, we have attempted to classify these methods based on their technical aspects. A formal discussion and comparison of recent review articles are also provided to have a benchmark for the presented survey mentioning their limitations. This paper also includes a proposed classification approach where a cascade approach of neural networks is presented for the classification of tumor, polyp, and ulcer jointly along with data set specifications and results.

Contents

* Corresponding author.
  E-mail addresses: tariqrahim@kumoh.ac.kr (T. Rahim), arslanusman@msn.com
(M.A. Usman), wdragon@kumoh.ac.kr (S.Y. Shin).

1. Introduction

Physical examination for diagnosing specific diseases in the small intestine leads physicians to suffer from a hectic and time-consuming process. This has led to the development of wireless capsule endoscopy (WCE) technology offering advantages to gastroenterologists for the diagnosis and examination of patients without the execution of complicated surgical operations. Although WCE technology can be used for inspection and assessment of the entire digestive system as an advantage, it can still become a time-consuming procedure, as the complete endoscopic process takes 8 h, generating 60,000 frames on average. So, an early detection of diseases such as tumors, polyps, and ulcers, is important for widening the scope of treatment options.

In 1989, a research group in Baltimore invented WCE that is later commercialized it through Given-Imaging Inc. (Wireless Capsule Endoscopy, 2020; Lin et al., 2005). Table 1 summarizes the companies producing WCE systems, certified by the Food and Drug Administration (FDA); these include MiroCam, IntroMedic Company Ltd., Seoul, Korea; Endocapsule, Olympus America, Inc., Center Valley, Pennsylvania; and PillCam SB2, Given Imaging, Ltd., Yoqneam, Israel. The addition of new technology such as OMOM® Jinshan science of technology capsule offers a good field of view along with the good resolution of images still unapproved by the FDA (Eliakim, 2017). The importance of the WCE device is noticeable as WCE allows the small bowel exploration, where significant fatal diseases can be explored outside and inside the small bowel that makes the necessity of an early automatic detection system. Conventional endoscopic techniques i.e. gastroscopy and colonoscopy are opted for diagnosing GI tract still, the major portion that includes the colon and small intestine remains inaccessible (Usman et al., 2016).

Diseases of the gastrointestinal tract (GI) are caused mainly by persistent or recurring bleed-inducing infections. The extracted frames from a WCE video (Usman et al., 2017), where a small portion of the small bowel is presented showing a yellowish coloration (Usman et al., 2017) along with the faintly pink color of the small intestine. Whenever bleeding occurs, the tissues become red and the affected regions become light pink and purple.

A typical WCE system comprises of three components: (a) capsule endoscope, (b) sensing pads or a belt for attaching the battery pack and data recorder to the patient, (c) a PC workstation having proprietary software. The WCE device shown in Figs. 1 and 2 is a pill-sized swallowable, capsule-shaped electronic device composed of various internally fixed optical and electronic components. The WCE device comprises a camera covering a definite angle of view reliant on the lens. Fig. 1 shows the MiroCam, IntroMedic Company Ltd., Seoul, South Korea, a device having a 170 degree of view (Usman et al., 2016).

Computer-aided technologies such as computer vision particularly offer services for automatic analysis of WCE videos resulting in the decrease of time taken by physicians for processing and examining videos. Current research has focused on two main issues related to the analysis of endoscopic images, namely, the detection and discrimination of malignant tissue. The former involves the detection

of malignant intestinal abnormalities such as tumors, polyps, ulcers etc. mainly caused by bleeding.

The motivation behind this survey is to provide a comprehensive review of the techniques adopted for the detection of tumors, polyps, and ulcers while considering purely wireless capsule endoscopy (WCE) source. Additionally, generalized anomalies found in WCE images followed by bleeding/lesion detection are provided to avoid the insight limitations of research. The focal point is to provide a comparative investigation of the current techniques, thereby creating possibilities for future research in this specific domain.

The rest of the paper is organized as follows. Section 2 describes the abnormality detection in WCE images. Sections 3–5 details a comprehensive description of detection for tumors, polyps, and, ulcers in WCE images respectively. Section 6 details the bleeding detection inside WCE images. Section 7 incorporates a proposed classification approach where a cascade approach of neural networks is presented for the classification of tumor, polyp, and ulcer jointly along with data set specifications and results. Section 8 covers a formal discussion in which attempt is made to enlighten different approaches made for detecting tumor, polyp, and ulcer with comparison to the proposed solution with proper references. Section 9 details some of the recent relevant review articles for WCE imaging. The section also attempts to explain the limitations of these review articles by bench marking it with our presented survey. Finally, the survey paper is wrapped in Section 10, provides a conclusion of the presented survey. Detailed information on the performance of each approach for the detection of the respective abnormalities is presented in tabular form. A diagrammatic view is provided that illustrates the technique made for each disease detection in the survey.

2. Abnormality detection in WCE images

This section attempts to details the work done for the detection of abnormalities in wireless capsule endoscopic images

A unique automatic approach for the detection and localization of the anomalies within the gastrointestinal (GI) from frame sequences of video endoscopic is proposed (Iakovidis et al., 2018). The proposed technique classifies the endoscopic frames into abnormal and normal frames by implementing a weakly supervised convolutional neural network (WCNN) architecture followed by deep saliency detection (DSD). The training of the model is done on weakly annotated images i.e. on image-level rather than pixel-level in the RGB color space. For the salient point detection, the deeper hidden layers of WCNN is used to extract the features maps by DSD algorithm. An iterative cluster unification (ICU) algorithm which is based on descriptor i.e. pointwise cross-feature-map (PCFM) extracted by DSD is used for the localization of GI anomalies. This involves both training and testing stage, wherein the training phase the ICU is trained with both normal and abnormal training images by clustering their salient points. The images classified by WCNN and detected salient points are received by ICU in the testing phase in a unified cluster form. Publicly available two data sets i.e. MIC-CAI comprising of 698 images out of which 465 images are used

Table 1
FDA/non-FDA-approved wireless capsule systems and specifications.

| WCE company/capsule	Size "mm"	Weight "g"	Field of view angle	Frames per second	Battery life "h"	Resolution "pixels"	Communication	Price/capsule
EndoCapsule; Olympus America, Inc., Center Valley, Pennsylvania	Length: 26 Diameter: 11	3.50	145°	2 fps	8 h or longer	512 × 512	Radio frequency communication	$500
PillCam SB3; Given Imaging, Ltd., Yoqneam, Israel	Length: 26.2 Diameter: 11.4	3.00	156°	2 fps or 2–6 fps	8 h or longer	340 × 340	Radio frequency communication	$500
PillCam SB2EX; Given Imaging	Length: 26 Diameter: 11	2.8	156°	2 fps	12 h or longer	256 × 256	Radio frequency communication	$500
MiroCam; Intromedic Co., Ltd., Seoul, Korea	Length: 24 Diameter: 11	3.3–4.70	170°	3 fps	11 h or longer	320 × 320	Human body communication	$500
PillCam ESO2; Given Imaging	Length: 26 Diameter: 11	<4	169°	18 fps	8 h or longer	256 × 256	Radio frequency communication	$500
OMOM® Jinshan Science and Technology	Length: 27.9 Diameter: 13	6.00	169°	2 fps	6–8 h or longer	640 × 480	Radio frequency communication	$250
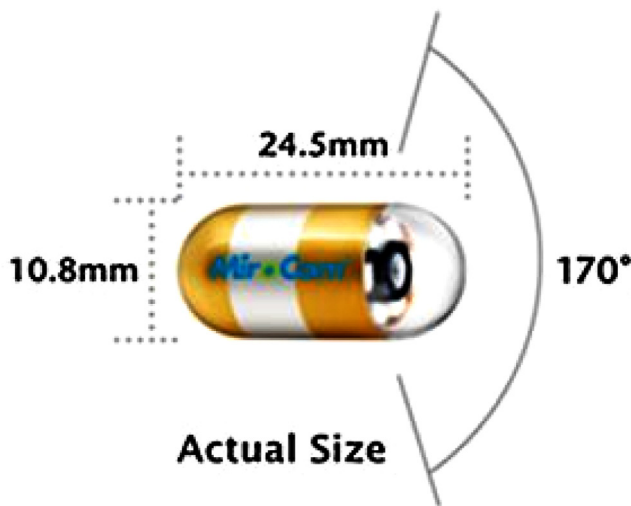
Fig. 1. Wireless capsule endoscopy (WCE) device. (Capsule MIROCAM by Intromedic Ltd.)
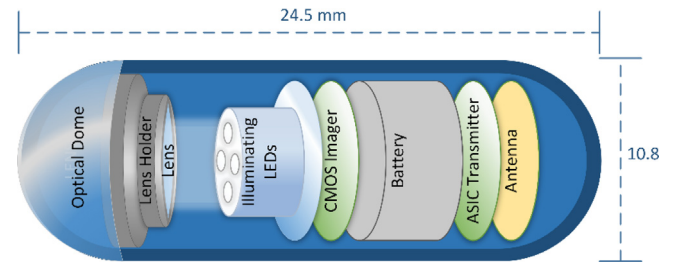

Fig. 2. Components of WCE device (Eliakim, 2017); CMOS: complementary metal oxide semiconductor, ASIC: application specific integrated circuit. (Dimensions specified for MIROCAM capsule by Intromedic Ltd.)

for training and 233 images are used for testing purposes. The second data set used is from KID dataset comprising of 2352 images. The classification performance of the proposed WCNN model is evaluated by utilizing different three diverse learning algorithms generating a high performance for both data sets. For the data set obtained MICCAI, the proposed classification method achieved an accuracy, sensitivity, and specificity of 90.90%, 93.00%, and 88.50% respectively, while for the data set obtained from KID, the corresponding accuracy, sensitivity, and specificity of 89.92%, 92.40%, and 85.58% is achieved respectively (Iakovidis et al., 2018). For the localization of the anomalies, the area under the curve (AUC) with the best performance value for MICCAI and KID is 84.48% and 87.70% respectively is achieved when cluster = 2.

For the detection of abnormalities inside WCE images, a technique i.e. local fuzzy patterns (LFP) is implemented that discriminate frames having normal and abnormal regions (Maghsoudi and Soltanian-Zadeh, 2013). The proposed method is based on two methods i.e. color, geometry (invariant moments), and texture features (GLCM, Gabor, LBP, and Laws' features). The LBP is used for the extraction of color texture features in four different color channels i.e. red, green, hue, and grayscale for the detection of abnormal and normal regions. To overcome with limitations of LBP, the LFP is implemented to extract features as unlike LBP, the LFP is handling the idea of partial truth i.e. completely false and completely true. The data set comprised of 44 normal and 100 abnormal for the analysis by maintaining the parameters for both LBP and LFP constant. The proposed approach for abnormalities regions detection inside

WCE images generated results in terms of accuracy, sensitivity, and specificity of 90.00%, 92.00%, and 91.00%, respectively (Maghsoudi and Soltanian-Zadeh, 2013).

Furthermore, an automatic technique where splitting of WCE image into several patches is done for the extraction of features. The speeded-up robust features (SURF) are extracted from each pixel of the image instead of utilizing the blob nature of SURF algorithm in CIELAB and "YCbCr" color space. To overcome the manual crafting of the feature vector of SURF interest points and lessening the burden of a large number of WCE images crafting, CNN is implemented that learn patterns of textural color and extract features from different regions of WCE image (Sekuboyina et al., 2017). The CNN gets the patches instead of the whole image and to avoid redundant patched information leading to overfitting, the WCE frame is spliced into nonoverlapping patches based on a percentage of malignant or benign pixels. Using a trained model, classification is achieved followed by assignment of each pixel to their respective patch. The proposed approach for the training data achieved sensitivity and specificity of 92.00% and 97.00% respectively while for testing data the sensitivity and specificity achieved is 21.00% and 98.00% respectively in CIELAB color space. The overall area under the curve (AUC), sensitivity, and specificity for all considered diseases in the specific study are 79.61% to 83.00%, 71.00% to 90.00%, and 72.00% to 75.00% respectively outperforming the compared approaches (Sekuboyina et al., 2017).

Moreover, for the discrimination of pathogenic endoscopic frames associated to ulcer within the gastrointestinal tract (GT), a unique method based on the bidimensional ensemble empirical mode decomposition (BEEMD) (Charisis et al., 2010) is implemented for the extraction of their intrinsic mode functions (IMFs). The BEEMD (Wu et al., 2009) which is an extension of EEMD that pursue and reveal any kind of inherited intrinsic oscillatory modes, named IMF by remaining in the time domain. The primary step for the proposed approach is the specification of the region of interest (ROI) which is manually done by cropping in the proposed study followed by a selection of RGB color space and channels extraction. Due to the nature of WCE images generation, noises are eliminated by applying BEEMB on each channel of color followed by decomposition of each image into eight IMFs per channel. Then for the quantification and extraction of textural features, lacunarity analysis (Lac) is applied on each channel of the RGB color space. The lacunarity window size curve is generated to examine the capability of classification of WCE images. Discriminant analysis-based classification and support vector machine (SVM) is used for classification. The total data set comprised of 80 images, out of which 40 pathogenic and 40 normal images. For both classifiers, the proposed approach (Charisis et al., 2010) is achieving high performance in the green channel, while for overall performance as benchmarked, the SVM outperforms by achieving an accuracy, sensitivity, and specificity of 95.75%, 95.00%, and 96.50% respectively.

Furthermore, a novel technique integrating deep convolutional neural network (DCNN) and geometric features for the detection and classification of GI tract diseases in WCE images is presented. A novel approach i.e. contrast-enhanced color features (CECF) and feature extraction via CNN is proposed for the extraction of lesion region (Sharif et al., 2019). The CECF gets along with two sub steps: (1) enhancing the local contrast of infectious region via hybrid contrast stretching method that comprised of median filter and top-bottom hat filtering; extraction of color features via hue, saturation, variation (HSV) transformation that will be used for similarity measures among pixels of HSV color image. The stretching phase consisting of top-hat filter enhances foreground contrast while bottom hat filter enhances background contrast removing unnecessary information and controlling global contrasts of the image. For the lesion detection CIELAB color transformation is done on the median filtered images for the extraction of color features

such as variance, standard deviation, mean, kurtosis, and skewness. These color features are then integrated in a matrix and mean deviation is computed by which weighted value is generated for making a cluster of the features extracted. Lastly, a conditional probability dependent threshold value is generated that is used for segmentation. Extraction of geometric and CNN features from improved RGB and segmented images is done. Pretrained model of VGG16 and VGG19 are used for the extraction of deep features from WCE images followed by 3×3 Max pooling operation for the selection of features with higher values and information. Then a unique method i.e. Euclidean Fisher vector (EFV) is proposed for the fusion of features extracted from both trained models. In the EFV method, fisher vector-based similar features are ordered into one separate cluster and other features are ordered into the different cluster. EFV method provides a unique feature vector which is fused parallel latter in one matrix. Following, the conditional entropy strategy is implemented on a fused vector for the selection of best features where features greater than 0.4 are selected for final classification. For the classification, the final selected features become the input of fine K-nearest neighbor (FKNN) (Larose and Larose, 2014). The proposed method is implemented on total of 4500 frames, out of which 2000 bleeding and health frames respectively, while 500 are ulcer images. For the proposed fused features approach, the KNN generated a highest performance in terms of accuracy, sensitivity, precision, and specificity of 99.42%, 100%, 99.51%, and 100% respectively (Sharif et al., 2019).

Furthermore, for the automatic detection of patterns in wireless capsule endoscopy (WCE) images, an integrated image processing with recognition analysis is presented. In particular, the method performed consists of a selection of techniques such as image registration, segmentation, local-global (L-G) graphs, region synthesis, and neural nets merged into a synergistic fashion for obtaining the desired outcomes (Bourbakis, 2005). For the image registration phase, the genetic algorithm optimization tactic is applied to finding the optimal registration parameter within the global search space. The region-based segmentation is achieved by applying fuzzy-like reasoning segmentation (FRS) technique which has three phases i.e. smoothing, edge detection, and segmentation. In the smoothing phase, a low pass filter will eliminate noise in a selective manner that preserves edges of the location of the image. The segmentation phase comprises of flood fill operations for the interior points of images that decide based on color similarity in the adjacent pixels to be filled or not. For the accurate representation and interest points, the L-G graph technique is adopted avoiding non-linear graph matching function. The recognition accuracy of the object can be improved by integrating the FRS with the L-G graph. A neural nets-based recognition developed by the ATL research lab is used for the detection purpose and the results are shown in a figure wise in the proposed work lacking the performance metrics (Bourbakis, 2005).

For addressing the problem related to GI walls that deforms in erratic way, a tracking technique for WCE images using triangular constraints via an affine transformation is proposed. The motivation is that the location of the target can be determined in the successive frames transition by using affine transformation (Yanagawa et al., 2017). The proposed approach after tracking the abnormal region both in backwards and forwards direction involves three stages. First, matching within successive frames for feature points in the abnormal (target) and surrounding (supporters) are done via SIFT/SURF and Kanade-Lucas-Tomasi (KLT) method (Tomasi and Detection, 1991) respectively. Second is the estimation of the target position is done based on voting process via affine matrix which is computed from triplet supporters. Adjustment of the relative location of the target is done by mapping of triangles among current and successive frames. Third, precise position determination of abnormal region is done, and to avoid the

error induced by the movement of successive frame; color information of the target is utilized. The proposed abnormality tracking approach is applied on 120 sequences at frame rate 2 fps having eight major abnormalities. The proposed method (Yanagawa et al., 2017) for abnormality tracking outperformed other efficient tracking algorithm such as mean-shifted technique (Collins, 2003) and the multi-domain network (MDNet) (Nam and Han, 2016) technique.

Table 2 summarizes the survey for abnormality detection in WCE images in terms of the performance of the respective methods. The accuracy, specificity, sensitivity, and precision for the abnormality detection by the respective approaches are shown in the table.

3. Tumor detection in WCE images

This section attempts to explain the work done for the detection of tumors in WCE images by pointing out both machine learning and other image processing approaches.

In order to characterize WCE images into abnormal, normal and tumor categories, an adaptive neuro-fuzzy interface system (ANFIS) is implemented (Alizadeh et al., 2017). The technique subsequently segregates images based on statistical analysis by following two steps i.e. extraction of the features and classification of the images based on 32 features. The features also include four statistical factors, i.e. contrast, homogeneity, correlation, and energy, obtained from co-occurrence metrics. Minimum redundancy among the features is achieved via mutual information selection of the features with maximum dependence on the target class generating a classification accuracy, specificity, and sensitivity of 94.20%, 96.27%, and 94.16% respectively.

Furthermore, a two-step automatic tumor detection procedure i.e. Region of Interest selection and classification is proposed inside WCE images. The detailed pipeline of the proposed approach comprises pre-processing, segmentation, feature extraction, and classification (Vieira et al., 2020). In the pre-processing stage, color conversion from "RGB" to "CIELab" color space is done where "a" and "b" channels are selected for segmentation. The automatic detection of ROI by employing a segmentation module i.e. maximum a posteriori (MAP) approach and the expectation-maximization (EM) for measuring the parameters of a multivariate Gaussian mixture model (GMM) is done. For convergence acceleration of the expectation-maximization (EM), a modified version of the Anderson algorithm is proposed. For the feature extraction histogram-based, statistical features such as variance, kurtosis, mean, and entropy are extracted. The second step utilizes the ensemble learning (EL) system grounded on SVMs where bagging tactic is utilized for obtaining a global classification. In the case of ES, the variety which ensures that per member of the ensemble is tuned for a subset of the training set can be obtained by using bootstrapping where the training data is randomly sampled with replacement. Two GMM are utilized i.e. one for normal data clustering (M clusters) and one for clustering abnormal data (N clusters) and combined into $M \times N$. Every subset is used to train a least squares-SVM classifier. At this point, the gating network i.e. MLP is trained for decoupling the collective maximization of both modules and same process done for testing step. The data set consisted of total 3936 frames out of which 3000 normal and 936 are pathological frames and are trained on WEKA for each ensemble elements (SVMs). The proposed technique of extracting features via ensemble classifier (Vieira et al., 2020) generated a high accuracy, specificity, sensitivity, and area under the curve (AUC) of 96.30%, 96.70%, 95.00%, and 95.10% respectively that also outperforms features extracted by wavelets and curvelets transforms in more than 5%.

For the detecting of small-bowel angioectasia, a deep learning-based approach of the CNN system is implemented. The CNN system based on single shot multibox detector (SSD) is trained where results are evaluated in terms of area under the receiver operating characteristic curve (ROC-AUC), specificity, sensitivity, positive and negative predictive values (Tsuboi et al., 2020). The deep neural network SSD unaltered is fed with annotated images via a Caffe deep learning framework having stochastic gradient descent and learning rate of 0.0001 for the proposed approach. Using the probability score ranging 0–1 where higher score mean high confidence, the capability of CNN to discriminate between image having angioectasia or not is evaluated. The data set is acquired at Hiroshima University Hospital, the University of Tokyo Hospital, and Sendai Kousei Hospital comprising a total of 12,685 WCE images out of which 2237 are used for training and 10,488 for testing purpose. The data set was partitioned into 12,237 normal small bowel images and 448 of images having angioectasia. The proposed approach for detecting ulcers inside WCE images generated an accuracy with a cutoff probability of 0.36 in terms of AUC, sensitivity, and specificity of 99.80%, 98.80%, and 98.40% respectively (Tsuboi et al., 2020).

To discriminate tumor from normal images, a unique set of features based on the texture that assimilates a multi-scale curvelet is operated in combination with fractal technology (Liu et al., 2016a). Executing a transformation known as inverse curvelet of certain scales, texture descriptors are computed of second and high order moments for different color channels. Genetic algorithm (GA) and support vector machine (SVM) is implemented in order to optimize set of features and classification respectively. The generated results in comparison with other approaches showed better performance in terms of sensitivity, specificity, and accuracy of 97.80% and 96.70%, and 97.30% respectively. Moreover, out of those 89 clinical patients under the study, 16.8% (15) were diagnosed (Liu et al., 2016a).

Due to the diverse appearances and shapes of images captured by capsule endoscopy, a unique textural feature based on wavelet and the multi-scale local binary pattern is implemented in (Li et al., 2011). Three classifiers i.e. multi-layer perception neural network (MLP), k-nearest neighbor, and SVM are implemented to facilitate classification and distinction of the diverse tumor images. The data set consists of 1200 images obtained from ten patients are used for this study resulting in detection accuracy, sensitivity, and specificity of 90.50%, 92.33%, and 88.67%, respectively.

A wavelet decomposition i.e. color wavelet co-variance (CWC) is implemented that extracts color features for the detection of tumors in endoscopic videos (Karkanis et al., 2003). Based on second-order textural computation of the co-variances, features are determined followed by optimal subset selection of features via the selected algorithm. Classification is achieved by linear discriminant analysis (LDA) in support of the proposed algorithm on color colonoscopic videos generating a high specificity and sensitivity of 97.00% and 90.00%, respectively (Karkanis et al., 2003).

Furthermore, an integrated approach of using discrete wavelet transform (DWT) and singular value decomposition (SVD) algorithm are implemented for tumor detection (Faghih Dinevari et al., 2016). Using DWT diversities in the appearance and illumination environment (which generally change with respect to time) are coped with. Finally, SVM is applied for classifying the WCE images to a data set constituting 400 normal and 400 abnormal WCE images. The integrated approach achieved a sensitivity, specificity, and accuracy of 94.00%, 93.00%, and 93.50%, respectively.

Furthermore, a computer-aided diagnostic technique for detection of small bowl tumor inside WCE images is presented in Li and Meng (2009a). The work take advantages of integrating wavelet transform and textural feature called local binary pattern as extractor in order to discriminate normal regions from tumor regions.

Table 2
Efficacy of existing methods for Abnormality detection in WCE images.

Ref.	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Iakovidis et al. (2018)	90.90	88.50	93.00	No
Maghsoudi and Soltanian-Zadeh (2013)	90.00	91.00	92.00	91.70
Sekuboyina et al. (2017)	No	75.00	90.00	No
Charisis et al. (2010)	95.75	96.50	95.00	No
Sharif et al. (2019)	99.42	100	100	99.51
Bourbakis (2005)	No	No	No	No
Yanagawa et al. (2017)	No	No	No	No

Implementing uniform local binary pattern for multi-resolution and rotation invariant textural features are extracted followed by three level of DWT on each channel of WCE image in "RGB" and "HSI" color spaces. For classification, support vector machine (SVM) is implemented on data set of 300 images obtained by PillCam SB2 yielding an accuracy, specificity and sensitivity of 96.67%, 96.00% and 97.33% in HSI color space (Li and Meng, 2009a) showing that efficiency of usage of wavelet based local binary pattern approach.

An integrated computer-aided scheme combining the textural features, uniform local binary pattern, wavelet, and SVM is implemented for the detection of tumors in the digestive tract (Li and Meng, 2012b). The local binary pattern is a descriptor of a local texture that defines the intensity distribution and is unaffected by illumination variances and offers multi-resolution of WCE images. SVM used for classification; nevertheless, the selected features do not guarantee high performance in terms of classification. Therefore, two additional feature selection tactics were implemented namely, sequential forward floating selection based on SVM (SVM-SFFS) and recursive feature elimination based on SVM (SVM-RFE) (Jain and Zongker, 1997). In the SFFS approach, some steps are made in reverse after the forward steps (Guyon et al., 2002) while in SVM-RFE, backward feature removal is used to rank the features based on the weight that is a function of the support vectors (Li and Meng, 2012b). This two-feature selection aided in improving the detection capability. The proposed technique was validated by conducting a wide range of experiments, resulting in accuracy, sensitivity, and specificity of 92.40%, 88.60% and 96.20% for recognition of the tumors in WCE images.

As mentioned, visualization of the entire GI tract is important and can be achieved by WCE. Human textural perception is dependent on the multi-scale investigation of patterns and was simulated using multi-resolution tactics (Barbosa et al., 2009). Moreover, the covariance of textural descriptors was successfully used in the classification of colonoscopy videos. A frame classification pattern that is dependent on the statistical textural features obtained by applying the discrete curvelet transform (DCT) domain was suggested (Collins, 2003). To address the limited directional sensitivity of DWT, DCT was used as it offers high directional sensitivity in many directions and is anisotropic in nature. These properties make DCT the best option for dealing with pattern complexities such as texture. The classification features obtained from the textural co-variance at multiple angles were applied to the proposed color curvelet co-variance scheme. The mean and standard deviation of the coefficients of DCT was the best descriptors of the textural features in this case (Barbosa et al., 2009). The classification is achieved by implementing an (MLP) neural network, resulting in a sensitivity of 97.20% and specificity of 97.40 %.

Due to the variations in the appearance, luminance, texture of the images of the intestine and variation of the shape among normal tissues and hemorrhages computerized detection technique becomes a hectic task. Owing to the astonishing number of image analysis errors, the convolutional neural networks (CNNs') model is preferred and has been used for the recognition of intestinal hemorrhage (Li et al., 2017). CNN models include LeNet, AlexNet, GoogLeNet, and VGGNet that offer variation in the format of the input, depth, and modules are explored in this study. Data augmentation using various transformations has been used to prevent over fitting, which is a severe problem when dealing with CNN models. The correlation between the image quality and detection accuracy of intestinal hemorrhage was also investigated. The CNN model is applied on a dataset comprising of 1300 of hemorrhage images and 40,000 normal images resulting highest precision for AlexNet of 98.06% and F-measure for VGG-Net of 98.87% showing the validity of the different models (Li et al., 2017).

A textural feature extraction technique followed by SVM to classify tumor/abnormal from normal images. The pre-processing is done by median filtering to denoise the images followed by DWT as feature extractors due to its multi-orientation characteristics. The textural features (mean and energy) are computed for diagonal components as a mean, while energy is considered for horizontal and vertical parts. These features are used for training the support vector machine recursive future elimination (SVMRFE) classifier that has the advantage of avoiding selection duplicate values in the extracted features (Ashokkumar and Subramanian, 2014). This method was applied to a small number of images, with 100% accuracy in detecting tumors in WCE images.

A novel approach that exploits the motion of animals i.e., mice, which is archetypal of neuroscience is used to track and detect frames having tumors, bleeding pixels, and other abnormal images. Quantification of the kinematics of running rodents is necessary in case of high frame rates (250 Hz). For this purpose, an automated technique is necessary to track the paws of animals, where the method constitutes the following steps: color-based segmentation, followed by classification using SVM plus neural network (NN) (Maghsoudi and Alizadeh, 2018). By coupling the kinematic features of running rodents and textural features, further classification and identification are performed to ensure precise analysis of the labeled paws. For detecting the frames with tumors in the dataset, features such as the geometry are extracted from the central region followed by Gabor filter for segmenting the features. Extraction of 22 gray level co-occurrence matrix (GLCM) features, 4 statistical features (mean, kurtosis, variance, and skewness), and seven invariant moments are done, leading to 990 features. Moreover, extraction of 75 Law's features, including the entropy, skewness, variance, mean, and kurtosis of 15 images is achieved by convoluting the images with Law's masks. Besides, the extraction of 88 features from GLCM at 4 varied angles along with the calculation of 7 invariant moments is done resulting in a total of 1160 features. Due to the use of dissimilar extraction techniques, normalization is performed. For recognition of the frames with tumors, thirty high-discriminant features are obtained by applying the Fisher test. Thereafter, for the classification of the WCE frames, multi-layer perceptron (MLP) neural network (Mohapatra et al., 2012) is implemented.

After detection of frames, sub-division of the images is performed and the local binary patterns LBP1 and LBP2 are used to extract 110 textural features, comprising 74 from LBP1 and 36 from LBP2. Then, extraction of the GLCM features at four multi-angles, and mean feature extraction is done making a total of 202 features. For seven samples, 21 masks having two dimensions are obtained

by implementing Law's kernels for five features, generating 105 features. The kurtosis, skewness, mean, variance, and entropy features are extracted via implementing eight Gabor filters at four angles and two frequency levels to the sub-images, resulting in 50 features in this step. The HSV color space is commonly used for the detection of objects in medical imaging. The kurtosis, skewness, mean, and variance were extracted for five color channels, thus generating a total of 381 features. The aforementioned normalization was performed, resulting in a reduction of features to 30, and subsequent implementation of the Fisher test enabled distinguishing of the chosen features into abnormal and normal regions within a frame (Barbosa et al., 2009). For distinction of tumors, other abnormalities, and bleeding from normal tissue, NN was used to classify the regions into normal and abnormal. Increasing the features while applying the Fisher test produced a direct relationship: a sensitivity of up to 1 and specificity of 0.928. The second technique for recognizing regions with abnormalities gave an accuracy for tumor, bleeding, and abnormality detection of 0.9092, 0.9747, and 0.9461, respectively, and a corresponding precision of 0.8945, 0.9465, and 0.8568; sensitivity of 0.9273, 0.9733, and 0.9671; and specificity of 0.9029, 0.9754, and 0.9381, respectively. For the recognition of tumors only, the second technique gave a specificity and sensitivity up to 0.9011 and 0.9263, respectively, for higher features (Maghsoudi and Alizadeh, 2018).

Textural analysis and representation are preferable while detecting tumors in the small intestine using endoscopic images. In the specific approach, wavelet transform is used to select the bands having the most important textural information for different color channels. From the chosen wavelet coefficients for each color channel, the feature set is computed as a co-occurrence matrix (Barbosa et al., 2008). These statistical descriptors are largely based on human insight and judgment of textures. MLP neural network called back propagation learner having 24 input neurons with 6, 8, or 12 hidden layers are used. For classification, two output layers are utilized to categorize the images into normal and tumor. The feasibility of the proposed technique is proven by application to capsule endoscopic data obtained from humans, where the sensibility and sensitivity achieved are 98.70% and 96.60%, respectively (Barbosa et al., 2008).

In order to detect tumors and colonic polyps in WCE images, an NN-based approach is presented for extraction of the textural features. Extraction of the features is based on scale invariant feature point (SIFT), which is used not only for a single key point but also for key points in the neighborhood (Sindhu and Valsan, 2017b). The Haralick texture features are extracted for individual 16×16 patches near key points. Combination of SIFT and the Haralick texture features resulted in successful extraction of the features. For classification neural network (NN) is trained on extracted features in order to accurately detect tumors or colonic polyps. The proposed technique achieved an overall accuracy level of 97.50% with a specificity and sensitivity of 93.40% and 98.80%, respectively, for the detection of polyp vs. tumor vs. normal images. In the case of tumor vs. normal detection, the accuracy was 95.10%, while the specificity and sensitivity were 92.10% and 96.50%, respectively.

A similar approach is used for detecting several inflammatory bowel diseases (IBDs), including Crohn's disease, ulcers, tumors, and other abnormalities. Similar to the former approach (Sindhu and Valsan, 2017b), extraction of the features is computed not only for a single key point but also for key points in the neighborhood via integrated SIFT and Haralick texture features implementation. A GLCM matrix having dimensions of 16×16 is generated, and an MLP neural network is then used as a classifier using supervised learning (Sindhu and Valsan, 2017a). The results obtained from this method prove the feasibility of the proposed technique, with accuracy values of 96.90%, 96.40%, 98.90%, and 97.40% for the detection of Crohn's disease, ulcers, polyps, and tumors, respectively. In addi-

tion, the overall accuracy of the proposed technique achieved is 88.60%.

Moreover, an automatic small bowel tumor detection technique is proposed by extracting features in CIELab color space. The discrimination of tumors from normal tissues is done via color histogram information in which a light-saturated region is categorized as a tumor (Vieira et al., 2015). Maximum a posteriori approach by using the expectation-maximization algorithm is used for the segmentation process, where parameters of a multivariate Gaussian mixture model (GMM) are estimated. To lessen the influence of light in medical images, CIELab color space is used where L channel is discarded remaining a and b channels. Five histogram-dependent statistical features i.e. mean, variance, entropy, kurtosis and the value reached by the cumulative function for 95.00% of the data are extracted. A total of 3200 frames are used for the analysis in which 700 and 2500 frames are a tumor and normal frames respectively. For classification multi-layer perception neural network (MLP) and a support vector machine (SVM) with cross-validation of 10-fold are implemented. The proposed detection technique using feature grouping of mean and variance over ROI resulted in accuracy, sensitivity, and specificity of 98.80%, 98.50%, and 98.80% respectively (Vieira et al., 2015).

Table 3 sums the survey for the detection of tumors in WCE images in terms of their respective performances for each method. The performance metrics such as accuracy, specificity, sensitivity, and precision for the detection of tumors by the respective approaches are listed in the table.

4. Polyp detection in WCE images

Apart from tumors, the detection of other pathological defects such as polyps, ulcers, and Crohn's disease is also essential for timely diagnosis. A unique technique for the detection of polyps and ulcers in the perforated form in WCE frames is attempted (Karargyris and Bourbakis, 2011). Polyps and ulcers share the common property of having different perceptible patterns and geometries, making them easily noticeable by experts. However, these abnormalities have different colors and different sharp edges, although converse to ulcers, the color of polyps does not vary from that of the normal digestive tract and wall, which makes it hard to extract color features in this case. Detecting polyps in such a scenario, as well as in WCE images where polyps can occur anywhere, is thus a difficult task. The preprocessing involving segmentation should achieve the goals of preserving the boundary information of the images and extraction of sharp segments, making it an important step. The log Gabor filter and Susan edge detector are thus used for segmentation and edge detection. The Gabor filter has the property of increased spectral information, Gaussian response and the ability for bandwidth optimization making it the best candidate for segmentation (Field, 1987). The smallest uni-value segment assimilating nucleus (SUSAN) edge detector is used to further improve the technique and the preprocessing step. Extraction of the geometric features is performed subsequently to obtain the output of the log Gabor and SUSAN edge process for detection. The color space and its transformation are important features for detecting ulcers in WCE images. HSV is recommended for medical imaging analysis, and transformation of RGB into HSV is preferred. The fuzzy region-growing tactic is the basis of the segmentation structure. Smoothing was used to remove noise from the input image while preserving the locality of edges in the areas where local contrast was inevitably low. A supervised learning technique i.e., SVM, was used to take advantage of the property of mapping inputs in n-dimensional space by finding the optimum hyperplane for separation of datasets. In addition to extraction of the pattern of ulcers, features such as the statistical texture were also computed

Table 3
Efficacy of existing methods for tumor detection in WCE images.

Ref.	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Alizadeh et al. (2017)	94.20	96.27	94.16	No
Vieira et al. (2020)	96.30	96.70	95.00	No
Tsuboi et al. (2020)	99.80	98.40	98.80	No
Liu et al. (2016a)	97.30	96.70	97.80	No
Li et al. (2011)	90.50	88.67	92.33	No
Karkanis et al. (2003)	No	97.00	90.00	No
Faghih Dinevari et al. (2016)	93.50	93.00	94.00	No
Li and Meng (2009a)	96.67	96.00	97.33	No
Li and Meng (2012b)	92.40	96.20	88.60	No
Barbosa et al. (2009)	No	97.40	97.20	No
Li et al. (2017)	No	No	No	98.06
Ashokkumar and Subramanian (2014)	100	No	No	No
Maghsoudi and Alizadeh (2018)	90.92	90.11	92.63	No
Barbosa et al. (2008)	NO	NO	96.60	NO
Sindhu and Valsan (2017b)	95.10	92.10	96.50	No
Sindhu and Valsan (2017a)	88.60	90.70	72.50	No
Vieira et al. (2015)	98.80	98.80	98.50	No

(Haralick et al., 1973) in word format with 14 values. This word pattern was introduced into the fuzzy support vector machine as an input in order to classify the input pattern (Inoue and Abe, 2001). Fifty frames of WCE video comprising 40 normal and 10 polyp frames, respectively, were subjected to the algorithm for calculating the sensitivity and specificity, resulting in a high sensitivity of 100% and low specificity of 67.50%. The results demonstrate the feasibility of the technique for detecting only polyp and ulcers, independent of tumor detection (Karargyris and Bourbakis, 2011).

Furthermore, a technique for the detection of polyp based on feature extraction has proposed that discard frames having the region of none-interest inside WCE videos. A sliding window of size 227×227 on the frame is slid from top to bottom and left to right generating window images (Billah et al., 2017). Using color wavelet (CW) transform, textural features are extracted due to the variation in polyp appearance. The co-occurrence matrix was computed in four different directions in order to find spatial relationships. Additional features like blobs and edges are extracted with a window size of 227×227 using a convolutional neural network (CNN) stimulated by the work done in Mesejo et al. (2016). Using multi-class SVM that provide sparse, noisy data and a fast linear solver, classification is performed on extracted features from window image. Compared to other techniques, the proposed technique showed better performance in terms of accuracy, sensitivity, and specificity, with values of 98.65%, 98.79%, and 98.52%, respectively (Billah et al., 2017).

Furthermore, a computer-aided polyp detection technique is proposed based on the amalgamation of features extracted from color wavelets and convolutional neural network (CNN) (Billah and Waheed, 2018). The multi-resolution analysis in the RGB color space by wavelet transform is done because of variation in the shape and size of polyps. Statistical features such as homogeneity, correlation, entropy, and energy are computed resulting in 144 color wavelet features. The CNN model at the beginning layers of the network extracts primitive features i.e. blobs and edges that are processed further in the deeper layers for obtaining a high level of image features. All the features extracted from both color wavelet and CNN model are fed to support vector machine (SVM) for the classification purpose. The data set comprised of 14,000 images in which 3500 and 10,500 images consisted of polyps and non-polyps respectively. The proposed features fusion scheme generated a high performance in terms of accuracy, sensitivity, and specificity of 98.34%, 98.67%, and 98.23% respectively (Billah and Waheed, 2018).

Moreover, an integrated approach of the Gabor filter and monogenic-local binary pattern (M-LBP) is implemented in order to detect polyp inside 55,000 images obtained from one patient (Yuan and Meng, 2014). By exploiting the excellent spatial local-

ity and selection of orientation provided by the Gabor filter, multi frequency textural information is extracted. To avoid over-simplification of the local image shapes, M-LBP (Zhang et al., 2010) was used in combination with LBP as a rotation invariant technique for each response image of the Gabor output. M-LBP comprises LBP operator information and a local surface and local phase. Dimension reduction is an important step to deal with a large number of elements and aids in the selection of the features that are the most important for the classification of polyps. Application of a supervised algorithm termed local discriminant analysis (LDA) (Karargyris and Bourbakis, 2011) is performed in order to achieve dimension reduction of the features. For classification, SVM was applied to a set of 872 images, half of which contained polyps and half comprising normal tissue images. The proposed technique resulted in a detection accuracy, sensitivity, and specificity values of 91.43%, 88.09%, and 94.78%, respectively.

Furthermore, computerized detection of polyps in WCE images is performed based on the unique textural features obtained by combining wavelet and uniform local pattern (Li and Meng, 2012a). Using SVM, classification is achieved. The LBP (Ojala et al., 1996), which is a textural operator robust to any kind of transformation extended (Ojala et al., 2002) to achieve wider multi-resolution characteristics for textural descriptors. The separation of operators (Ojala et al., 2002) is done defined by parameters such as (P) and (R) which control quantization of the region in the angular domain and radius of a circle that controls the spatial resolution of this descriptor (Li and Meng, 2012a). DWT is implemented for multi-resolution texture analysis at different sub-energy level bands by integrating multiple operators having different values of (P) and (R). During the experiment, cross-validation was performed to avoid overfitting (Li and Meng, 2012a). The classification accuracy achieved in detecting polyps with proposed technique by integrating the features extracted from different sub-image levels compared to a single level is high. By implementing (P) and (R) with values of 24 and 3, respectively, the classification accuracy in RGB color space achieved is 91.60%, while in HSI color space it is 91.0%, showing the performance of technique (Li and Meng, 2012a).

For detecting polyp inside WCE images, a unique technique termed sparse auto-encoder with image manifold constraint (SSAEIM) is proposed (Yuan and Meng, 2017). Although the shape features implemented in Iwahori et al. (2015) and Bae and Yoon (2015) and (HOG) (Dalal and Triggs, 2005) are all-inclusive, still they neglect intrinsic information. The discriminative model results in the suggested technique robust and accurate for distinguishing polyps from normal images (Yuan and Meng, 2017). SSAEIM for polyp detection was stimulated by sparse autoencoder (SAE), which is an unsupervised learning network for used for learning

the features of images automatically. To avoiding pitfalls such as neglecting intrinsic information, and in order to achieve a discriminative approach, implementation of the image manifold model in combination with theories such as graph and the latest manifold learning is preferred (Ma et al., 2015). For such modification, SAE can be altered by integrating the image manifold results (SAEIM), which ensures similarity of the features in the images of a given category. Compared to simple neural networks, the deep neural network provides better performance for image characterization. The stacking of several SAEIM layers was employed in the proposed method, in which three SAEIM layers were stacked. Image resizing from 256×256 to $256 \times 256 \times 3$ was undertaken to reduce the computational complexity, which changed the size of the first layer to 12,288 units, followed by 6000 and 1200 units for the second and third layers, respectively. The output layer is the fourth layer comprising 128 units, which produced a discriminative model capable of characterization of the features at a high level. In every layer of SSAEIM, image manifold was applied to maintain high inter-variances and low intra-variances between features. The output of the SSAEIM model involves characterization of each wireless capsule image with dimensions of 128 as a feature, followed by the use of a supervised learner called Softmax classifier for polyp detection. In the model, five parameters (the sparsity penalty control (β), image manifold constraint (γ), weigh decay cost (α), learning rate (μ), and sparsity parameter (ρ)) were set to obtain the best results (Yuan and Meng, 2017). Evaluation of the SSAEIM technique was performed with 3000 normal wireless capsule images each having 1000 of the above three categories. The overall recognition accuracy (ORA) of the proposed computational technique for analysis of the wireless capsule images was 98.00%, while for each category (polyp, clear images, turbid, and bubbles) the accuracy was found to be 98.00%, 95.50%, 99.00%, and 99.50%, respectively (Yuan and Meng, 2017).

In contrast with the outdated bag of feature (BoF) technique, an improved bag of features (BoF) is implemented for classifying polyps inside WCE images (Yuan et al., 2016). Features such as the complete local binary pattern (CLBP), visual words of the length of 120, 8×8 path size, and SVM are used in line in the proposed technique for automatic detection of polyps in WCE images. The key points, called salient points having useful information is the focus of the proposed technique, around which extraction of the features are done (Yuan et al., 2016). The SIFT features are extracted due to their rotation and scale in variance (Guo et al., 2010). The five features considered are the SIFT (Lowe, 2004), LBP (Ojala et al., 1996), uniform LBP (Ojala et al., 2002), CLBP (Guo et al., 2010), and HOG (Dalal and Triggs, 2005). Feature integration was performed after acquiring the above textural features by defining the key point as p_i and (x, y) using tactic, where p_i and (x, y) define the pixel and location in the original image, respectively. By selecting a region having a given patch size denoted by p_i, the SIFT + LBP descriptor can be generated, represented by the combined vector that defines the whole patch.

$$SIFT + LBP = [SIFT_i, LBP_i] \qquad (1)$$

Conversion of vector patches into a visual of words is done in order to produce a vocabulary by applying K-means to the data set. After that, the combined features of the testing and training images are coded by assigning the closest visual words that form a histogram, which later becomes the input for the classifier. The improved BoF technique is tested using two classifiers, i.e., SVM (LibSVM (Chang and Lin, 2011) and FLDA (Bishop and Nasrabadi, 2007). A total of 2500 WCE images used, of which 2000 were normal and 500 were polyp images. The training is performed with 1000 normal images and 250 polyp images, and the remaining 1000 normal images were used for testing with 250 polyp images (Yuan et al., 2016). The analysis is performed based on patch sizes of all dimensions for all combinations of features, i.e., [(SIFT + LBP), (SIFT + uniLBP), (SIFT + CLBP and (SIFT + HOG)], and implemented using two classifiers, i.e., SVM and FLDA. The results achieved by using these combinations and classifiers showed that by using SIFT integrated with CLBP and SVM as a classifier, a specificity of 90.88%, sensitivity of 94.54%, and accuracy of 93.20% is achieved, indicating performance superior to that in recent work done in Li and Meng (2012a).

For the differentiating normal and polyp images inside WCE images and integrated approach based on histogram chromaticity and Zernike moment is proposed. The Zernike moment which is invariant to rotation, scale, and translation is used as a feature extractor in HSI color space. For the representation of the color features, a two-dimensional histogram in HSI color space is applied for the HS channel; thus, the histogram provided chromaticity information and is called a chromaticity histogram (Li et al., 2009). Another important factor that impedes recognition (as a low-level feature of the image) is the shape of the polyp. For shape illustration, two techniques can be applied, i.e., contour-based and region-based (Zhang and Lu, 2003). However, due to the background in the capsule images and the fact that is hard to achieve clear and accurate contour, a shape descriptor, which is region-based, was implemented in the proposed technique. In the case where the shape features are invariant to rotation, translation, and scaling, Zernike moments (Teague, 1980) can be applied as they satisfy the mentioned properties. Three-hundred images from the gastrointestinal tract are selected, consisting of 150 images of polyps and 150 normal images. Two classifiers, i.e., the machine learning perception (MLP), having three MLP layers plus two nonlinear outputs, and the support vector machine (SVM) inspired by the work are used. By using Zernike moments with 5 orders, the proposed technique employing the machine learning perception (MLP) classifier resulted, amazingly, in even better results than the support vector machine (SVM), yielding an accuracy, sensitivity, and specificity of 94.20%, 93.33%, and 95.07%, respectively (Li et al., 2009).

Moreover, the detection of colorectal polyps inside WCE images by unsupervised learning is attempted. The proposed technique consists of watershed segmentation employing a unique primary marker selection procedure that is dependent on features such as the Gabor texture and clustering using K-means. To avoid illumination issues, a selection technique termed initial marker dependent on Gabor filters and K-means is implemented (Hwang and Celebi, 2010). K-means clustering used to generate clusters having similar properties and the output of the Gabor filters was used to obtain a number of segmented regions based on the elevation levels. Based on these levels, regions chosen as local minima were selected as initial markers. The watershed transform is applied using the selected markers and regions that are more deeply segmented were obtained with the controlled marker. The selected polyp candidates exhibited an elliptical or spherical shape; thus, the curvature center ratio introduced for each region. A total of 128 images used for the proposed technique, consisting of 64 normal and 64 polyp images. An optimal value of "K" chosen, and the correct polyp segmentation ratio (CPSR) is computed with 100% sensitivity and 81.00% specificity (Hwang and Celebi, 2010).

An automatic polyp detection based on the decomposition of WCE images via wavelet and curvelet transform is proposed starting with a pre-processing stage to eliminate encircled black regions followed by the data augmentation (Souaidi et al., 2018). For the multi-resolution analysis, transform like the 2-D discrete wavelet transform, dual-tree complex wavelet transforms (DTCWT), Gabor wavelet transforms, and curvelet transforms is exploited. The distribution of the sub band coefficients is shaped concerning the characteristics of the wavelet-based transform weather it produced a real or complex values coefficient by utilizing Weibull, generalized Gaussian distribution (GGD), Rayleigh or gamma distributions. Processing of each channel of RGB color space is done to collect

all the features to generate a final feature vector. Support vector machine classifier (SVM) is used for classifying the images into polyps and normal images. The data set comprised of 5926 polyp patches and 1864 normal patches after data augmentation. The proposed scheme while using gamma distribution and DTCWT generated a high performance in terms of accuracy, sensitivity, and specificity of 96.00%, 96.00%, and 96.00% respectively (Souaidi et al., 2018).

As WCE generate a lot of images that result in an imbalance data set. This imbalance of the data set creates a bias toward the non-polyp class in the detection process. A structure for achieving enhanced data sampling is proposed, i.e., Adaboosting, in which unbiased polyp detection is learned. The learning arrangement consists of numerous weak classifiers integrated, having re balanced data sets based on down/up sampling to obtain a polyp detector. Further partial least squares (PLS) is applied in order to improve the capability for feature discrimination among non-polyps and polyps, and high dimensionality (Bae and Yoon, 2015). Old techniques based on imbalanced learning are classified into re weighting and re sampling approaches (Galar et al., 2012; Hoens and Chawla, 2013). These techniques suffer from a cost sensitivity problem, and to overcome this cost sensitivity boosting (CSB) algorithms such as AdaBoost (Fan et al., 1999) and CSB1/CSB2 (Ting, 2000) is implemented. Moreover, for slanted data sets, partial least squares (PLS) is used to achieve asymmetric classification (Qu et al., 2010). Modeling of relationships among groups with latent variables is performed via PLS. The goal of this analysis is to establish latent vectors or orthogonal score vectors by maximization of the covariance among dissimilar variable sets (Hwang et al., 2007). The proposed technique uses the latest improvement in the area of imbalanced learning to develop an integrated framework of AdaBoost, re sampling, and partial least squares (PLS), termed as data sampling-based boosting.

Despite the enhanced classification achieved with PLS, even with imbalanced datasets, there is still a need to cope with the issue of imbalance. The synthetic minority over-sampling technique (SMOTE) for the generation of samples in feature space is adopted. When the data are free of noise, down-sampling is performed with SVM, whereas in the case of noisy data, Tomek links analysis (Kubat and Matwin, 1997) is performed. To address the diverse appearance of polyps and imbalanced data sets, a framework design called data sampling-based boosting via integration of suggested feature learning plus data sampling was utilized with algorithms such as AdaBoost.M1 (Polikar, 2006). The entire technique was implemented with 1263 images after checking the data set with CVC-ColonDB (Bernal et al., 2012). Benchmarking of the detection performance using different features (color histogram, LBP, HOG and VHOG) for the same classifier showed that better performance was achieved when VHOG was used in combination with the proposed PLS. Extensive quantitative benchmarking of diverse detectors based on the AdaBoost algorithm using different re sampling approaches employing imbalanced learning techniques was also performed. These detectors include:

- SVM: Single linear support vector machine (SVM) classifier
- ESM: Ensemble classifier (No sampling)
- ESM-UP: Ensemble classifier (Up-sampling)
- ESM-DW: Ensemble classifier (Down-sampling)
- ESM-UP-DW: Ensemble classifier (Up/down-sampling)

The results showed better detection performance in terms of a lower miss rate and false positive per image (FFPI), which were achieved by implementing detectors with the VHOG-PLS feature, using a resampling-based technique, i.e., ESM-UP-DW, irrespective of the size of the dataset and data variations. The cascaded approach (Viola and Jones, 2001) was applied for imbalanced datasets via ensemble learning, despite the use of AdaBoost. The proposed technique outperformed the cascaded approach in terms of the recall and precision-recall area under the curve (PR-AUC), yielding higher accuracy and lower miss-detection (Bae and Yoon, 2015).

For detection of polyp inside WCE images, a technique is proposed based on the extraction of local fractal dimension (LFD) features above detected SIFT key points. Concatenation of extracted features with complete local binary pattern (CLBP) or uniform local binary pattern (LBPu) is done in order to assimilate texture information. Due to the appearance, color, and texture of polyps, extraction of geometrical information was performed by using LFD. Textural features that are robust to RST are extracted by choosing the CLBP or LBPu from the WCE by concatenation with LFD. The dataset consisted of 2433 capsule images comprising 1700 and 733 normal and polyp images, respectively. The classification is performed with classifiers, such as MLP, SVM, and random forest, in order to differentiate polyp images from normal images (Ansari and Charfi, 2017). For the integration of SIFT + LFD + LBPu and SIFT + LFD + CLBP, the performance achieved with the three classification techniques is assessed, demonstrating the better performance of SVM. The better classification was achieved with SIFT + LFD + LBPu in comparison with SIFT + LFD + CLBP, resulting in precision, accuracy, specificity, and sensitivity of 96.19%, 97.98%, 95.86%, and 100%, respectively. The area under the curve (AUC) for both combinations (i.e., SIFT + LFD + LBPu for all three classifiers) are computed, and values of 0.9793, 0.9779, and 0.9765 were respectively obtained for SVM, random forest, and MLP, whereas for SIFT + LFD + CLBP with all three classifiers, the AUC was 0.9632, 0.9703, and 0.9632 for SVM, random forest, and MLP, respectively (Ansari and Charfi, 2017).

A modified region-based convolutional neural network (R-CNN) by utilizing Keras deep learning model is proposed for detecting polyps inside WCE images by producing masks around polyps. The localization of polyps is done by segmentation followed by detection of most likely pixels points of polyps in the mask areas and determining the centroid depicting the location of the polyp (Sornapudi et al., 2019). Data augmentation of WCE images is done by rescaling, rotation, histogram equalization, and Gaussian blurring to avoid overfitting during the model. For the features extraction, ResNet (He et al., 2016) with Feature Pyramid Network (FPN) (Lin et al., 2017) is implemented that extracts feature from different hierarchical levels with different scales. The Region Proposal Network (RPN) along with feature extraction make the bounding box or proposals where the RPN works in a convolution form by scanning the features received from FPN. The binary masks are generated by assigning the region proposals to various distinct regions of features maps obtained from FPN. These mapped regions are supplied to the RoIAlign module (He et al., 2017), supported by convolutional layers and fully connected layers for predicting the location and size of the predicted mask to fit the object. The final binary mask (28×28) matches to specific region proposal, and the regions that have a class probability of 0.8 or greater are rated true predictions. For the optimization of the loss function on specific ROI region, stochastic gradient descent (SGD) is used. Fine-tuning of the model is done on the pre-trained models of COCO weights and Flicker's balloon by using the data set of ImageNet dataset and Microsoft COCO, where the learning rate was kept 0.001, epochs of 1000, and with empirically chosen hyperparameters. The proposed model is implemented for different scales of region proposals such i.e. 8×8, 16×16, 32×32, 64×64, and 128×128 to detect poly inside WCE images of different sizes resulting in precision and F1 score of 98.46% and 96.10% respectively (Sornapudi et al., 2019).

Table 4 sums the survey for the detection of polyps in WCE images in terms of their respective performances for each method. The performance metrics such as accuracy, specificity, sensitiv-

Table 4
Efficacy of existing methods for polyp detection in WCE images.

Ref.	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Karargyris and Bourbakis (2011)	No	67.50	100	No
Billah et al. (2017)	98.65	98.52	98.79	No
Billah and Waheed (2018)	98.43	98.23	98.67	No
Yuan and Meng (2014)	91.43	94.78	88.09	No
Li and Meng (2012a)	91.60	No	No	No
Yuan and Meng (2017)	98.00	No	No	No
Yuan et al. (2016)	93.20	90.88	95.54	No
Li et al. (2009)	94.20	95.07	93.33	No
Hwang and Celebi (2010)	No	81.00	100	No
Souaidi et al. (2018)	96.00	96.00	96.00	No
Bae and Yoon (2015)	No	No	No	70.67
Ansari and Charfi (2017)	97.98	95.86	100	96.19
Sornapudi et al. (2019)	No	No	No	98.46

ity, and precision for the detection of tumors by the respective approaches are listed in the table.

5. Ulcer detection in WCE images

An automatic technique using a convolutional neural network (CNN) based on single short multi-box detector (SSD) is presented to detect erosions and ulceration in WCE images (Aoki et al., 2019). For training the model 5360 WCE images having erosions and ulceration collected from 115 patients, while 10,440 free images for testing collected from 64 different patients. In the 10,440 free images, 440 contain erosions while remaining 10,000 images are normal small bowl images. Single short multi box detector (SSD) (Liu et al., 2016b), is implemented for AI diagnostic system having input in the form of input labeled images via framework of deep learning named Caffe. The stochastic gradient with 0.0001 learning rate is implemented after resizing each image to 300×300. The image annotation is done manually and symbolized by green as a (true box) and yellow as a (CNN box) followed by a condition that if the overlapping area between true and CNN box enclosed is greater than either true box or CNN box; Define it as the accurate answer. The second condition is for the existence of multiple CNN boxes in a single image, then detection of an ulceration or erosion is true even 1 boxes accurately detect rather than that all CNN boxes should fulfill criteria. The area under the curve (AUC) for CNN model to detect ulcerations or erosion after confidential interval (CI) adjustment is 95.80%. The optimal value of cut-off for score of probability according Youden index was 0.481, so areas with score of probability greater than 0.481 are identified as ulcerations or erosions by CNN mode. At this cut-off value, the CNN model accuracy for detection of erosions or ulcerations in terms of accuracy, sensitivity and specificity are 90.81%, 88.20% and 90.90% respectively (Aoki et al., 2019).

To address the issue in training a deep learning model with respect the quality of data, a data augmentation for ulcer via feature space extrapolation technique is proposed. A dual encoder network integrating two VGG19 nets fully connected is made for one-dimensional encoded feature space (Lee et al., 2019). Initial training classification is done employing a normal dataset and imbalanced ulcer dataset. Then based on the respective closest normal sample, extrapolation of training data of ulcer in the feature space encoded is done. For extrapolation, the assumption is made for the availability of a good enough number of normal images which can render the whole visual features of a normal class. Based on extrapolation ratio R, every pair of ulcer-normal creates a single extrapolated sample. Fine-tuning of fully connected layers for classification is done after creating a balanced ulcer and normal dataset in the feature space. The data set consists of 14,500 WCE images out of which 1000 are an ulcer and 4500 are normal images followed by basic data augmentation technique. The network is implemented using TensorFlow having learning rate of 5e−07, batch size of 10, and optimizer like Adam on Nvidia TITAN xp. For the comparison four different frames with VGGNet are implemented where DN-Extra2000 generated higher performance in terms of accuracy, precision, and recall of 96.20%, 95.10%, and 92.60% respectively (Lee et al., 2019).

An automatic ulcer detection system based on color and texture feature extraction is proposed. The proposed system attempt pre-processing by applying median filtering to remove noises from the neighboring pixels (Charfi et al., 2019). For segmentation purpose, simple linear iterative clustering (SLIC) technique is applied. After the SLIC, color and textural saliency mapping are computed. The color saliency map features are computed based on the mean of the colors of sub-regions of all pixels by utilizing the Euclidean distance. For the textural saliency map features a local binary pattern (LBP) is implemented followed by the fusion process of both maps. Features are extracted by applying color LBP (CLBP) by estimating the LBP for each channel of RGB color channel. For the spatial layout and local shape of the image, pyramid of histograms of orientation gradients (PHOG) is employed for extracting descriptors at edges in comparison to HOG. Also, the bag of the visual word is analyzed to extract features by setting the no. of visual words to 500. For classification, every feature extracted from the segmented regions is separately classified by SVM via a radial basis function kernel, MLP and RF classifiers. In contrast, to consider the most suitable class that observation should fit the label, the probability scores obtained from each classifier are selected and are given as observations to a first-order discrete hidden Markov model (HMM). The proposed detection approach is implemented on two data sets; one having a total of 446 WCE images in which 287 are an ulcer and 159 are normal images. The second data set comprised of a total of 2170 images, out of which 570 are normal and 1600 are ulcer images. The proposed automatic detection system generated a high performance in terms of accuracy, sensitivity, and specificity for both data sets. For the first data set, the accuracy, sensitivity, and specificity achieved is 95.30%, 96.20%, and 96.20% respectively, while for the second data set, the accuracy, sensitivity, and specificity archived is 94.80%, 96.20%, and 95.50% respectively (Charfi et al., 2019).

A multi-scale method that is dependent on the local binary patterns (LBP) and Laplacian pyramid (MS-CLBP) for the detection of ulcer abnormalities inside WCE images is proposed in YC_bC_r and RGB color space. Additional information is captured as a robust descriptor i.e. illumination variance by the proposed method in WCE images (Souaidi et al., 2019). For multi-resolution analysis, the LBP is implemented with neighborhood circular symmetry as (P, R), where P and R represent the pixels and radius respectively. The CLBP is an extended form of LBP specified by the local difference of sign-magnitude transform (LDSMT) and the local central pixel. Two CLBP variants i.e. $(CLBP/M^r_{(P,R)}i)$ and $(CLBP/S^r_{(P,R)}i)$ having R=1 and LBP neighbors P = 8 are applied on each channel of

color space i.e. YCbCr and RGB. The $CLBP/M^r_{(P,R)}i$ and $CLBP/S^r_{(P,R)}i$ are the two constituents of LDSMT that represent sign component S_p and M_p components respectively. The generated CLBP's coded map i.e. $CLBP/M_8, 1^r i$ and $CLBP/S_8, 1^r i$ are reflected as input for five-level of Laplacian pyramid decomposition for coarse-to-fine detection approach of computing features at any spatial resolution followed by calculation of histogram from each sub-bands. Taking four sub-bands coefficients histograms, extraction of six statistical features are done that are concatenated further to generate a single feature vector. For classification, support vector machine (SVM) is applied on two data sets (Souaidi et al., 2019). The first data set obtained from Cheik Zaid Hospital, Rabat, Morocco comprised of 240 images out of which having 160 normal and 80 ulcer images while the second dataset comprised of 212 images out of which 133 images having ulcer and 79 normal images. The performance of classification for both multi-scale CLBP/Sign and the multi-scale CLBP/Magnitude are evaluated using SVM both RGB and YCbCr color space. For Cr component of YC_bC_r color space on operator $CLBP/M_8, 1^r i$, the proposed approach applied on first data set achieved an accuracy, sensitivity, and specificity of 93.88%, 97.21%, and 87.22% respectively. For the second data set, the proposed approach using operator $CLBP/S_8, 1^r i$ in RGB color space achieved an accuracy, sensitivity, and specificity of 95.11%, 95.64%, and 94.83% respectively (Souaidi et al., 2019).

Automatic detection and classification of ulcers proposed for WCE images using deep learning modes of CNN architectures i.e. Alexnet and GoogLeNet. Features such as LBP, color histogram and color coherence are extracted for detailed simulation for checking the performance of CNN model (Alaskar et al., 2019). The CNN layers consisting of neurons arranged in 3 dimensions for converting each 3D input into respective 3D output of activation neuron. The pretrained models used are fine-tuned through the weight freezing of the initial layer. The pre-trained model GoogLeNet (Krizhevsky et al., 2012) in which every layer work as a kernel for extracting finest features is used for object detection and classification. The proposed work use GoogLeNet with addition of four new layers, 50% dropout probability, fully connected layer (FCN), softmax layer and output layer of classification. The AlexNet (Yuan and Meng, 2017) used ReLU as an activation function with 50% dropout probability where feature extractor is used as a first layer plus AlexNet has fewer layers compared to GoogLeNet. For the modification, the addition of one FCN layer and equaling the 23rd layer to output layer and no. of classes. The data set consisted of 1875 images out of which 250 images are normal and 1525 images are ulcer images that are resized accordingly to train both models. Comprehensive experimental results are obtained for both models by selecting different learning rates that show that with a learning rate of 0.0001, a 100% classification accuracy with sensitivity and specificity of 1 for both GoogLeNet and AlexNet is achieved. However, AlexNet is outperforming GoogLeNet in the training time by almost 50% due to lesser no. of layers (Alaskar et al., 2019).

A unique technique for the detection of ulcer and erosion of small intestine is proposed using framework of deep learning for WCE images. 144 patients are selected for the dataset collection, which include 65 normal, 47 ulcer and 32 erosion cases (Fan et al., 2018). For detection of lesions two independent models i.e. one ulcer and one for erosion detection are trained. For the ulcer detection 3250 ulcer and 5000 normal images are used, while for erosion detection, 8000 normal and 4910 erosion images are used. AlexNet (Yuan and Meng, 2017) is implemented for the detection purpose with both models trained with stochastic gradient descent, 50 batch size, 0.01 of learning rate, 0.9 of momentum and 0.001 of weight decay. The training accuracy after changing six time learning rate for erosion and ulcer model achieved high accuracy level of 94.68% and 96.36% respectively. The ROC curve is also evaluated to check the veracity of both models generating AUC value for erosion and ulcer model of 0.9863 and 0.9891 respectively. For further performance demonstration of the technique proposed, comparison with other approaches such as SVM and gray scale histogram is done yielding outperforming results. The overall performance for both erosion and ulcer detection model generated high performance values in terms of sensitivity of 93.67% and 96.80%, an accuracy of 95.34% and 95.16%, a specificity of 95.98% and 94.79%, an AUC of 0.9904 and 0.9805 correspondingly (Fan et al., 2018).

For detecting multi-abnormalities inside WCE image, a novel technique consisting of a stage called learning of feature to generate visual words. From polyps, bleeding, ulcers, and samples of normal WCE images, computation of SIFT feature independently. Segmentation is done to remove the background, highlighting the critical areas of the WCE images for all patches followed by textural and color information extraction via HSC-SIFT (Yuan et al., 2017). Then the K-means clustering algorithm is applied on extracted features to get visual words. Integrating these four kinds of visual words used later for classification. At the feature stage coding, a unique saliency and adaptive locality-constrained linear coding (SAALC) is proposed for encoding the images. Encoding of patch features was achieved with SAALC based on adaptive coding computed using the difference in the distance between features and visual words. Compared to traditional coding such as LLC, implementation of adaptive coding bases removes the issue of selecting every patch feature based on static visual words. The max-pooling yield representation of the WCE image was applied after coding of every patch descriptor. For recognition of abnormal images, SVM is used as a classifier (Chang and Lin, 2011) using a Gaussian radial basis as a kernel. The data set consists of total 1650 images out of which 500 images each for normal tissues, polyps, and bleeding while rest 150 images are ulcers. Compared to other techniques for multi-abnormality classification, the proposed technique generated the best results, having an overall recognition accuracy (ORA) of 88.61%. The proposed technique also outperformed the bench marked technique for detecting bleeding, polyps, ulcers, and normal images, with an accuracy of detection for each image type of 96.60%, 83.50%, 80.33%, and 88.20%, respectively (Yuan et al., 2017).

In order to detect ulcer in WCE images, a novel textural extraction process is suggested using the multi-resolution property of discrete curvelet transform (DCT) (Candès and Donoho, 2000; Candes and Donoho, 2000; Demanet and Ying, 2007). From the subbands of DCT, the lacunarity index is computed to obtain textural information (Eid et al., 2013). Employing curvelet transforms for WCE images, a better accuracy has been achieved using HSV and YCbCr color spaces in comparison to the hardware-oriented RGB color space. From examinations reported on specific data, YCbCr has proven to generate the best performance among all the three-color spaces. Both YCbCr and HSV color space provides a close perception to human, while YCbCr being more preferred color space in uncontrolled lighting conditions such as during WCE process (Li and Meng, 2009b). DCT is applied on each color channel via wrapping technique in order to remove wavelet weakness, as wavelet cannot adequately signify image singularities like curves and lines while dealing with medical data (Candès and Donoho, 2000; Candes and Donoho, 2000; Demanet and Ying, 2007). For the DCT analysis parameters such as No. of scales and angles are selected carefully to avoid losing sub-band information. Differential lacunarity analysis is used to compute the filling of space while dealing with a large homogeneous data set. In spite of using the traditional gliding box algorithm (GBA) for computing the lacunarity, that is only suited for binary images, the technique used to implement the differential lacunarity analysis (LAC) suitable for gray scale images (Allain and Cloitre, 1991; Dong, 2000). By choosing a fixed box size of "$r = 3$" and a window size of "$w = 4 - box_m$", where "box_m" is equivalent to the minimum aspect of every sub-image the LAC was calculated.

As this computation can result in a high dimensional vector feature, leading to degradation of the accuracy and performance speed, the hyperbolic function is implemented to model LAC-w curve, defined as (Allain and Cloitre, 1991; Dong, 2000):

$$\Lambda = \frac{b}{w^a} + c \tag{2}$$

where the independent variables *a,b,c* are computed as least squares problem solution; that represents the LAC-based textural features of a WCE image (Charisis et al., 2012). The *w* represents window size.

For classification, SVM having the kernel of radial basis function with the scalar unit factor is implemented. Ten-fold cross-validation is applied for testing and training of the classifier. The experimental process is performed with 130 normal and 130 abnormal images. The results of using different angles and color channels are evaluated, demonstrating that the "Cr" channel gave good results, whereby choosing an optimal angle in the "Cr" channel by trying different combinations led to an accuracy of 86.54%, the specificity of 88.56%, and sensitivity of 84.51% (Eid et al., 2013).

In order to classify normal and abnormal images inside WCE, a color wavelet transform (CWC) based on DWT is proposed. Textural color and analysis are done due to variance in the appearance of ulcer and color of blood. For multi resolution and textural analysis, three-level DWT is used. Four second-order statistical models (correlation, entropy, angular the second moment, and inverse difference) are used at different orientations ($0°$, $45°$, $90°$, and $135°$) to provide discrimination with a high level of accuracy (Liu et al., 2012). Seventy-two CWC vectors were computed after calculating the resemblance among the wavelet features in order to avoid waste of computational time due to repeated information. Human textural perception defined by Texton (Julesz, 1981) is specifically used for categorization and is symbolized by the texton map "T". For a pair of textural layout filters defined by (*r*, *T*) where "*r*" is the region of the image and "*t*" is the texton, at location "*i*", the response of the feature is proportional to the pixels for the region of offset "*r* + 1″" with the texton index "*t*", represented as.

A modified joint boost algorithm (Torralba et al., 2004) is implemented for iterative selection of the highest discriminative textural layout as a "weak learner' for every texton. This selection is then integrated into the classifier, allowing every single weak learner to be used for classification of a number of classes at one time (Torralba et al., 2004). The data set collected from Jin Shan Corporation consisted of 100 images, out of which 45 images are abnormal tissues. The training and testing data consisted of 45 random samples each, while the remaining 10 samples are used for validation. The experimental analysis is divided into two phases comprised of selecting the best color space in order to define the best textural properties, followed by bench marking with relative techniques such as co-occurrence matrix (CM) and wavelet transform (WT). The results achieved with the proposed technique indicated better performance in terms of the sensitivity 82.30% and specificity 89.10% (Liu et al., 2012).

Furthermore, textural feature analysis is done for the detection of ulcers inside WCE images using curvelet-based pattern i.e. LBP (Li and Meng, 2009b). LBP shows resistance to illumination variances and it can discriminate small structures such as dark and bright spots. The features extracted by LBP are the mean, standard deviation, energy, kurtosis, skew, and entropy. The multi-resolution analysis is performed using wavelet theory as it yields smooth functions when dealing with one-dimensional data. To capture more features in more directions, curvelet transforms is applied, which intimates a curve in the shape followed by a precise scaling rule (Do and Vetterli, 2003). The data set collected from 20 patients consisted of 3600 images, out of which 1800 normal and 1800 ulcer images. To prevent over-fitting, 4-fold cross-validation was used in this specific technique. SVM and MLP neural network are used to verify the performance of the extracted features for distinguishing normal from ulcerated regions. For performance analysis, the proposed technique with suggested CLBP features is bench marked in "RGB" and "YCbCr" color space using different types of features, i.e., LBP features, CWC, and CM. Implementation of MLP neural network in "YCbCr" color space for the specific data set yield better performance, with an accuracy, sensitivity, and specificity of 92.37%, 93.28%, and 91.46%, respectively (Li and Meng, 2009b).

Moreover, detection of ulcer inside the WCE image is done by using local feature integration. The proposed method is based on the bag-of-words (BOW) model and feature fusion method. Implementation of LBP and SIFT as feature extractor done on each patch after patching scheme. Clustering for every patch the feature which resulted in a vocabulary dictionary is performed with the K-means clustering algorithm (Yu et al., 2012). In the proposed technique, the spatial pyramid kernel is used to obtain spatial pyramid BOW histograms to consider spatial layout information (Lazebnik et al., 2006). The main addition in this feature is an increment of the coarser grid over the space of the image at 1...L resolution so that the grid at level 1 has $2l - 1$ cells beside every dimension. A total of 344 WCE images comprising 172 normal and 172 ulcer images were utilized. Additionally, 120 ulcer images and 120 normal images were chosen for testing. For classification, SVM is implemented for training these features. For feature fusion, linear classifier dependency modeling (LCDM) based on Bayesian theory is used to avoid large intra-class and small inter-class variations (Ma and Yuen, 2011). For extraction of local regions for the BOW model, techniques such as grid and Harris affine are also used, followed by the construction of vocabularies of different sizes for pyramid kernel. The results showed that by using the best classifier with feature fusion in combination with LBP produced a higher accuracy, sensitivity, and specificity of 89.58%, 99.17%, and 80.00%, respectively than those of the feature fusion in combination with SIFT (Yu et al., 2012).

To detect ulcers in WCE images, a two-stage fully automatic technique is proposed. In the first stage, a saliency detection technique based on multi-level super pixel illustration was proposed for shaping the ulcer. Salient regions that are semantically significant and perceptible were found via image segmentation using multi-level super-pixel. The analogous saliency was estimated according to the textural and color features for every level, as these features are related to different primary areas of super-pixel sizes. A final saliency is generated by fusing all levels of saliency maps. The second stage involved the characterization of the images from the saliency map in order to efficiently encode the features for ulcer detection via locality constrained linear coding (LLC) (Yuan et al., 2015b). The suggested saliency detection technique comprised three major steps, starting with multi-level segmentation using simple linear iterative clustering (SLIC), which segment the WCE images from rough to a fine level into numerous super-pixels. Thereafter, saliency regions are detected based on textural and color contrast for every level of super-pixel. The Leung-Malik (LM) filter bank is used for extraction of the textural features, as LM is a multi-resolution, multi-orientation, and multi-scale filter comprising 48 filters (Leung and Malik, 2001). For a given image at the *l*th level, the texture saliency texture$\hat{\mathbf{S}}_l \in \mathbb{R}^{M \times N}$ is given as:

$$\text{texture}\hat{\mathbf{S}}_l(x, y) = \sum_{j=1, j \neq i}^{N} \frac{\mathbf{D}(i, j)}{\max_{i,j}(\mathbf{D}(i, j))} \tag{3}$$

where $\mathbf{D}(i, j)$ is the (i, j) entry of the distance matrix $\mathbf{D}$ representing the Euclidean distance among the texture feature vectors (Yuan et al., 2015b). The texture$\hat{\mathbf{S}}_l \in \mathbb{R}^{M \times N}$ entry (x, y) defines a pixel inside the *i*th superpixel region. So by the combining different level

saliency map for lth superpixel level such as $l = 1, \ldots, L$, the final texture saliency texture$\mathbf{S}$ is obtained as:

$$\text{texture}\mathbf{S} = \frac{1}{L}\sum_{l=1}^{L}\text{texture}\hat{\mathbf{S}}_l \qquad (4)$$

where L is the different level saliency map of lth superpixel level and texture$\hat{\mathbf{S}}_l$ is the texture saliency.

For color-based saliency detection, numerous trial, and error experiments are performed in different color spaces to preserve useful information and HSV and CMYK are found to be optimal for ulcer recognition. The mean and variance values from HSV (i.e., S plane) and for CMYK (i.e., M) are selected for generating the color feature matrix. Regions with variations in the color distribution are chosen to obtain color saliency. Thereafter, by taking the mean of these different levels, the final saliency color$\mathbf{S}$ is calculated as:

$$\text{color}\mathbf{S} = \frac{1}{L}\sum_{l=1}^{L}\text{color}\hat{\mathbf{S}}_l \qquad (5)$$

Finally, a multi-level texture and color super-pixel saliency is proposed by fusion of the saliency regions. The final saliency final$\mathbf{S}$ map was obtained as:

$$\text{final}\mathbf{S} = \text{color}\mathbf{S} \circ \text{texture}\mathbf{S} \circ \mathbf{K} \qquad (6)$$

where "o" is the Hadamard product matrix while $\mathbf{K} \in \mathbb{R}^{M \times N}$ representing Gaussian kernel focused at the center of the image which slowly declines toward the edges in order to imitate human perception (Yuan et al., 2015b). The color$\mathbf{S}$ represents the final color saliency while texture$\mathbf{S}$ represents the final texture saliency. In the second stage, the modified locality constrained linear coding (LLC) technique is used to combine the acquired saliency map with the max-pooling technique for WCE images. It consists of three steps, starting with the extraction of local descriptors (dense scale-invariant feature transform dSIFT (Lowe, 2004), a dense histogram of oriented gradients (dHOG) (Dalal and Triggs, 2005), and dense uniform local binary patterns (duniLBP) (Ojala et al., 2002) for each image independently to generate a code book or visual vocabulary via the K-means clustering technique. After obtaining three code-books, the mapping of each descriptor matrix is performed with code books for image illustration. Initially, the original LLC was used to encode every single descriptor obtained for each testing and training image. After coding, all codes were pooled together to obtain a representation of the WCE image. The images are finally represented final$_\mathbf{y} \in \mathbb{P}^{3M}$ by the concatenation of the results generated from three descriptors as:

$$\text{final}_Y = [dSIFT_yT \quad dHOG_yT \quad duniLBPyT]^T \qquad (7)$$

The dataset for the experimental analysis consisted of 170 normal and 170 ulcer images taken from 20 patients. 1225 patches are extracted, having dimensions of 16×16 with three kinds of features, i.e., dSIFT), dHOG and duniLBP. SVM having a Gaussian radial basis function is then applied as a classifier. For testing and training, 20% and 80% of the images are respectively used, with 5-fold cross-validation. Tuning of the vocabulary size (M) for the proposed technique is carefully executed, and the results generated with $M = 50$ are the best, with mean specificity, sensitivity, and accuracy values of 91.18%, 94.12%, and 92.65%, respectively indicating the validity of the proposed algorithm (Yuan et al., 2015b).

Table 5 sums the survey for the detection of ulcers in WCE images in terms of their respective performances for each method. The performance metrics such as accuracy, specificity, sensitivity, and precision for the detection of tumors by the respective approaches are listed in the table.

6. Bleeding detection in WCE images

Bleeding inside the gastrointestinal (GI) tract is an important factor that needs to be detected timely. However, the images generated from WCE are numerous and having different luminance qualities at different places, making a hectic task for physicians to diagnose. This bleeding inside the GI tract can lead to further harmful diseases, such as a tumor, poly, and ulcer, etc. making the necessity of techniques to detect efficiently bleeding inside WCE. This section details the approaches attempted to detect bleeding inside WCE images.

To detect bleeding inside a frame of WCE videos, a color, and textural features extraction technique is proposed. Color analysis is done for removing dark and bright blocks to compute luminance by taking the square root of a sum of single RGB blocks (Pogorelov et al., 2019). Edges detection and removal for avoiding false results using the canny operator followed by morphological dilation which dilates the information detected. Color features are extracted by computing the red ratio for single pixels in the R channel of the RGB color space. The extracted features based on color is fed to SVM that uses three types of kernel i.e. radial basis function (RBF), polynomial, and linear for classification purpose. By applying a statistical approach of computing textural features, the gray-level co-occurrence matrix (GLCM) is utilized that extracted 22 textural features. A classification approach based on color and texture integration is implemented for the discrimination of bleeding and non-bleeding pixels. Random tree (RT), random forest (RF), and logistic model tree (LMT) classifiers are used having different combinations of color and texture features as an input vector. The data set was collected by endoscopists from Endoscopy unit at University of Malay Medical Center (UMMC), Kuala-Lumpur, Malaysia consisting of 1200 frames out of which 800 are bleeding while 400 are non-bleeding frames. For the frame level of bleeding detection, SVM using RBF as a kernel resulted in high accuracy, specificity, sensitivity, and F1 score, of 97.77%, 95.55%, 97.76%, and 97.80% respectively. For the entire pixel-level bleeding detection the proposed approach of integrated features classification generated high results in terms of accuracy, specificity, sensitivity, and F1 score, of 97.76%, 95.59%, 97.76%, and 97.76% respectively (Pogorelov et al., 2019).

Furthermore, a computer-aided diagnosis (CAD) system i.e. saliency-aware hybrid network (SHNet) for the detection of GI bleeding inside WCE is proposed. Precisely, the SHNet includes two densely connected convolutional networks (DenseNets) and a guided deconvolutional network (GDN) which are regarded as global image stream and saliency-aware stream sequentially respectively (Xing et al., 2019). The GDN is used to encode the most discriminative features maps of the last layers from DenseNets and generate saliency maps. Polar transformation is proposed to lessen the noise from the background and highlighting information of the image. Finally, the ensemble learning approach is employed that optimize the ultimate diagnosis result jointly. The ensemble learning system in the training phase sums the losses of two dense models and optimize eventually updating all the parameters due to back-propagation. The data set consisted of 1200 images out of which 600 normal images and 600 bleeding images. For the training the network, the algorithm such as Adam Optimizer with a learning rate of 0.01 and batch size of 8 is used that resulted in high performance in terms of precision, recall, and F1 score of 94.70%, 97.30%, and 95.90% respectively (Xing et al., 2019).

A unique approach based on textural feature extraction is proposed in which classification is achieved by using probabilistic neural network (PNN). For detection of bleeding areas, an intelligent programming technique that are tested on image level and pixel level (Pan et al., 2011). Inside WCE images, non-bleeding areas are dissimilar bleeding regions even in colored or gray images. Color

Table 5
Efficacy of existing methods for ulcer detection in WCE images.

Ref.	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Aoki et al. (2019)	90.81	90.90	88.20	No
Lee et al. (2019)	96.20	No	No	95.10
Charfi et al. (2019)	94.80	95.50	96.20	No
Souaidi et al. (2019)	95.11	94.83	95.64	No
Alaskar et al. (2019)	100	100	100	No
Fan et al. (2018)	95.16	94.79	96.80	No
Yuan et al. (2017)	88.61	No	No	No
Candès and Donoho (2000)	86.54	88.56	84.51	No
Eid et al. (2013)	No	89.10	83.20	No
Dong (2000)	92.37	91.46	93.28	No
Liu et al. (2012)	89.58	80.00	99.17	No
Do and Vetterli (2003)	92.65	91.18	94.12	No

textural feature extraction is done for RGB and HSI color spaces making 6D color feature vector x = (R, G, B, H, S, I). The PNN used as a classifier is feed forward three layered neural network that include, input layer, radial basis layer, and competitive layer. The color feature vector is the input for PNN classifier, which recognize $m = 1$ and $m = 0$ as bleeding and non-bleeding pattern respectively. After that, a bleeding detection technique is applied after removing dark pixels from WCE image followed by forming a feature vector of each pixel as: x = (R, G, B, H, S, I). A pixel by pixel scanning is done on WCE image and recognition is done by PNN for bleeding and non-bleeding pattern. The software for bleeding detection is test both for pixel and image level. For pixels level 930,575 pixels containing 339,850 bleeding and 590,725 non-bleeding pixels are manually detected. For image level, 3172 bleeding images and 11,458 non-bleeding images are manually detected. The results generated by image-level testing generated a higher performance in terms of sensitivity and specificity of 93.10% and 85.60% respectively compared to pixel level testing as it's not necessary for whole image level to be so sensitive for pixel-level of minor bleeding (Pan et al., 2011).

Furthermore, a novel technique is proposed for detection of bleeding inside WCE video that is based on pixel grouping via pixel segmentation (Fu et al., 2013). Except edge pixels, non-bleeding and bleeding pixels mostly have similar luminance and vary in redness resulting to apply canny detector in L channel. The canny edge detector detects the pixels of edge while pixels of bleeding can be survived followed by morphological operation to dilate the image. Due to high luminance of WCE images, pixel to pixel segmentation is a time consuming and computational cost task, therefore adaptive grouping of pixels based on location and color via super pixel segmentation is performed. The proposed super pixel segmentation is the modified approach from Levinshtein et al. (2009), Achanta et al. (2012) introducing a cluster center as an initial seed growth that depend on similarity of pixel for finding a cluster rather than evolution of curve, making it faster than (Fu et al., 2013). Furthermore, Gaussian filter is applied to remove noise from WCE images and to find connectivity for bleeding region, new label assigning is done on disjoint segment resulting a single super pixel for even a small bleeding region. A similarity measure among color similarity and spatial distance is computed based on Euclidean distance and spatial position of pixel respectively as shown by,

$$S = S_{color} + \lambda \times S_{spatial} \tag{8}$$

The λ controls the two factors in above equation.

Implementing SVM (Chang and Lin, 2011) for training, detecting patterns of bleeding and non-bleeding is done for super pixel to super pixel of WCE image. Color is the main feature while bleeding detection and SVM is trained with RGB color. Training data set for SVM constitute of 60,000 feature vectors comprising of 20,000 bleeding pixels and 40,000 pixels of non-bleeding. The whole data set comprised of 5000 WCE image, containing 4000 non-bleeding

and 1000 bleeding images collected from 20 patients. Different radial basis of SVM are implemented such as RBF, Linear and Poly. The results achieved using RBF as radial basis for SVM generate a higher accuracy, specificity and sensitivity having values of 95.00%, 94.00% and 99.00% respectively (Fu et al., 2013).

Moreover, examination and detection of bleeding in RGB normalized color space is done for WCE images. The proposed technique is grounded on variation profile of inter plane intensity that extract effective region of interest (ROI) (Kundu et al., 2018). Normally, the region surrounding in WCE images are having unnecessary information and analysis in RGB color space have certain limitation due to the high correlation of channels making the detection of bleeding hard. For remedy, this study is proposing a normalized color scheme represented by (rgb) having range of (0, 1) after transformation of coordinates of each pixel (Cheng et al., 2001; Andreadis et al., 1990). A pre-processing step for ROI extraction is done to have more discriminative features that indicate only bleeding areas rather entire image pixels consideration. The ROI identification is most crucial step in the proposed study that is done on normalized plane. Based on linear separable boundary, a condition is made in order to distinguish between bleeding and non-bleeding areas as given by,

$$r(i,j) \geq n \times g(i,j) \tag{9}$$

Here, "n" indicate a slope of discriminative line r-g plane. The pixels fulfilling the condition are considered to be ROI. Segmentation of ROI is done to provide certain of bleeding image for testing yet not assuring that exact bleeding region, leading to extract features from ROI for further classification. Features extraction is done at the ROI via histogram from values of pixel of normalized planes of WCE image as it can reproduce the areas of bleeding even if it is small or large in some bins. From the intensities of pixels of the corresponding plane of g, histograms are generated followed bin frequency values suggestions as bleeding region features. The features extracted are inputted to the K-nearest neighbor (KNN) classifier to classify non-bleeding and bleeding images. A distance function is calculated among test and train data sets that become test data set for the classifier. After bleeding image detection, marking the bleeding region automatically is helpful for physicians to identify the area. Morphological operation of two steps is done in order to detect bleeding in the ROI frames selected and classified. The proposed method outperformed other techniques and generated detection accuracy, sensitivity and specificity of 97.86%, 95.20% and 98.32% respectively (Kundu et al., 2018).

An automatic technique for the detection and segmentation of bleeding using local features based on block-based segmentation is proposed. The kvasir data set that has features such as ColorLayout, JCD, EdgeHistogram, Tamura, PHOG, and AutoColorCorrelogram are all global features that are only suitable for image classification; however unable to localize the bleeding region within an image (Obukhova et al., 2019). The block-based segmentation dividing

the image into 16×16 block followed by the extraction of features (color or brightness information, HOG (block size of 2×2), the amount of detail (Ros-Rosenfeld-Toy metric)) from the individual block in RGB, CieLAB, and HSCV color space. For eliminating the features having high mutual linear reliance and to filter the features that do not vary from different classes, the covariance matrix was calculated leaving only H, S, a, b, and Ros features. For the classification, linear discriminant analysis (LDA), support vector machine (SVM), random decision forests (RDF), and ADA boost are implemented with 10-fold of cross-validation. The data set comprised of 110 images having 70 images of bleeding and 20 of normal images generating overall high performance in terms of accuracy, sensitivity, and specificity of 94.40%, 84.70%, and 96.80% respectively for RDF (Obukhova et al., 2019).

Furthermore, a color histogram of block statistics, namely CHOBS approach is proposed to detect bleeding inside WCE video. Preprocessing is done in the first step of the analysis to remove black pixels of edges and corner areas. The CHOBS extract block-based local feature from each color channel offering better feature representation in contrast to single pixel-based feature (Ghosh et al., 2018b). By integrating local block of three channels of RGB color space, an index value is determined. For obtaining the index value of per pixel, the binary values of L most significant bits (MSBs) of the individual color plane are used. From the extracted local features, extraction of color histogram-based global features is done. A color histogram, which is extracted from those index values, presents a distinct color texture feature. To select bin with the higher color histogram and information for bleeding images, feature selection and dimensionality reduction are done. For the former, histogram pattern is utilized while for the latter principal component analysis (PCA) is applied. To classify the non-bleeding and bleeding WCE images, K-nearest neighbor (KNN) classier is employed. The proposed technique also localize the bleeding region in addition to the detection of bleeding by classifying all blocks of an image into non-bleeding and bleeding classes. Then recognizing the label of every pixel within a block for bleeding or non-bleeding followed by morphological operation fine-tuning. The dataset consisted of 2350 WCE images out of which 1900 non-bleeding and 450 bleeding images. For the detection of the bleeding frame, the proposed technique generated an accuracy, specificity, and sensitivity of 97.85%, 99.15%, and 99.47% respectivelyGhosh et al. (2018b).

Another work that is extending Bag of Words technique (Sivic and Zisserman, 2003; Csurka et al., 2004) taking advantage of color feature extraction to classify bleeding images from non-bleeding images is proposed (Yuan et al., 2015a). Avoiding whole image analysis, region of interest (ROI) is outlined by considering extreme square inscribed in the circular image representing main image features. Color features extraction is done in "YCbCr", "RGB", "HSV" and "LAB" via opting 10% normal images and 10% bleeding images. Using k-means clustering algorithm (Kanungo et al., 2002) that get input represented image pixel vectors in the color space, consistent cluster centers are calculated. The cluster centers that are concatenated generated from normal WCE data set and bleeding data set become vocabulary of visual words. Mapping 3D color data on each image point to closest visual words and computing the amount of each visual word resulting a histogram (w, d). In K-size clusters, wi indicate the ith visual word while di count occurring frequency. SVM (Chang and Lin, 2011) and KNN used as a classifier for discriminating bleeding form normal images. For localization of bleeding area, a two-stage saliency extraction strategy is applied. The first stage saliency comprises of different weight assignments particularly in CMYK and CIELAB color space. The second stage is in "RGB" color space, where bleeding region is defined by R value threshold. Regions similar to value of red colors are considered to be bleeding region which are then integrated for all three channels saliency maps. At the end, the two saliencies are fused together having bleeding localization information with different aspects, and is given as:

$$S_{final} = w_1 \times S_{stage1} + w_2 \times S_{stage2} \tag{10}$$

where S_{final} is the fused saliency map $w_1 + w_2 = 1$ The data set consisted of 2400 WCE images obtained from 10 different patients, out of which 2000 are normal images and 400 are bleeding images. The proposed technique generated best results for setting cluster centers 80 having accuracy, sensitivity and specificity of 95.75%, 92.00% and 96.50% respectively (Yuan et al., 2015a).

Moreover, an efficient extraction of features scheme is suggested to detect bleeding frame in WCE video automatically. Pre-processing is done in order to remove the black pixels in the outer regions of WCE images having low intensities resulting desired shape in the form of circular image. As red and green channels greatly influence the bleeding and non-bleeding regions, so instead of red (R), green (G), and blue (B) analysis independently a transform plane is constructed based on ratio of pixel intensity of (R/G) planes (Ghosh et al., 2015). Due to illumination variation in WCE images, analysis of single pixel can lead to distortion and computational burden, and extracted features might not be offer reliable characteristics. So extraction of features is done based on block pixels as the R/G ration is high for red (R) for bleeding pixels. The study suggested an R/B domain maximum value of block pixels for extracting potential features. For specific color spatial distribution, block feature histogram is computed after construction of R/G plane followed by feature extraction for each block. For bleeding images the patterns of histogram vary significantly from non-bleeding images. For classification, the K-nearest neighbor (KNN) is used as a function based on Euclidean distance for the calculation of features in the test image and K neighboring patterns in the data set of training. The data set consisted of 1000 WCE images, out of which 800 are non-bleeding and 200 are bleeding images. The proposed technique after setting K=10 for testing achieved accuracy, sensitivity and specificity of 96.10%, 96.48% and 96.01% respectively (Ghosh et al., 2015).

A semantic segmentation with combination of CNN approach is implemented for detection of gastrointestinal angiectasia (GIA) inside small-bowl (SB) capsule endoscopy (Leenhardt et al., 2019). The SB capsule endoscopic frames are collected from Computer Assisted Diagnosis for Capsule Endoscopy database (CAD-CAP) containing 6360 still frames. Manual annotation of 6360 still frames is done, in which extraction of 2946 of still frames having vascular lesion. The non-hemorrhagic still frames are chosen by experts same as for vascular lesion. Two data sets are made i.e. A1 and A2 that are used for training and testing purpose respectively. Another two data sets i.e. N1 and N2 comprising of 20,000 normal still frames are extracted which is used for training and testing purpose collected from SB-CEs database. Initially morphological features such as color, shape, contour and size are extracted by designing a handcrafted colorimetric segmentation method based on image saturation variation and red over green ratio. For more satisfactory results, CNN is implemented as it extracts first the finest features and then classification is done based extraction of features. An acceptable distinction during training and learning phase is necessary among GIA still frames by using dataset NI as a control for A1. Similarly, for N2 and A2 for testing phase. Two data sets are created after identification of 600 non-hemorrhagic GIA splitting into 300 still frames each (A1 and A2) from 2946 still frames having vascular lesions. Similarly, 600 normal still frames are chosen for creating control data sets i.e. (N1 and N2). The proposed technique of segmentation with a CNN model for extraction of deep features generated a specificity and sensitivity of 96.00% and 100% for detection of GIA respectively (Leenhardt et al., 2019).

Furthermore, an automatic scheme is implemented to detect angiodysplasis lesions by categorizing latent region of interest

and then classifying them by combination features such as color-based, statistical, texture and morphological operation. To solve the issue of unbalanced sampling among non-pathological and pathological regions, a boosted decision tree classification technique is opted. Image pre-processing is done by applying a contrast-limited adaptive histogram equalization (CLAHE) for enhancing the local contrast among intestinal background and an angiodysplasia regions followed by application of "RGB" de-correlation stretching for detecting local heterogeneity (Noya et al., 2017). Potential region selection is started by analyzing that green channel have high color contrast among background and lesion, so normalized threshold of 0.15 is selected to generate a mask which act as kernel for segmentation scheme. For homogeneous regions, hole filling algorithm which is morphological reconstruction is applied followed by erosion process and finally division of regions into smaller regions which are reflected as initial candidate regions. Filtering for the initial nominee regions is done heuristically by determining a threshold of heuristic for the region, potential degree of regions and perimeter via their geometric properties. To find weather the region is a lesion or not, the degree of region is the ratio among pixels in the area and circumscribed rectangle pixels. Splitting of ROI is done to avoid the addition of non-lesion pixels by using intensities variance in each channel defining as a scale for region heterogeneity. The features extracted for an angiodysplasia lesion are color, texture and geometric and the proposed work extracted 11 statistical, 8 geometrical and 5 texture features. The features are extracted in 4 different color spaces which are "YCbCr, "RGB", "HSV" and "CIElab" resulting 200 total features. A step wise regression analysis to avoid feature redundancy which reduce features set to 80. Classification of latent ROI is done as 97.70% region is recognized as non-pathological regions by the ROI detection scheme leading to unbalanced data set. To alleviate this effect, RUSBoosted (Seiffert et al., 2009) trees classifier is implemented yielding results in the form of receiver operating characteristic (ROC) and area under the curves (AUC). Total data set consisted of 45,664 and 1029 non-pathological and pathological samples respectively for training and testing in which 22,832 and 514 regions are non-lesion and with lesion respectively. Different classifiers with 10-fold cross validation are evaluated for unbalance data sets yielding high accuracy, still the RUSBoosted approach in comparison generated much more higher accuracy of 96.80% with AUC of 98.69% (Noya et al., 2017).

An automatic detection for obscure bleeding using a novel color feature extraction method is presented using support vector machine (SVM) as a classifier (Liu and Yuan, 2009). The three matrices represented by M1, M2, and M3 are used to store three channels intensities of RGB color space. Rather than extracting features from whole WCE image, down sampling is done by block division phenomena that have only ROI followed by pixel transformation into vectors called "raw vectors". The proposed feature selection is based on the ratio of intensities between red and green as M1/M2 for bleeding as blue channel (M3) have no such influence for bleeding region. SVM from Chang and Lin (2001) with different function as kernel is used for classification purpose having input fed from extracted feature step. The data set consisted of 800 images out of which 400 are normal and 400 are bleeding images. The proposed technique using ratio features generated sensitivity and specificity more than 99% compared to raw vectors and color histogram features. Implementing SVM using linear as a kernel generated a high average specificity while a high average sensitivity is achieved while using RBF as a kernel (Liu and Yuan, 2009).

Furthermore, a universal extraction of feature based technique is proposed that use a transform color domain analysis to discriminate bleeding region in WCE images. Pre-processing is done to remove the dark sides of WCE images leaving only the central portion. Due to illumination variation and high correlation among RGB channels, bleeding discrimination become hard from normal images as it share the same color property (Ghosh et al., 2018a). So a narrow band histogram is constructed on the pixel intensity ration of G/R as blue color is less significant for bleeding case. Features are extracted on G/R ratio plane overcoming the illumination problems. High and low order statistical features such as median, variance, and kurtosis are extracted from that transformed color domain. A count based feature from pixel intensity ratio is done, that make a threshold value T_G/R for discriminating bleeding and non-bleeding region. Based on the threshold, the pixel intensity for bleeding should have G/R intensity of pixel ratio less than T_G/R. For classification, supervised with linear kernel and Sequential minimization optimization (SMO) is used for binary classification. Additionally, a detection approach is provided for WCE videos using a post processing process from the results obtained from classifier. An investigation zone is formed based on the analysis that bleeding and non-bleeding frames occur continuously in consecutive (L) frames. Frames are identified and using edge detection algorithm like Sobel operators, edges are removed followed by same transformation and ratio of G/R which is here 0.5 for bleeding. Morphological dilation and erosion is applied to find the connectivity and remove small regions. The data set comprised of 2350 WCE images obtained from 20 different videos, out of which 1900 are non-bleeding while 450 are bleeding images. The proposed technique using SVM with linear function as kernel generated a high sensitivity, accuracy and specificity of 97.75%, 97.96% and 97.99% respectively (Ghosh et al., 2018a).

Based on segmentation and choosing channels of color for feature analysis having a lot of information for bleeding regions, a novel technique is proposed for bleeding detection. The color space conversion and selection of channel is considered as a pre-processing. The hue, RGB and CIE lab are chosen for specific feature analysis where by a naïve approach for every value of pixel in each channel, an index is considered making a look up table (LUT) (Hajabdollahi et al., 2018). Increased in the LUT content in their corresponding indices can be seen for bleeding regions in the data set. Using this approach, superlative color features such as saturation, gray scale, and "a" is chosen. A multi-layer perception (MLP) artificial neural network (ANN) with hidden layer as 40–20–8 is implemented on three selected color channels for classification. Patches having dimensions of 5×5 are extracted from three colors spaces becoming input for MLP where central pixel class during training phase is reflected as network output. The sigmoid function is used for binary decision and later quantized by weights adjustment. The network is trained and tested with 50 images with 5 fold cross validation and results are achieved in terms of DICE of high average value of 84.03% (Hajabdollahi et al., 2018).

A novel approach for automatic bleeding detection in WCE images assimilating features like handcrafted (HC) and convolutional neural network (CNN). The proposed technique efficiently discriminate bleeding region from non-bleeding even on small training data set. The technique comprised of three steps, extraction of features, combination of these features and finally classification process (Jia and Meng, 2017). For HC features, k-means clustering is applied to get information in the form of 2-D histogram for each image. Obtaining HC feature vector in 1-D for each WCE input image by saving values of y-axis that represent pixel number fitting to each cluster. Lab color space is suited for such feature extraction having length of 50 for HC feature vector (Jia et al., 2016). Feature extraction via CNN (Krizhevsky et al., 2012) is done by making 8 layered CNN model comprising of 3 conv layers, 3 pooling layers and 2 fully connected layers with leaning rate, momentum and weight decay of 0.01, 0.9 and 0.004 respectively. The data set comprised of 1500 WCE images obtained from 80 different patients, out of which 1200 are normal images while 300 are bleeding images having dimension of $240 \times 240 \times 3$. The proposed integrated approach resulted in higher values of precision, recall and F1 score of 94.79%,

Table 6
Efficacy of existing methods for bleeding detection in WCE images.

Ref.	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Ma and Yuen (2011)	97.76	95.59	97.76	No
Yuan et al. (2015b)	No	No	No	94.70
Leung and Malik (2001)	No	85.60	93.10	No
Lowe (2004)	95.00	94.00	99.00	No
Fu et al. (2013)	97.86	98.32	95.20	No
Chang and Lin (2011)	94.40	96.80	84.70	No
Kundu et al. (2018)	97.85	99.15	99.47	No
Obukhova et al. (2019)	95.75	96.50	92.00	No
Sivic and Zisserman (2003)	96.10	96.01	96.48	No
Csurka et al. (2004)	No	96.00	100	No
Yuan et al. (2015a)	96.80	No	No	No
Ghosh et al. (2015)	>99	>99	>99	No
Noya et al. (2017)	97.96	97.99	97.75	87.20
Seiffert et al. (2009)	No	No	No	98.00
Liu and Yuan (2009)	No	No	No	94.79

91.00% and 92.85% compared to other feature extraction based technique (Jia and Meng, 2017).

Table 6 sums the survey for the detection of bleeding in WCE images in terms of their respective performances for each method. The performance metrics such as accuracy, specificity, sensitivity, and precision for the detection of tumors by the respective approaches are listed in the table.

Furthermore, detailed information is provided in Table 7, describing the specifications and source of the collected data set for Tumor, Polyps, and Ulcers. Thus, we have provided a comprehensive details with in-depth probing of the methods underlying data set collection and the number of frames selected. Details specifying division of the data set into testing and training to conduct experimental analysis is also been provided, and lastly, cross-validation information to avoid over-fitting is presented.

Fig. 3 shows a comprehensive overview of how each anomaly is detected in WCE images based on the technique implemented. The techniques for tumor, polyp, ulcer, and bleeding detection are divided into machine learning (ML), and ML is further divided into SVM and deep learning; if the method is not a machine-learned approach, it is categorized as "other techniques".

7. Proposed cascaded classification approach

Based on the above studies done, we came up with possible future work by enlisting some of the key problems related to the images generated by wireless capsule endoscopy. The motivation behind for providing a possible future work is to offer a solution to such key issues.

7.1. Motivation

Primarily wireless capsule endoscopy is hectic and time consuming task and the output generated should be refined up to the perceptual quality of physicians to diagnose the diseases timely. Firstly, the output generated from WCE is compressed frame due to the limitation of battery life and storage capacity of WCE device resulting in the degradation of the quality of images. Secondly, transmitting WCE frames to a remotely placed physicians can also degrade the quality of the images as a communication channel are vulnerable to noises. Finally, the technique implemented previously either focusing the conventional way of detecting (tumor, polyp, and ulcers) comprising of pre-processing step followed by specific classifiers. Moreover, mostly the WCE frames while transmission to remote physicians gets affected by additive white Gaussian noise (AWGN), and compression artifacts, etc. leading the necessity of denoising approach before the frames are classified into normal and abnormal frames. To cope with such issues, we proposed a cascade approach of a neural network consisting of two neural networks as explained below.

7.2. Proposed approach for joint classification of tumors, polyps, and ulcers

The proposed approach for the classification of tumors, polyps, and ulcers comprise a cascaded approach of neural network where two neural networks in a feed-forward way are implemented. To the best of knowledge, the proposed cascaded approach is have not been implemented nor suggested for the classification of GI diseases when dealing with WCE data.

7.2.1. Denoiser model

The proposed cascaded approach is illustrated in Fig. 4 where it can be seen that there are two neural networks, i.e. denoising convolutional neural network (DnCNN) (Zhang et al., 2017) and convolutional neural network (CNN).

The proposed cascaded model for joint classification of tumors, polyps, and ulcers shown in Fig. 4 has the input of WCE images having dimensions of 320×320. The WCE frames during the generation phase get compressed and while transmitting over a communication channel to a remote physician gets degraded in terms of perceived quality leading to a time-consuming task for diagnosis. Therefore, the WCE images is passed through a machine-learned denoiser DnCNN that consists of training the network with noisy WCE (tumor, polyps, and ulcers) images. The concept of implementing DnCNN as a residual learner to compute the amount of noise and subtracting that noise from noisy model by a feed-forward convolutional neural network (CNN). To achieve this, a feed-forward convolutional network is fed with a WCE (tumor, polyps, and ulcers) images. The architecture of DnCNN comprises of three layers i.e. Conv+ReLU, Conv+BN+ReLU, and Conv with a depth of "D" as shown in Fig. 5. For the first layer of the network Conv+ReLU, 64 filters having a size of $3 \times 3 \times c$ are implemented to produce 64 feature maps; for non-linearity, rectified linear units (ReLU max (0,.)) are used. Here "c" shows the number of channels, three in this case, concerning the colored WCE (tumor, polyps, and ulcers) of the network. For layer 2, Conv+BN+ReLU with 64 filters size of $3 \times 3 \times 64$ is chosen with the inclusion of batch normalization between convolution and ReLU. For reconstructing the WCE (tumor, polyps, and ulcers) images, the final layer with Conv having $3 \times 3 \times 64$ is used inside the network (Zhang et al., 2017).

Here the DnCNN model gets an input in the form of noisy WCE (tumor, polyps, and ulcers) images as $y = x + v$, where adoption of residual learning is done to train $R(y) = v$ as a residual mapping generating $x = y - R(y)$. The averaged mean squared error between the desired residual mages and estimated ones from the noisy input can

Table 7

Specifications of the collected data set of tumor, polyp, and ulcer.

| Ref.	Data set origin	No. of frames in data set	Testing data	Training data	Cross validations
Alizadeh et al. (2017)	Shariati Hospital in Tehran, Iran (M2A Micro capsule)	315	120	110	Unspecified
Vieira et al. (2020)	Capucho's Hospital in Lisbon	936	Unspecified	Unspecified	10
Tsuboi et al. (2020)	Hiroshima University Hospital, the University of Tokyo Hospital, and Sendai Kousei Hospital.	14,922	10,488	2237	Unspecified
Liu et al. (2016a)	Obtained by four Gastroenterologists (Capsule data)	1800	Unspecified	Unspecified	10 fold
Li et al. (2011)	Obtained by experts (Given Imaging CE)	1200	900	300	4 fold
Karkanis et al. (2003)	Olympus CF-100 HL endoscope	1380	1200	180	Unspecified
Faghih Dinevari et al. (2016)	PillCam Company http://www.givenimaging.com/	800	200	600	Unspecified
Li and Meng (2009a)	PillCam SB2	300	Unspecified	Unspecified	3 fold
Li and Meng (2012b)	PillCam SB2, J. Lau and Y. Chan, two experts in Prince of Wales Hospital, Hong Kong	1200	60	540	10 fold
Barbosa et al. (2009)	Hospital dos Capuchosin Lisbon by Doctor Jaime Ramos	400	200	200	Unspecified
Li et al. (2017)	Unspecified	12, 090	20% (2418)	80% (9672)	Unspecified
Ashokkumar and Subramanian (2014)	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified
Maghsoudi and Alizadeh (2018)	Dr. Hossein Asl Soleimani and Shariati Hospital	60	12	48	Unspecified
Barbosa et al. (2008)	Hospital dos Capuchos in Lisbon	396	192	204	Unspecified
Sindhu and Valsan (2017b)	Pillcam^R SB, device by Given Imaging	435	15	70	15 fold
Sindhu and Valsan (2017a)	Pillcam^R SB, device by Given Imaging	1385	15	70	15 fold
Vieira et al. (2015)	Capucho's Hospital in Lisbon	3200	Unspecified	Unspecified	15 fold
Karargyris and Bourbakis (2011)	Digestive Specialists, Dayton, OH	50	Unspecified	Unspecified	Unspecified
Billah et al. (2017)	(http://www.depeca.uah.es/colonoscopy_dataset/) and (https://polyp.grand-challenge.org/databases/)	14,000	7000	7000	Unspecified
Billah and Waheed (2018)	(http://www.depeca.uah.es/colonoscopy_dataset/) and (https://polyp.grand-challenge.org/databases/)	14,000	4200	9800	K-fold
Yuan and Meng (2014)	Unspecified	872	272	600	Unspecified
Li and Meng (2012a)	M2A WCE, Given Imaging Company	2500	1250	1250	10 fold
Yuan and Meng (2017)	James Lau in Prince of Wales Hospital in Hong Kong, Pillcam SB WCE system	3000	Unspecified	Unspecified	3 fold
Yuan et al. (2016)	Qilu Hospital, Shandong University M2A WCE, Given Imaging Company	2500	1250	1250	Unspecified
Li et al. (2009)	James Lau in Prince of Wales Hospital in Hong Kong, Pillcam SB WCE system	300	Unspecified	Unspecified	3 fold
Hwang and Celebi (2010)	Unspecified	128	Unspecified	Unspecified	Unspecified
Souaidi et al. (2018)	WEO Clinical Endoscopy Atlas	6.790	1358	5432	10 fold
Bae and Yoon (2015)	CVC-ColonDB and Augmented Dataset	1263	Augmented Data set	CVC-ColonDB	5 fold
Ansari and Charfi (2017)	www.capsuleendoscopy.org	2433	1283	1150	Unspecified
Sornapudi et al. (2019)	(Colonoscopy): CVC-ColonDB, CVC-PolypHD, and ETIS-Larib WCE: Mayo Clinic	Colonoscopy (552) WCE (1800)	Unspecified	Unspecified	K-fold
Aoki et al. (2019)	Ethics committee of The University of Tokyo (no. 11931) and Japan Medical Association (ID JMA-IIA00283) Pillcam SB2 or SB3 WCE device (Given Imaging, Yoqneam, Israel).	15,800	10,440	5360	Unspecified
Lee et al. (2019)	Unspecified	5500	Unspecified	Unspecified	Unspecified
Charfi et al. (2019)	(1) WEO Clinical Endoscopy Atlas, (2) C. Endoscopy: 'Capsule endoscopy products'	1 (446) 2 (2170)	1 (235) 2 (1120)	1 (211) 2 (1050)	3 fold
Souaidi et al. (2019)	(1) PillCam SB3 (Cheik Zaid Hospital, Rabat, Morocco) (2) James Lau in Prince of Wales Hospital in Hong Kong,	1 (240) 2 (212)	1 (48) 2 (42)	1 (192) 2 (170)	5-fold
Alaskar et al. (2019)	Dr Khoroo'sMedical Clinic/Trust. Available online: http://www.drkhuroo.in/#	1875	105	421	Unspecified
Fan et al. (2018)	Unspecified	8250	500	2000	Unspecified
Yuan et al. (2017)	Pillcam SB WCE system and the Rapid Reader 6.0 software (both from Given Imaging Co., Ltd.)	1650	20% (330)	80% (1320)	5 fold
Candès and Donoho (2000)	Unspecified	260	Unspecified	Unspecified	10 fold
Eid et al. (2013)	Jin Shan Corporation (Yang Qi, Tang Du Hospital of Fourth Military Medical University)	100	45	45	10 fold
Dong (2000)	James Lau in Prince of Wales Hospital in Hong Kong, Pillcam SB WCE system	3600	Unspecified	Unspecified	4 fold
Liu et al. (2012)	Given Imaging Ltd.	584	240	344	Unspecified
Do and Vetterli (2003)	Pillcam SB WCE system and the Rapid Reader 6.0 software (both from Given Imaging Co., Ltd.)	340	20% (68)	80% (272)	5 fold
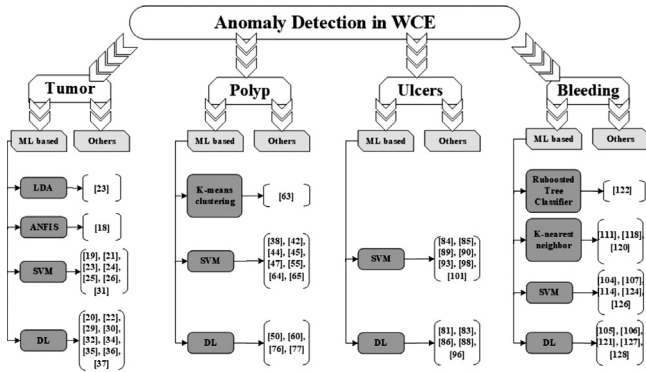
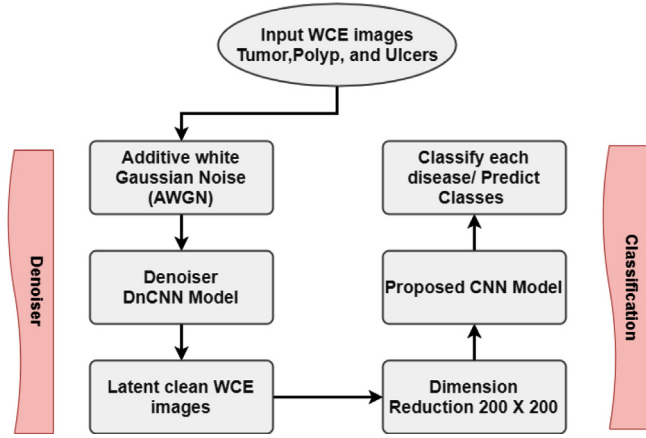**Fig. 3.** Detailed information of anomaly detection in WCE images.

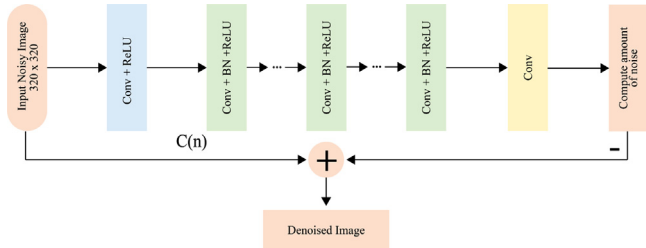**Fig. 4.** Proposed cascaded model for joint classification of tumors, polyps, and ulcers.

Fig. 5. Denoising convolutional neural network for noisy WCE images.

be adopted as losses in the DnCNN's function to learn the trainable parameters "θ" (Zhang et al., 2017).

$$l(\theta) = \frac{1}{2N} \sum_{i=1}^{N} \|R(y_i; \theta) - (y_i - x_i)\|_R^2 \qquad (11)$$

where $\{(y_i, x_i)\}_{i=1}^{N}$ indicates "N" noisy-clean training patch (images) pairs.

We implemented a pre-trained model of DnCNN that uses BSD68 and Berkeley segmentation data set for testing and training respectively under the Gaussian noise ranging from $\sigma \in [0\text{–}55]$. For training the network, 128×3000 patches having a size of 50×50 are cropped. To obtain sufficient spatial information for denoising, the network depth is set to 20 and used SGD as an optimizer with 0.0001 of weight decay, 128 of mini-batch size and 0.9 momentum. Exponential decay in the learning rate was achieved from 0.01 to 0.0001 for epochs of 50 (Zhang et al., 2017).
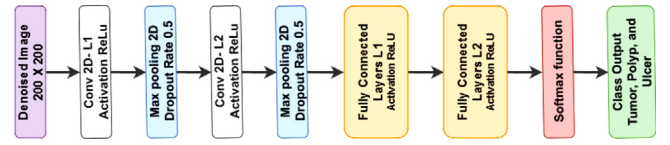

Fig. 6. Convolution neural network (CNN) for classification of (tumor, polyp, and ulcer).

Table 8
Simulation parameters used for CNN model.

Network parameters	Configuration values
Input image size	200×200
Optimizer	Adam
Learning rate, η	0.001
Batch size	32
Epochs	30

7.2.2. Proposed CNN model for joint classification

The denoised images obtained from DnCNN operation are then used for the classification purpose; starting with the reducing of denoised images from 320×320 to 200×200 which are to be fed to CNN model. The network architecture is illustrated in Fig. 6, comprising of 2 layers (L1 and L2) of convolution, 2 layers of max pooling having dropout of 0.5, 1 dense layer having activation function (Softmax), and 1 output layer for classification output as (tumor, polyp, and ulcer). For the first layer (L1), input image with dimension 200×200 will be passed having a filter size of 32 with rectified linear unit (ReLU) as an activation function. The same configuration will be used for L2 reducing the dimension of the image to 99×99 followed by the Max pooling layer of size 2×2 with drop out of 0.5. Flattening via two fully connected layers (FC1 and FC2) is achieved for a single continuous linear vector that is connected to a Softmax layer. The last layer comprise of class output layer for classification purposes that predict the class of the disease. The simulation parameters used for the joint classification of tumor, polyp, and ulcer are given in Table 8.

7.2.3. Data set and experimental results

The acquisition of the WCE data set is a crucial task when classification is of multi-class; three different data sets for tumor (Usman et al., 2016), polyp (Bernal et al., 2015), and ulcer (Sharif et al., 2019) are obtained from three different data set sources. For the implementation, 100 images for each disease are acquired making the whole data set of 300 WCE images after making the dimension equal in size of 320×320. The training and testing data are divided into 80% and 20% respectively, keeping each set comprising of an equal number of pathogenic and normal images. Inspired by the previous work (Nawarathna et al., 2014), the experimental analysis for the proposed joint classification model was conducted on a computer with Intel(R) Core i7 2600K, 64-bit, 3.40 GHz processor, and 16 GB memory. For the simulation environment, we used MATLAB R. 2018b. For the frame per second (fps), mean processing time per frame (MPT) and reaction time (RT) (Shin et al., 2018) in terms of frames and seconds, which is a detection time taken by our method in order to process and display the output is computed. The fps achieved by running the code 5 times and then taking the average resulted in 4 fps or 0.22 sec for the proposed approach.

The performance of the proposed joint classification of the tumor, polyp, and ulcer is evaluated separately for each disease in terms of accuracy, sensitivity, specificity, and precision in tabular form. Table 9 show the classification performance for tumor, polyp, and ulcer respectively after the implementation of proposed cascaded models. In the case of joint classification of WCE images under artifacts such as AWGN noises, the proposed cas-

Table 9
Performance evaluation for each disease.

	Disease name	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Proposed joint classification model	Tumor	96.98	96.40	97.56	96.39
	Polyp	97.51	97.12	97.91	97.13
	Ulcer	98.30	98.30	98.01	98.31

caded approach of using DnCNN as a denoiser and CNN as a classifier resulting in better performance as shown in Table 9.

8. Formal discussion

This section comprises of a formal discussion throwing light on the approaches made to detect various anomalies found in WCE images. The formal discussion section also encompasses the bench marking of our possible proposed solution with the approaches been adopted for each anomaly detection.

For tumor section, the techniques adopted to detect tumor inside WCE images are mostly based on SVM (Liu et al., 2016a; Li et al., 2011; Faghih Dinevari et al., 2016; Li and Meng, 2009a, 2012b; Ashokkumar and Subramanian, 2014) and deep learning (DL) (Alizadeh et al., 2017; Karkanis et al., 2003; Barbosa et al., 2009, 2008; Li et al., 2017; Mohapatra et al., 2012; Sindhu and Valsan, 2017a, 2017b) after extracting features. Preprocessing step for each technique is an essential step followed by a classification approach. For instance, in Li and Meng (2012b) LBP is used as a texture feature extractor with two more feature extractor i.e. SVM-SFFS and SVM-RFE on 1200 images yielding good performance. Similarly, in Li et al. (2017) the data set is large, implementation of CNN models like LeNet, etc., and high-performance results. Albeit, each technique is following a generalized rule of pre-processing followed by a classification approach for classifying only one disease at a time, none of the studies are attempting to address the issues such as JPEG compression during the generation of WCE videos and AWGN effects while transmitting these frames to a remote physician. The proposed joint classification model is concentrating on the cascaded approach for denoising these artifacts via DnCNN (Zhang et al., 2017) and then implementing the proposed CNN model that categorizes tumor, polyp, and ulcer jointly.

Similarly for polyp section, in the proposed survey paper are based on SVM (Karargyris and Bourbakis, 2011; Billah et al., 2017; Yuan and Meng, 2014; Li and Meng, 2012a; Yuan et al., 2016; Ansari and Charfi, 2017) and DL (Yuan and Meng, 2017; Li et al., 2009; Bae and Yoon, 2015) where features are extracted and those features are used as an input for classification for each approach. Take the case in Karargyris and Bourbakis (2011) where SUSAN and Gabor filters are used for edge detection process in HSV color spaces. Then SVM as supervised learning is implemented for the classification of polyp and ulcer as the data set consisted of both diseases. Also in Yuan and Meng (2014), M-LBP is used as a feature extractor for spatial locality followed by SVM. Furthermore, for the DL approach, the Zernike moment analysis is done for feature extraction in Li et al. (2009) in HSI color space followed by MLP as a classifier. Although each method is following a generalized rule of preprocessing accompanied by a classification method for classifying only one disease at a time, none of the studies are striving to address the concerns such as JPEG compression during the generation of WCE videos and AWGN effects while transmitting these frames to a remote physician. The proposed joint classification model is concentrating on the cascaded approach for denoising these artifacts via DnCNN (Zhang et al., 2017) and then implementing the proposed CNN model that categorizes tumor, polyp, and ulcer jointly.

For ulcer detection research in specific survey comprises of SVM (Yuan et al., 2017, 2015b; Eid et al., 2013; Yu et al., 2012) and DL (Aoki et al., 2019; Alaskar et al., 2019; Fan et al., 2018; Li and Meng, 2009b) approaches. The attempt made as in Yuan et al. (2017) to use HSC-SIFT for segmentation and textural and color information extraction followed by SVM. Moreover, a DL approach in Li and Meng (2009b) for ulcer detection using LBP for textural analysis followed by the MLP neural network for classification purpose. The same justification is provided as given above for the comparison of the proposed possible solution with work done previously.

Furthermore, efforts are made to reflect the performance analysis of each disease and section for their respective technique in the survey along with performance metrics and with their references, as shown in Tables 2–6. Table 7 illustrates in detail the source of data set collection along with training and testing of images with cross-validation to show the trade-off between the number of images with the approach being made along with its performance.

9. Recent review articles and comparison

This section details some of the recent relevant review articles for WCE imaging. The section also attempts to explain the limitations of these review articles by bench marking it with our presented survey.

An attempt is made to review the issues facing while detecting lesion and interpretation in WCE as the earlier capsule imaging (CE) was having frame rates of 2 fps such as PillCam SB2 (Kim et al., 2018). This can lead to miss detection of the small bowel as CE moves inside the GI tract with peristalsis. The new PillCam SB3 provides a frame rate of 2–6 fps reducing this issue to some extent while giving more burden to process due to a large number of frames generation. The study also presented an interpretation constrain for trainees need to perform at least 20 supervised capsule interpretations before they are approved in addition to localization approaches of the lesion in the CE frame (Kim et al., 2018).

In comparison, our work is attempting to present a detailed overview of the state of the art approaches made for the detection of diseases such as a tumor, polyp, and ulcer by providing the performance of each technique in the tabular form. Moreover, our work is also providing the study done for the anomalies and bleeding detection that can be the reason for diseases such as a tumor, polyp, and ulcer. Complete detailed information for the dataset source, detailing the total number of images, testing, training, and cross-validation in tabular form followed by division of techniques on the basis of approaches are provided that are lacking in Kim et al. (2018). The study also misses a possible proposed future model (Kim et al., 2018), that is included in our proposed work by detailing the architecture of the model proposed.

Moreover, an extensive survey is presented for medical imaging analysis in CE to present research work done by dividing the approaches into image and video compression, video summarization and redundant image elimination, segmentation and region identification, image enhancement and interpretation, computer-aided abnormality detection in capsule endoscopy, and segmentation and region identification (Jani and Srivastava, 2019). In addition to that, the survey also provided experimental setup and results.

Though the study provided a detailed survey for medical image analysis, still it did not provide the performance of each technique based on their approaches for the detection of a specific disease (Jani and Srivastava, 2019) that can be found in our presented sur-

vey. No formal discussion section is there that can enlighten the insights of each technique, and there is no proposed future model with the system architecture (Jani and Srivastava, 2019) for each approach that in contrast can be found in our presented survey.

Deep learning is applied in many areas that greatly improved performance, particularly in image-based classification and recognition (Hwang et al., 2018). This study reviewed recent articles focusing on deep learning-based approaches for diagnosis and provided a future direction for the application of artificial intelligence (AI) in CE. The study focused on lesion detection by reviewing deep learning articles by limiting the study to polyp/ulcer, and hookworms in addition to models like Alexnet, CNN, and VGG-Net. The study provided performance analysis results in the form of tabular form for those models in terms of accuracy and precision (Hwang et al., 2018).

In comparison to the survey that limits the study for the detection of polyp and hookworm focussing only models like Alexnet, CNN, and VGG-Net, and having no tabular analysis for the performance of each approach (Hwang et al., 2018); our survey is providing a detailed overview of the state of the art approaches made for the detection of diseases such as a tumor, polyp, ulcer, anomalies, and bleeding in WCE images by providing the performance of each technique in the tabular form. Complete detailed information for the dataset source, detailing the total number of images, testing, training, and cross-validation in tabular form followed by division of techniques on the basis of approaches are provided that are lacking in Hwang et al. (2018). The study also misses a possible proposed future model (Hwang et al., 2018), that is included in our proposed work by detailing the architecture of the model proposed.

Various research work has been done for the detection of polyp employing automatic image processing, machine learning approached, and computer vision. An attempt is made to review several approaches done for the detection of polyp in video capsule endoscopy images followed by challenges encountered by conventional image processing and computer vision techniques (Prasath, 2017). The study provided the insight of previous works adopted to detect polyp focusing on techniques that are based on the segmentation in VCE. For the segmentation based detection of polpy in VCE, the study focused on methods adopted such as KNN, SVM, MLP, Fisher's linear discriminant analysis (FLDA), and HMM by enlisting the performance of each approach and study. Moreover, localization within the VCE frame and holistic system based detection of polyp is also provided followed by the discussion and outlook discussing the challenges encountered by traditional image processing and computer vision methods (Prasath, 2017).

In contrast to the review that restricts the research for the detection of polyp concentrating on approaches based on conventional image processing and computer vision techniques (Prasath, 2017); our survey is providing a detailed overview of the state of the art approaches made for the detection of diseases such as a tumor, polyp, ulcer, anomalies, and bleeding in WCE images by providing the performance of each technique in the tabular form. Comprehensive information for the dataset source, detailing the total number of images, testing, training, and cross-validation in tabular is provided in our presented survey that is lacking in Prasath (2017). The study also misses a possible proposed future model (Prasath, 2017), that is included in our proposed work by detailing the architecture of the model proposed.

Furthermore, a detailed survey is provided that reviewed research work for endoscopy and WCE imaging. In addition, the goal of this survey is to provide the maturity of each technique adopted based on its possible development. The survey provided a comprehensive evaluation of methodologies by categorizing into three categories i.e. artificial intelligence techniques, signal-processing techniques, and color- and image-processing techniques (Karargyris and Bourbakis, 2010). The AI section covered tools based on machine learning aptitudes, such as artificial neural networks (ANNs) and clustering. The signal-processing section covered all techniques that include processing information in the frequency domain such as bandpass filters, Fourier transforms and, wavelet transforms. The color- and image-processing section encompassed systems regarding color processing such as color and segmentation transformations. The survey also provided a future research direction by mentioning the point that a single solution for WCE technology is unrealistic due to the circumstance of the problem for a specific disease. However, the improvement in the hardware for capturing the better quality of videos and software engineers to create an intelligent software system for dealing with the current unsolved issues is a necessary direction (Karargyris and Bourbakis, 2010).

Although the survey provided a comprehensive overview of the research work for endoscopy and WCE imaging by limiting the approaches to artificial intelligence techniques, signal-processing techniques, and color- and image-processing techniques and in tabular form. The survey missed (Karargyris and Bourbakis, 2010) the approaches adopted recently for the detection tumor, polyp, ulcer, anomalies, and bleeding in WCE images. Moreover, there is no information for the data set availability source, No. of training and testing images with No. of cross validation information that can affect that performance of specific approach that in contrast can be found in our survey (Karargyris and Bourbakis, 2010). The study also lack a possible proposed future model (Karargyris and Bourbakis, 2010), that is included in our proposed work by detailing the architecture of the model proposed.

10. Conclusion

This paper presents a comprehensive survey of state-of-the-art computer-aided methods that are used to identify, detect and classify anomalies in the GI tract of living beings capture through the WCE process. Several anomalies occur in the GI tract by this review paper only encompasses published methods related to tumor, polyp, and ulcer. The survey also details the detection approaches for the general anomalies and bleeding that can be the symptoms of the aforementioned diseases found inside WCE images. Additionally, a classification based on learning-based methods and other methods is provided as well for each of the three anomalies.

A formal discussion section is provided that also encompasses the benchmarking of our possible proposed solution with the approaches been adopted for each anomaly detection. Moreover, a detailed section of comparing relevant review articles for WCE imaging to show the limitations of previous attempts in this regard.

This paper aims to provide the readers with a single platform to gain access to the recently published articles related to computer-aided detection of diseases in WCE videos. According to the authors' extensive review and the works presented in this paper, there has been no work that provides a joint classification of all the three mentioned anomalies. Finally, we have proposed an approach for joint classification of tumor, polyp, and ulcer in WCE videos, where a cascaded approach of neural networks is employed along with details of the experimental setup, data set, and results.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34 (11), 2274–2282.

Alaskar, H., Hussain, A., Al-Aseem, N., Liatsis, P., Al-Jumeily, D., 2019. Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. Sensors 19 (6), 1265.

Alizadeh, M., Maghsoudi, O.H., Sharzehi, K., Hemati, H.R., Asl, A.K., Talebpour, A., 2017. Detection of small bowel tumor in wireless capsule endoscopy images using an adaptive neuro-fuzzy inference system. J. Biomed. Res. 31 (5), 419.

Allain, C., Cloitre, M., 1991. Characterizing the lacunarity of random and deterministic fractal sets. Phys. Rev. A 44 (6), 3552.

Andreadis, I., Browne, M., Swift, J., 1990. Image pixel classification by chromaticity analysis. Pattern Recognit. Lett. 11 (1), 51–58.

Ansari, M.E., Charfi, S., 2017. Computer-aided system for polyp detection in wireless capsule endoscopy images. 2017 International Conference on Wireless Networks and Mobile Communications (WINCOM), 1–6.

Aoki, T., Yamada, A., Aoyama, K., Saito, H., Tsuboi, A., Nakada, A., Niikura, R., Fujishiro, M., Oka, S., Ishihara, S., et al., 2019. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointest. Endosc. 89 (2), 357–363.

Ashokkumar, B., Subramanian, S.P.S., 2014. Automatic detection of tumor in wireless capsule endoscopy images using energy based textural features and SVM based RFE approach. Int. J. Eng. Res. Appl. 4 (4), 50–54.

Bae, S.-H., Yoon, K.-J., 2015. Polyp detection via imbalanced learning and discriminative feature learning. IEEE Trans. Med. Imaging 34 (11), 2379–2393.

Barbosa, D.J., Ramos, J.C.C., Lima, C.S., 2008. Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform. 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 3012–3015.

Barbosa, D.J., Ramos, J., Correia, J.H., Lima, C.S., 2009. Automatic detection of small bowel tumors in capsule endoscopy based on color curvelet covariance statistical texture descriptors. 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 6683–6686.

Bernal, J., Sánchez, F.J., Vilariño, F., 2012. Towards automatic polyp detection with a polyp appearance model. Pattern Recognit. 45, 3166–3182.

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. Comput. Med. Imaging Graph. 43, 99–111.

Billah, M., Waheed, S., 2018. Gastrointestinal polyp detection in endoscopic images using an improved feature extraction method. Biomed. Eng. Lett. 8 (1), 69–75.

Billah, M., Waheed, S., Rahman, M.M., 2017. An automatic gastrointestinal polyp detection system in video endoscopy using fusion of color wavelet and convolutional neural network features. Int. J. Biomed. Imaging.

Bishop, C.M., Nasrabadi, N.M., 2007. Pattern recognition and machine learning. J. Electron. Imaging 16, 049901.

Bourbakis, N., 2005. Detecting abnormal patterns in WCE images. In: Fifth IEEE Symposium on Bioinformatics and Bioengineering (BIBE'05). IEEE, pp. 232–238.

Candès, E.J., Donoho, D.L., 2000. Curvelets: A Surprisingly Effective Nonadaptive Representation for Objects with Edges.

Candes, E.J., Donoho, D.L., 2000. Curvelets, multiresolution representation, and scaling laws. In: Wavelet Applications in Signal and Image Processing VIII, vol. 4119, International Society for Optics and Photonics, pp. 1–12.

Chang, C.-C., Lin, C.-J., 2001. LIBSVM: A Library for Support Vector Machines [EB/OL].

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) 2 (3), 27.

Charfi, S., El Ansari, M., Balasingham, I., 2019. Computer-aided diagnosis system for ulcer detection in wireless capsule endoscopy images. IET Image Process. 13 (6), 1023–1030.

Charisis, V., Hadjileontiadis, L.J., Liatsos, C.N., Mavrogiannis, C.C., Sergiadis, G.D., 2010. Abnormal pattern detection in wireless capsule endoscopy images using nonlinear analysis in RGB color space. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE, pp. 3674–3677.

Charisis, V.S., Hadjileontiadis, L.J., Liatsos, C.N., Mavrogiannis, C.C., Sergiadis, G.D., 2012. Capsule endoscopy image analysis using texture information from various colour models. Comput. Methods Prog. Biomed. 107 (1), 61–74.

Cheng, H.-D., Jiang, X.H., Sun, Y., Wang, J., 2001. Color image segmentation: advances and prospects. Pattern Recognit. 34 (12), 2259–2281.

Collins, R.T., 2003. Mean-shift blob tracking through scale space. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2003. Proceedings, vol. 2. IEEE, pp. II–234.

Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C., 2004. Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, vol. 1, ECCV, Prague, pp. 1–2.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: International Conference on Computer Vision & Pattern Recognition (CVPR'05), vol. 1. IEEE Computer Society, pp. 886–893.

Demanet, L., Ying, L., 2007. Curvelets and wave atoms for mirror-extended images. In: Wavelets XII, vol. 6701, International Society for Optics and Photonics, p. 67010J.

Do, M.N., Vetterli, M., 2003. The finite ridgelet transform for image representation. IEEE Trans. Image Process. 12 (1), 16–28.

Dong, P., 2000. Test of a new lacunarity estimation method for image texture analysis. Int. J. Remote Sens. 21 (17), 3369–3373.

Eid, A., Charisis, V., Hadjileontiadis, L.J., Sergiadis, G.D., 2013. A curvelet-based lacunarity approach for ulcer detection from wireless capsule endoscopy images. Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, 273–278.

Eliakim, R., 2000. Where do I see minimally invasive endoscopy in 2020: clock is ticking. Ann. Transl. Med. 5 (9).

Faghih Dinevari, V., Karimian Khosroshahi, G., Zolfy Lighvan, M., 2016. Singular value decomposition based features for automatic tumor detection in wireless capsule endoscopy images. Appl. Bionics Biomech.

Fan, W., Stolfo, S.J., Zhang, J., Chan, P.K., 1999. AdaCost: misclassification cost-sensitive boosting. ICML.

Fan, S., Xu, L., Fan, Y., Wei, K., Li, L., 2018. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. Phys. Med. Biol. 63 (16), 165001.

Field, D.J., 1987. Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A: Opt. Image Sci. 4 (12), 2379–2394.

Fu, Y., Zhang, W., Mandal, M., Meng, M.Q.-H., 2013. Computer-aided bleeding detection in WCE video. IEEE J. Biomed. Health Inform. 18 (2), 636–642.

Galar, M., Fernández, A., Tartas, E.B., Bustince, H., Herrera, F., 2012. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) 42, 463–484.

Ghosh, T., Fattah, S., Shahnaz, C., Kundu, A., Rizve, M., 2015. Block based histogram feature extraction method for bleeding detection in wireless capsule endoscopy. TENCON 2015 – 2015 IEEE Region 10 Conference, 1–4.

Ghosh, T., Fattah, S.A., Wahid, K.A., 2018a. Automatic computer aided bleeding detection scheme for wireless capsule endoscopy (WCE) video based on higher and lower order statistical features in a composite color. J. Med. Biol. Eng. 38 (3), 482–496.

Ghosh, T., Fattah, S.A., Wahid, K.A., 2018b. CHOBS: color histogram of block statistics for automatic bleeding detection in wireless capsule endoscopy video. IEEE J. Transl. Eng. Health Med. 6, 1–12.

Guo, Z., Zhang, L., Zhang, D., 2010. A completed modeling of local binary pattern operator for texture classification. IEEE Trans. Image Process. 19, 1657–1663.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46 (1–3), 389–422.

Hajabdollahi, M., Esfandiarpoor, R., Soroushmehr, S.M.R., Karimi, N., Samavi, S.A., Najarian, K., 2018. Segmentation of Bleeding Regions in Wireless Capsule Endoscopy Images an Approach for Inside Capsule Video Summarization, CoRR abs/1802.07788.

Haralick, R.M., Shanmugam, K.S., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. 3, 610–621.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. Proceedings of the IEEE International Conference on Computer Vision, 2961–2969.

Hoens, T.R., Chawla, N.V., 2013. Imbalanced Datasets: From Sampling to Classifiers.

Hwang, S., Celebi, M.E., 2010. Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 678–681.

Hwang, S., Oh, J.H., Tavanapong, W., Wong, J.S., de Groen, P.C., 2007. Polyp detection in colonoscopy video using elliptical shape feature. 2007 IEEE International Conference on Image Processing, vol. 2, pp. II–465–II–468.

Hwang, Y., Park, J., Lim, Y.J., Chun, H.J., 2018. Application of artificial intelligence in capsule endoscopy: where are we now? Clin. Endosc. 51 (6), 547.

Iakovidis, D.K., Georgakopoulos, S.V., Vasilakakis, M., Koulaouzidis, A., Plagianakos, V.P., 2018. Detecting and locating gastrointestinal anomalies using deep learning and iterative cluster unification. IEEE Trans. Med. Imaging 37 (10), 2196–2210.

Inoue, T., Abe, S., 2001. Fuzzy Support Vector Machines for Pattern Classification.

Iwahori, Y., Hattori, A., Adachi, Y., Bhuyan, M.K., Woodham, R.J., Kasugai, K., 2015. Automatic detection of polyp using Hessian filter and hog features. Procedia Comput. Sci. 60, 730–739.

Jain, A., Zongker, D., 1997. Feature selection: evaluation, application, and small sample performance. IEEE Trans. Pattern Anal. Mach. Intell. 19 (2), 153–158.

Jani, K.K., Srivastava, R., 2019. A survey on medical image analysis in capsule endoscopy. Curr. Med. Imaging 15 (7), 622–636.

Jia, X., Meng, M.Q.-H., 2017. Gastrointestinal bleeding detection in wireless capsule endoscopy images using handcrafted and CNN features. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 3154–3157.

Jia, X., Cai, L., Liu, J., Dai, W., Meng, M.Q.-H., 2016. GI bleeding detection in wireless capsule endoscopy images based on pattern recognition and a MapReduce

framework. In: 2016 IEEE International Conference on Real-Time Computing and Robotics (RCAR). IEEE, pp. 266–271.

Julesz, B., 1981. Textons, the elements of texture perception, and their interactions. Nature 290, 91–97.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. (7), 881–892.

Karargyris, A., Bourbakis, N., 2010. Wireless capsule endoscopy and endoscopic imaging: a survey on various methodologies presented. IEEE Eng. Med. Biol. Mag. 29 (1), 72–83.

Karargyris, A., Bourbakis, N.G., 2011. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. IEEE Trans. Biomed. Eng. 58, 2777–2786.

Karkanis, S.A., Iakovidis, D.K., Maroulis, D.E., Karras, D.A., Tzivras, M., 2003. Computer-aided tumor detection in endoscopic video using color wavelet features. IEEE Trans. Inf. Technol. Biomed. 7 (3), 141–152.

Kim, S.H., Yang, D.-H., Kim, J.S., 2018. Current status of interpretation of small bowel capsule endoscopy. Clin. Endosc. 51 (4), 329.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 1097–1105.

Kubat, M., Matwin, S., 1997. Addressing the curse of imbalanced training sets: one-sided selection. ICML.

Kundu, A.K., Fattah, S.A., Rizve, M.N., 2018. An automatic bleeding frame and region detection scheme for wireless capsule endoscopy videos based on interplane intensity variation profile in normalized RGB color space. J. Healthc. Eng.

Larose, D.T., Larose, C.D., 2014. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons.

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2. IEEE, pp. 2169–2178.

Lee, C., Min, J., Cha, J., Lee, S., 2019. Feature space extrapolation for ulcer classification in wireless capsule endoscopy images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 100–103.

Leenhardt, R., Vasseur, P., Li, C., Saurin, J.C., Rahmi, G., Cholet, F., Becq, A., Marteau, P., Histace, A., Dray, X., et al., 2019. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. Gastrointest. Endosc. 89 (1), 189–194.

Leung, T., Malik, J., 2001. Representing and recognizing the visual appearance of materials using three-dimensional textons. Int. J. Comput. Vis. 43 (1), 29–44.

Levinshtein, A., Stere, A., Kutulakos, K.N., Fleet, D.J., Dickinson, S.J., Siddiqi, K., 2009. Turbopixels: fast superpixels using geometric flows. IEEE Trans. Pattern Anal. Mach. Intell. 31 (12), 2290–2297.

Li, B., Meng, M.Q.-H., 2009a. Small bowel tumor detection for wireless capsule endoscopy images using textural features and support vector machine. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 498–503.

Li, B., Meng, M.Q.-H., 2009b. Texture analysis for ulcer detection in capsule endoscopy images. Image Vis. Comput. 27 (9), 1336–1342.

Li, B., Meng, M.Q.-H., 2012a. Automatic polyp detection for wireless capsule endoscopy images. Expert Syst. Appl. 39 (12), 10952–10958.

Li, B., Meng, M.Q.-H., 2012b. Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection. IEEE Trans. Inf. Technol. Biomed. 16 (3), 323–329.

Li, B., Fan, Y., Meng, M.Q.-H., Qi, L., 2009. Intestinal polyp recognition in capsule endoscopy images using color and shape features. 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), 1490–1494.

Li, B., Meng, M.Q.-H., Lau, J.Y., 2011. Computer-aided small bowel tumor detection for capsule endoscopy. Artif. Intell. Med. 52 (1), 11–16.

Li, P., Li, Z., Gao, F., Wan, L., Yu, J., 2017. Convolutional neural networks for intestinal hemorrhage detection in wireless capsule endoscopy images. In: 2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE, pp. 1518–1523.

Lin, O., Kozarek, R., Schembre, D., Spaulding, W., 2005. Blinded comparison of esophageal capsule endoscopy (ECE) versus conventional esophagogastroduodenoscopy (EGD) for identification of esophagitis and Barrett's esophagus in patients with chronic gastroesophageal reflux disease (GERD). In: Proceedings of the 4th International Conference on Capsule Endoscopy, Miami, Florida.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2117–2125.

Liu, J., Yuan, X., 2009. Obscure bleeding detection in endoscopy images using support vector machines. Optim. Eng. 10 (2), 289–299.

Liu, X., Gu, J., Xie, Y., Xiong, J., Qin, W., 2012. A new approach to detecting ulcer and bleeding in wireless capsule endoscopy images. Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics, 737–740.

Liu, G., Yan, G., Kuang, S., Wang, Y., 2016a. Detection of small bowel tumor based on multi-scale curvelet analysis and fractal technology in capsule endoscopy. Comput. Biol. Med. 70, 131–138.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016b. SSD: single shot multibox detector. In: European Conference on Computer Vision. Springer, pp. 21–37.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.

Ma, A.J., Yuen, P.C., 2011. Linear dependency modeling for feature fusion. In: 2011 International Conference on Computer Vision. IEEE, pp. 2041–2048.

Ma, L., Crawford, M.M., Yang, X., Guo, Y., 2015. Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 53 (5), 2832–2844.

Maghsoudi, O.H., Alizadeh, M., 2018. Feature Based Framework to Detect Diseases, Tumor, and Bleeding in Wireless Capsule Endoscopy, CoRR abs/1802.02232.

Maghsoudi, O.H., Soltanian-Zadeh, H., 2013. Detection of abnormalities in wireless capsule endoscopy frames using local fuzzy patterns. In: 2013 20th Iranian Conference on Biomedical Engineering (ICBME). IEEE, pp. 286–291.

Mesejo, P., Pizarro-Perez, D., Abergel, A., Rouquette, O., Béorchia, S., Poincloux, L., Bartoli, A., 2016. Computer-aided classification of gastrointestinal lesions in regular colonoscopy. IEEE Trans. Med. Imaging 35, 2051–2063.

Mohapatra, S., Patra, D., Kumar, S., Satpathy, S., 2012. Lymphocyte image segmentation using functional link neural architecture for acute leukemia detection. Biomed. Eng. Lett. 2, 100–110.

Nam, H., Han, B., 2016. Learning multi-domain convolutional neural networks for visual tracking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4293–4302.

Nawarathna, R., Oh, J., Muthukudage, J., Tavanapong, W., Wong, J., De Groen, P.C., Tang, S.J., 2014. Abnormal image detection in endoscopy videos using a filter bank and local binary patterns. Neurocomputing 144, 70–91.

Noya, F., Alvarez-Gonzalez, M., Benitez, R., 2017. Automated angiodysplasia detection from wireless capsule endoscopy. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 3158–3161.

Obukhova, N., Motyko, A., Timofeev, B., Pozdeev, A., 2019. Method of endoscopic images analysis for automatic bleeding detection and segmentation. In: 2019 24th Conference of Open Innovations Association (FRUCT). IEEE, pp. 285–290.

Ojala, T., Pietikä inen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. 29 (1), 51–59.

Ojala, T., Pietikä inen, M., Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. (7), 971–987.

Pan, G., Yan, G., Qiu, X., Cui, J., 2011. Bleeding detection in wireless capsule endoscopy based on probabilistic neural network. J. Med. Syst. 35 (6), 1477–1484.

Pogorelov, K., Suman, S., Azmadi Hussin, F., Saeed Malik, A., Ostroukhova, O., Riegler, M., Halvorsen, P., Hooi Ho, S., Goh, K.-L., 2019. Bleeding detection in wireless capsule endoscopy videos-color versus texture features. J. Appl. Clin. Med. Phys. 20, 141–154.

Polikar, R., 2006. Ensemble based systems in decision making. IEEE Circuits Syst. Mag. 6, 21–45.

Prasath, V., 2017. Polyp detection and segmentation from video capsule endoscopy: a review. J. Imaging 3 (1), 1.

Qu, H.-N., Li, G., Xu, W.-S., 2010. An asymmetric classifier based on partial least squares. Pattern Recognit. 43, 3448–3457.

Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2009. RUSBoost: a hybrid approach to alleviating class imbalance. IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. 40 (1), 185–197.

Sekuboyina, A.K., Devarakonda, S.T., Seelamantula, C.S., 2017. A convolutional neural network approach for abnormality detection in wireless capsule endoscopy. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, pp. 1057–1060.

Sharif, M., Attique Khan, M., Rashid, M., Yasmin, M., Afza, F., Tanik, U.J., 2019. Deep cnn and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. J. Exp. Theor. Artif. Intell., 1–23.

Shin, Y., Qadir, H.A., Aabakken, L., Bergsland, J., Balasingham, I., 2018. Automatic colon polyp detection using region based deep cnn and post learning approaches. IEEE Access 6, 40950–40962.

Sindhu, C.P., Valsan, V., 2017a. A novel method for automatic detection of inflammatory bowel diseases in wireless capsule endoscopy images. 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), 1–6.

Sindhu, C.P., Valsan, V., 2017b. Automatic detection of colonic polyps and tumor in wireless capsule endoscopy images using hybrid patch extraction and supervised classification. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 1–5.

Sivic, J., Zisserman, A., 2003. Video Google: a text retrieval approach to object matching in videos. IEEE, pp. 1470.

Sornapudi, S., Meng, F., Yi, S., 2019. Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps. Appl. Sci. 9 (12), 2404.

Souaidi, M., Charfi, S., Abdelouahad, A.A., El Ansari, M., 2018. New features for wireless capsule endoscopy polyp detection. In: 2018 International Conference on Intelligent Systems and Computer Vision (ISCV). IEEE, pp. 1–6.

Souaidi, M., Abdelouahed, A.A., El Ansari, M., 2019. Multi-scale completed local binary patterns for ulcer detection in wireless capsule endoscopy images. Multimed. Tools Appl. 78 (10), 13091–13108.

Teague, M.R., 1980. Image Analysis via the General Theory of Moments.

Ting, K.M., 2000. A comparative study of cost-sensitive boosting algorithms. ICML.

Tomasi, C., Detection, T.K., 1991. Tracking of Point Features. Tech. Rep. CMU-CS-91-132. Carnegie Mellon University.

Torralba, A., Murphy, K.P., Freeman, W.T., 2004. Sharing visual features for multiclass and multiview object detection. IEEE Trans. Pattern Anal. Mach. Intell. 29, 854–869.

Tsuboi, A., Oka, S., Aoyama, K., Saito, H., Aoki, T., Yamada, A., Matsuda, T., Fujishiro, M., Ishihara, S., Nakahori, M., et al., 2020. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. Digest. Endosc. 32, 382–390.

Usman, M.A., Satrya, G.B., Usman, M.R., Shin, S.Y., 2016. Detection of small colon bleeding in wireless capsule endoscopy videos. Comput. Med. Imaging Graph. 54, 16–26.

Usman, M.A., Usman, M.R., Shin, S.Y., 2017. Quality assessment for wireless capsule endoscopy videos compressed via HEVC: from diagnostic quality to visual perception. Comput. Biol. Med. 91, 112–134.

Vieira, P.M., Ramos, J., Lima, C.S., 2015. Automatic detection of small bowel tumors in endoscopic capsule images by ROI selection based on discarded lightness information. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 3025–3028.

Vieira, P.M., Freitas, N.R., Valente, J., Vaz, I.F., Rolanda, C., Lima, C.S., 2020. Automatic detection of small bowel tumors in wireless capsule endoscopy images using ensemble learning. Med. Phys. 47, 52–63.

Viola, P.A., Jones, M., 2001. Robust Real-Time Object Detection.

Wireless Capsule Endoscopy, 2020. Given-Imaging is now Medtronic, Available at http://medtronicsolutions.medtronic.com/given-imaging.

Wu, Z., Huang, N.E., Chen, X., 2009. The multi-dimensional ensemble empirical mode decomposition method. Adv. Adapt. Data Anal. 1 (03), 339–372.

Xing, X., Yuan, Y., Jia, X., et al., 2019. A saliency-aware hybrid dense network for bleeding detection in wireless capsule endoscopy images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 104–107.

Yanagawa, Y., Echigo, T., Vu, H., Okazaki, H., Fujiwara, Y., Arakawa, T., Yagi, Y., 2017. Abnormality tracking during video capsule endoscopy using an affine triangular constraint based on surrounding features. IPSJ Trans. Comput. Vis. Appl. 9 (1), 3.

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

Yuan, Y., Meng, M.Q.-H., 2014. A novel feature for polyp detection in wireless capsule endoscopy images. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 5010–5015.

Yuan, Y., Meng, M.Q.-H., 2017. Deep learning for polyp recognition in wireless capsule endoscopy images. Med. Phys. 44 (4), 1379–1389.

Yuan, Y., Li, B., Meng, M.Q.-H., 2015a. Bleeding frame and region detection in the wireless capsule endoscopy video. IEEE J. Biomed. Health Inform. 20 (2), 624–630.

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

Yuan, Y., Li, B., Meng, M.Q.-H., 2016. Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images. IEEE Trans. Autom. Sci. Eng. 13, 529–535.

Yuan, Y., Li, B., Meng, M.Q.-H., 2017. WCE abnormality detection based on saliency and adaptive locality-constrained linear coding. IEEE Trans. Autom. Sci. Eng. 14, 149–159.

Zhang, D., Lu, G., 2003. Evaluation of MPEG-7 shape descriptors against other shape descriptors. Multimed. Syst. 9, 15–30.

Zhang, L., Zhang, L., Guo, Z., Zhang, D., 2010. Monogenic-LBP: a new approach for rotation invariant texture classification. 2010 IEEE International Conference on Image Processing, 2677–2680.

Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L., 2017. Beyond a Gaussian denoiser: residual learning of deep cnn for image denoising. IEEE Trans. Image Process. 26 (7), 3142–3155.

Tariq Rahim is a Ph.D. student at Wireless and Emerging Network System Laboratory (WENS Lab.) in Department of IT Convergence Engineering, Kumoh National Institute of Technology, Republic of Korea. He has completed Master in Information and Communication Engineering from Beijing Institute of Technology, PR China in 2017. His research interests include image and video processing and quality of experience for high resolution videos.

Muhammad Arslan Usman received B.S. degree in Electrical Engineering (Telecommunications) from COMSATS University, Lahore, Pakistan, in 2010 and M.S. Degree in Electrical Engineering (Signal Processing) from Blekinge Tekniska Högskola (BTH), Karlskrona, Sweden, in 2013. I completed my PhD degree in the field of IT Convergence Engineering from Kumoh National Institute of Technology, South Korea, in January 2018, and later I continued in the same research group as a postdoctoral research fellow for a period of 4 months. My research interests include quality of experience (QoE) estimation, modeling and management for medical and general videos, object classification and detection in computer vision applications, and next generation wireless networks including non-orthogonal multiple access. Currently I am working as a postdoctoral research fellow with Wireless Multimedia and Networking (WMN) research group, Kingston University, United Kingdom.

Soo Young Shin received his BS, MS, and PhD degrees in Electrical Engineering and Computer Science from Seoul National University, Korea in 1999, 2001, and 2006, respectively. He was a visiting scholar in FUN Lab at University of Washington, US, from July 2006 to June 2007. After 3 years working in WiMAX Design Lab. of Samsung Electronics, he is now associate professor in School of Electronics in Kumoh National Institute of Technology since September 2010. His research interests include wireless LAN, WPAN, WBAN, wireless mesh network, sensor networks, coexistence among wireless networks, industrial and military network, cognitive radio networks, and next generation mobile wireless broadband networks.