

## Detection of abnormality in wireless capsule endoscopy images using fractal features

Samir Jain<sup>a</sup>, Ayan Seal<sup>a,b,\*</sup>, Aparajita Ojha<sup>a</sup>, Ondrej Krejcar<sup>b,c</sup>, Jan Bureš<sup>d</sup>, Ilja Tachecí<sup>d</sup>, Anis Yazidi<sup>e</sup>

<sup>a</sup> PDPM Indian Institute of Information Technology Design and Manufacturing, Jabalpur, 482005, India

<sup>b</sup> Center for Basic and Applied Research, Faculty of Informatics and Management, University of Hradec Kralove, Hradecka, 1249, Hradec Kralove, 50003, Czech Republic

<sup>c</sup> Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100, Kuala Lumpur, Malaysia

<sup>d</sup> Second Department of Internal Medicine-Gastroenterology, Charles University, Faculty of Medicine in Hradec Kralove, University Hospital Hradec Kralove, Sokolska 581, Hradec Kralove, 50005, Czech Republic

<sup>e</sup> Artificial Intelligence Lab, Oslo Metropolitan University, 460167, Norway

### ARTICLE INFO

#### Keywords:

Wireless capsule endoscopy  
Anomaly detection  
Fractal dimensions  
Differential box-counting

### ABSTRACT

One of the most recent non-invasive technologies to examine the gastrointestinal tract is wireless capsule endoscopy (WCE). As there are thousands of endoscopic images in an 8–15 h long video, an evaluator has to pay constant attention for a relatively long time (60–120 min). Therefore the possibility of the presence of pathological findings in a few images (displayed for evaluation for a few seconds only) brings a significant risk of missing the pathology with all negative consequences for the patient. Hence, manually reviewing a video to identify abnormal images is not only a tedious and time consuming task that overwhelms human attention but also is error prone. In this paper, a method is proposed for the automatic detection of abnormal WCE images. The differential box counting method is used for the extraction of fractal dimension (FD) of WCE images and the random forest based ensemble classifier is used for the identification of abnormal frames. The FD is a well-known technique for extraction of features related to texture, smoothness, and roughness. In this paper, FDs are extracted from pixel-blocks of WCE images and are fed to the classifier for identification of images with abnormalities. To determine a suitable pixel block size for FD feature extraction, various sizes of blocks are considered and are fed into six frequently used classifiers separately, and the block size of  $7 \times 7$  giving the best performance is empirically determined. Further, the selection of the random forest ensemble classifier is also done using the same empirical study. Performance of the proposed method is evaluated on two datasets containing WCE frames. Results demonstrate that the proposed method outperforms some of the state-of-the-art methods with AUC of 85% and 99% on Dataset-I and Dataset-II respectively.

### 1. Introduction

Nowadays, wireless capsule endoscopy (WCE) based gastrointestinal endoscopy (GIE) is preferred by medical practitioners or endoscopists to examine some parts of the gastrointestinal tract due to non-invasive nature compared to conventional endoscopy procedures such as gastroscopy and colonoscopy with associated discomfort observed in patients [1–4]. WCE has become the standard diagnostic method in the field of small intestinal diseases.

Generally, a capsule of size  $10 \times 25$  mm equipped with a tiny optical camera, LED light source, signal transmitter, and the battery is

swallowed by a patient that travels through the esophagus, stomach, small intestine, and large bowel. However, the size of a capsule may vary from manufacturer to manufacturer. The tiny camera placed inside the capsule captures data in the form of a video which is sent to a sensing device tied over the waist of the patient. Generally, the camera records videos at a rate of 2–6 frames per second, in a typical WCE video that can be of 8–15 h long. A WCE video may consist of more than 50,000 frames in general [5] which makes it a tedious task to carefully watch the whole video and search for abnormalities. Computer-aided diagnosis (CAD) tools can assist medical practitioners to identify abnormal image(s) by exploiting computer vision algorithms with high precision. Due to

\* Corresponding author. PDPM Indian Institute of Information Technology Design and Manufacturing, Jabalpur, 482005, India.

E-mail address: [ayan@iiitdmj.ac.in](mailto:ayan@iiitdmj.ac.in) (A. Seal).

<https://doi.org/10.1016/j.combiomed.2020.104094>

Received 5 July 2020; Received in revised form 23 October 2020; Accepted 23 October 2020

Available online 27 October 2020

0010-4825/© 2020 Elsevier Ltd. All rights reserved.

significant growth of artificial intelligence based vision techniques, their applications in bio-medical imaging, and especially in WCE video processing have also been explored by many researchers (see for e.g. Ref. [6–13]). In WCE images, complexity and diversity of texture and color information in different types of pathologies pose challenges in designing a fully automatic CAD diagnostic system with high reliability. Nonetheless, semi-automatic CAD systems help significantly reduce the time to analyze a video by a professional evaluator in-order to eliminate normal frames. The task of an evaluator is then primarily the second reading of findings marked as abnormal (pathological) by the CAD system and elimination of false-positive frames. Various approaches have been presented over the years for specific types of abnormality detection such as bleeding [9,14], ulcers [12,15], and polyps [10,16]. These methods range from image processing, to machine learning and more recent methods based on deep learning architectures [7,8,10–12,17–20]. Some recent approaches focus on covering a broad range of abnormalities instead of only specific type of abnormalities, which are more useful for devising automatic CAD diagnostic tools for investigation of WCE frames [10,20]. In this work, a CAD tool is designed and developed to identify abnormal image(s) in WCE videos that cover a broad range of abnormalities. A random forest (RF) ensemble-based classifier is trained to identify abnormal images using the fractal dimension (FD) of the image pixel blocks. Main contributions of the present work are as follows:

- Differential Box Counting (DBC) based FD method is employed to quantify the textures of 24-bit color RGB images of WCE videos, as the anomalies exhibit different textural properties. This improves the efficiency of the CAD tool in detecting anomalies.
- In addition to RGB channels, the *L* component of the CIE-Lab color space model of images is utilized for enriching the textural information as it closely matches the human perception of lightness [21].
- To compensate for class imbalance in the WCE datasets which are generally highly imbalanced and small, synthetic minority over-sampling technique (SMOTE) [22] is applied on the feature set.
- Selection of the RF ensemble classifier is made through an empirical study on six most frequently used classifiers in the literature, namely k-nearest neighbor (KNN), support vector machine (SVM), extreme learning machine (ELM), naive bayes (NB), decision tree (DT), and RF. The CAD tool using the DBC based FD and RF classifier outperforms some of the state-of-the-art methods as illustrated in section 4.4.

The rest of the work is organized into five sections. Section 2 presents a brief overview of existing methods. The proposed methodology with its background is described in Section 3. Section 4 is devoted to experimental analysis and performance evaluation of the proposed approach. In this section, a comparison with existing approaches is also presented. Section 5 presents the concluding remarks and future scope of the proposed work.

## 2. Related work

Detection of abnormalities in WCE images using computer vision techniques is not a new area of research; it started ever since the WCE systems were introduced for healthcare in 2001. Growth of WCE market led to the development of sophisticated computer vision techniques for detection of abnormal frames using image processing and machine learning techniques. Over the last two decades, numerous techniques have been presented by researchers using different machine learning algorithms and feature extraction methods [23].

It is well known that feature extraction plays a crucial role in machine learning algorithms and finding a suitable feature extraction method for a particular problem is always a challenge. Early studies on WCE image analysis were focused on detection of bleeding abnormality [6,7,24,25]. Recent trend shows a paradigm shift towards design of CAD

systems that cover a broad range of abnormalities like vascular, ulcers, polyps, instead of a single or similar type of abnormalities. A brief overview of some of the existing methods is presented in Table 1, where we give details about the problem, feature extraction methods, classifiers, and the performance of the methods on different metrics. It may be noticed that SVM, ANN, and, KNN have been the popular choices of classifiers for abnormal frame detection. Various feature extraction methods like speed-up robust feature (SURF) [15], color and word histograms [24–27], local binary pattern (LBP) [6], Log-Gabor wavelet transform [28], and, gray level co-occurrence matrix (GLCM) [12] have been used for WCE image analysis.

In the WCE images, color information plays a crucial role in identification of abnormality. For example, when bleeding images are to be identified, the Red channel in RGB space is more important over other channels. Al-Rahayfeh et al. [18] suggested range-ratio-color condition for classifying bleeding frames. They identified that bleeding regions contain pixels with higher red channel values and lower values of green and blue channels. Therefore, a range of pixel values in each channel was considered for the identification of bleeding locations with 98% accuracy. Similar to the range-ratio-color method, Kundu et al. [8] fetched region of interest (ROI) on normalized RGB planes for bleeding frame classification. Pixels were selected based on a linear separability condition between red, blue, and green planes. The histogram was then computed on extracted ROI and the bleeding frames were identified with 98.3% accuracy using KNN classifier. The images were extracted from videos provided by Medtronic [29].

Often the presence of noise in WCE images adversely affects accuracy of the system, based on individual pixel features which get distorted due to noise. To overcome this problem, Liu et al. [24] considered block-based color histogram features and used SVM classifier for the identification of abnormal images. Specificity and sensitivity were reported to be 99% in Ref. [24]. Lv et al. [25] combined color and spatial information for extracting features using a pyramid of color invariant histograms over the HSI color space. They employed the whole image and its extracted features to identify bleeding images with an accuracy of 97.9% on dataset of videos [29]. The capability of method being invariant to light intensity was made possible with the help of hue channel of HSI color space. Ghosh et al. [14] have introduced the concept of color histogram of block statistics (CHOBS) that computes statistical features over blocks of pixels for creating histograms to identify bleeding images with an accuracy of 99.2% on the publicly available WCE video dataset provided by Medtronic [29].

Sainju et al. [26], extracted statistical features like mean, standard deviation, skew, and energy from each channel of the RGB color space of pixel intensities. Bleeding images were classified by calculating the first-order histogram yielding a classification accuracy of 89%. Yuan et al. [9] contemplated the idea of a word-based color histogram by applying k-means clustering on pixels of normal and abnormal images. Clusters were created using YCbCr color space representation of images. Individual pixels were then mapped to different clusters according to their distance with a cluster. The classification accuracy of 95.8% was reported by authors in Ref. [9] with 80 clusters. As this method involved clustering, the performance varied with cluster size in a convex fashion where optimal cluster size was determined and tuned.

Histograms based on concrete color plane pixel values were common in all the aforementioned methods. Since different color models have their own advantages, WCE images have also been experimented by researchers using CMYK, HSI, YCbCr, and, CIE-Lab color space models ([9,15,17,25]). Iakovidis et al. [15], have employed image transformation to CIE-Lab color space and applied SURF on color channels to select the salient points. The minimum and the maximum values of each color component are calculated over a square neighborhood of salient pixels for extraction of features. An average value of 89.2% of the Area under Curve (AUC) was reported in Ref. [15] with a minimum AUC of 69.9% in case of angiectasias on KID Dataset 1 [30]. A customized color model close to the CMYK color model was devised by Novozamsky et al.

**Table 1**

A brief overview of state-of-the-art methods for detecting abnormality in WCE images.

Authors	Purpose	Feature Extraction Method	ML classifier	Classification Results
Karkanis et al. [7], 2001	Lesion detection	Statistical information of second order discrete wavelet transform	ANN	Accuracy: 95.50%
Liu et al. [34], 2008	Bleeding detection	Raw color pixel values	SVM	Sensitivity: 99.64%, Specificity: 99.58%
Li et al. [6], 2009	Bleeding detection	Chrominance moment and LBP is computed in HSI color space	ANN	Accuracy: 92.40%, Sensitivity: 93.20%, Specificity: 91.60%
Al-Ryalfeh et al. [18], 2010	Bleeding detection	Range of color pixel values and ratio of them	Pixel count as threshold	Accuracy: 98.00%
Lv et al. [25], 2011	Bleeding detection	Pyramid of color invariant histogram	SVM	Accuracy: 97.90%, Sensitivity: 97.80%, Specificity: 98.00%
Sainju et al. [26], 2013	Bleeding detection	Histogram	ANN	Accuracy: 89.00%
Iakovidis et al. [15], 2014	Lesion detection	Probability WCE images are converted into CIE-Lab color space and salient points are extracted using speedup robust features (SURF) method	SVM	Accuracy: 94.50%, Sensitivity: 96.00%, Specificity: 84.60%
Yuan et al. [9], 2015	Bleeding detection	Histogram on bag-of-visual-words using K-size color clusters	KNN	Accuracy: 95.75%, Sensitivity: 92.00%, Specificity: 96.50%
Novozamsky et al. [17], 2016	Bleeding detection	Threshold set on pixel values based on new color model close to CMYK color model in which red color is enhanced	Threshold on pixel count	TP:96.41%, FP:21.81%
Jia et al. [31], 2017	Bleeding detection	CNN combined with handcrafted (HC) features in which histogram is computed over the count of pixels in each centroid when the image pixels are clustered using K-means clustering algorithm	Softmax	Recall: 91.00%, Precision: 94.79%, F1-score: 92.85%
Chen et al. [28], 2017	Bleeding detection	Log-gabor filter bank, histogram and color features	SVM	Accuracy: 98.97%, Sensitivity: 94.07%, Specificity: 99.15%
Yuan et al. [16], 2017	Polyp detection	Sparse autoencoder	Softmax	Accuracy: 98.17%

**Table 1 (continued)**

Authors	Purpose	Feature Extraction Method	ML classifier	Classification Results
		with hybrid loss function employing KL divergence with an image manifold constraint.		
Kundu et al. [8], 2018	Bleeding detection	Histogram computed over normalized green planes	KNN	Accuracy: 97.86%, Sensitivity: 95.20%, Specificity: 98.32%
Iakovidis et al. [10], 2018	Anomaly detection	5-layer CNN, Activation Maps	CNN	Accuracy: 89.90%, Sensitivity: 90.70%, Specificity: 88.20%
Sadasivan et al. [13], 2019	Anomaly detection	A CNN trained on normal and abnormal patches of size $64 \times 64$ obtained from WCE images of size $320 \times 320$	CNN	AUC: 95.36%
Xiao et al. [11], 2020	Lesion detection	CNN based object detection using YOLOv3 network.	Softmax (CNN)	Mean Average Precision (mAP): 93.50%
Khan et al. [12], 2020	Ulcer and bleeding detection	CNN with transfer learning on VGG-16 and GLCM which are fused and relevant features are selected by applying particle swarm optimization technique.	Cubic SVM	Accuracy: 98.40%, Sensitivity: 98.33%, Precision: 98.36%, F1-score: 98.34%, AUC: 100%

[17] in which, both color and spatial properties were adopted for the identification of bleeding images. They have achieved a true positive percentage of 98% in classifying bleeding images on a private dataset.

For the identification of anomalies like ulcers, polyps, villous oedema, aphthae etc., textural features are quite useful, since some kind of abnormalities can only be identified using their texture. Color features of such abnormal regions in an image are quite like those of normal regions in some of the abnormalities. In view of this many researchers have considered a combination of color and textural features for WCE image and video analysis. Meng et al. [6] have dealt with the problem of differentiating normal images from bleeding images by exploiting LBP as the textural feature extractor. These LBP features alongwith with the chrominance moments extracted from the HSI color plane were used to classify abnormal WCE images. Classification performance was evaluated with zeroth, first, and second-order chrominance moments and the authors have reported that the zero-order chrominance moment provided the highest accuracy of 92.4% on a private WCE dataset provided by Prince of Wales Hospital, Hong Kong. Karkanis et al. [7] employed the wavelet transformation of second-order statistics on each WCE image for the extraction of texture features to detect polyps where 97.2% accuracy was yielded. When the WCE capsule travels through the digestive tract, there is a possibility of lower luminance which affects the quality of the WCE image. This problem was addressed by Chen et al. in Ref. [28], where the Log-Gabor wavelet transform was applied to rectify the problem of non-uniform luminance. In addition to the Log-Gabor

filter, gray-level histogram and color features were combined for the detection of bleeding as an abnormality which reached to classification accuracy of 98.97% where the WCE dataset was provided by Ankon Incorporation, Wuhan, China. Devising algorithms to extract features need thorough knowledge about the behavior of abnormalities which sometimes becomes difficult if some of the observations are missed. It can be reduced if an algorithm is capable of extracting the features automatically.

In the last decade deep learning has become very popular due to its outstanding performance in some of the most challenging computer vision tasks. Deep learning is a kind of representation learning in which data features are extracted automatically. Due to huge success of deep learning in other domains, it has attracted the attention of people working on WCE image and video analysis too. Convolutional neural networks (CNNs), which are the basic building blocks of deep learning algorithms, have been proposed by many researchers for accurate classification of WCE images (see, e.g. Refs. [19,20,31], and references therein). However, the current deep learning algorithms face challenge if the datasets are small and imbalance, which is the case with WCE image datasets. As a result, CNNs often get trapped in overfitting, and do not generalize well [32,33]. Nonetheless, deep learning architectures are being explored in the WCE image and video analytics and some promising results have been reported recently [10,11,13].

The related work discussed above manifest the importance of color and texture features for WCE image analysis in the block based approach. The present work, focuses on efficient feature extraction using the concept of DBC based FD on blocks of WCE images, computed over the individual color channel, exploiting both color and textural (FD) information. The problem of highly imbalanced small dataset is also addressed by applying the well known SMOTE method for data augmentation. It is worthwhile to mention that the proposed method covers a broad range of abnormalities in WCE images and can classify abnormal images with high accuracy and, AUC.

In the next section, the proposed method is discussed in detail.

### 3. Methodology

In this section, a method is proposed for abnormality detection in WCE images using FD and RF algorithm. The choice of machine learning classification algorithm is based on an empirical study with six different ML algorithms on two different datasets, one of which is a publicly available dataset, namely KID dataset [30] and the other being private dataset [17]. Each algorithm uses the FD of image pixel blocks and the best performing method is chosen for abnormality detection in WCE images. To begin our discussion on the methodology of the proposed work, we first give a brief introduction to the concept of FD in the context of images.

#### 3.1. Fractal dimension

FD has been an active research topic over the past decades owing to its noticeable applications such as texture segmentation [35], bio-metrics [36,37], cancer diagnosis [38], pattern recognition [39], MRI image analysis [40], image fusion [21,41] and others [42]. FD measures the self-similar contents of an image [43]. One may notice that abnormalities in WCE images are complex objects with varying shapes, sizes, color, and textures. Moreover, anomalies are highly irregular natural shapes which can be correlated with fractal features (see Fig. 4). There exist many methods in the literature to measure the FD of images. The DBC method is one of the most frequently deployed methods due to its simplicity and computational efficiency [44].

The DBC method was introduced by Sarkar et al. [44] to estimate the FD of a gray-scale image. According to the DBC method, FD  $F_d(K)$  of a fractal set  $K$  is defined in an  $n$ -dimensional Euclidean space by Eq. (1).

$$F_d(K) = \frac{\log(B_s)}{\log(1/s)}, \quad (1)$$

such that  $s \neq 1$  and  $B_s$  is the number of boxes at scale  $s$  required to cover the fractal set  $K$ . An image  $I$  of size  $N \times N$  is mapped into three-dimensional space as shown in Fig. 1, where  $x$  – coordinate and  $y$  – coordinate are used to denote the dimensions of  $I$  while  $z$  – coordinate represents the gray-level intensity of  $I$ . The image plane is divided into non-overlapping grids of size  $g \times g$  pixels. Here,  $g$  is an integer and it varies from 2 to  $N/2$ . If  $g$  is not divisible by  $N$  then zeros are padded in the boundary of  $I$ . A grid  $s$  will have a scale of value  $g/N$ . Let  $b_s$  be the number of boxes of size  $g \times g \times h$  each representing gray-level variations of a particular grid, where  $h$  is the height which can be calculated using Eq. (2), where  $Q$  is the total number of gray-levels. Let  $q_{max}$  and  $q_{min}$  be the maximum and the minimum intensity values of the  $(i,j)^{th}$  grid respectively. The number of boxes  $b_s(i,j)$  required to cover the  $(i,j)^{th}$  grid can be computed using Eq. (3).

$$b_s(i,j) = \left\lceil \frac{q_{max}}{h} \right\rceil - \left\lceil \frac{q_{min}}{h} \right\rceil + 1, \quad (3)$$

where  $\left\lceil \frac{q_{max}}{h} \right\rceil$  and  $\left\lceil \frac{q_{min}}{h} \right\rceil$  denote the box numbers containing the maximum and minimum gray-levels on the  $(i,j)^{th}$  grid respectively. Therefore, the total count of boxes termed as  $B_s$  at scale  $s$  can be calculated using Eq. (4).

$$B_s(i,j) = \sum_{i,j} b_s(i,j), \quad (4)$$

where  $B_s$  represents the surface roughness in image  $I$ . Now, FD of  $I$  is the slope of the line obtained by fitting the points  $\left( \log\left(\frac{1}{s}\right), \log(B_s) \right)$  vs using linear least squares regression (LLS). Since the value of FD is a scalar quantity, it may not be sufficient to use as a feature for a classification problem. In the present paper, an image is divided into square pixel

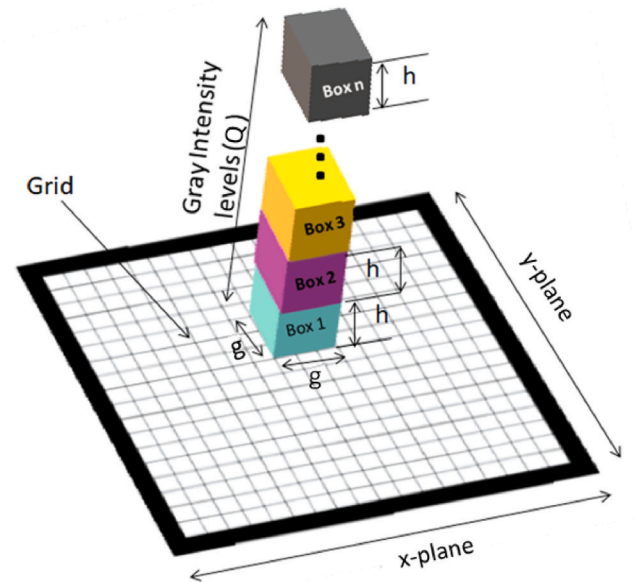


Fig. 1. A sketch showing corresponding a grid with computable parameters to find the number of boxes using DBC method.

$$h = \frac{gQ}{N}, \quad (2)$$



blocks of equal size and instead of considering the value of FD as a feature, the number of boxes  $b_s$  of each block  $s$  in an image  $I$  are used as features for the identification of abnormal images. Several methods [35, 43, 45–51] exist in literature to estimate FD of an image and each has its own merits and limitations. A brief overview of some of these methods is reported in Table 2. For more details on DBC based FD estimation methods, interested readers are referred to Ref. [42].

### 3.2. Proposed method

A schematic block diagram of the proposed CAD system is shown in Fig. 2. As shown in the figure, the proposed method consists of two steps: feature extraction and classification. Class balancing is performed on the dataset containing extracted features after the first step. In the first step,  $R$ ,  $G$ ,  $B$ , and  $L$  channels are extracted from WCE images. The  $L$  channel is obtained by converting the RGB color image into its CIE-Lab color space representation [52]. Then a DBC method is applied on the blocks extracted from each channel to extricate features. To select a suitable DBC based FD estimation method for WCE images, eight different DBC based FD methods listed in Table 2 are initially chosen and their performances are compared on two WCE image datasets used in the present work. The best performing DBC method given in Ref. [43] is selected for FD feature computation in the present study (see Tables 3–6). In color theory, the lightness component also known as a tone represents the brightness of color in an image. It is used as one of the color appearance parameters. In relation to GI anomalies, the lightness component plays an important role since some anomalies are very close to normal regions in the case when color is used as a discriminator. For example, light red color in the intestinal walls may represent normal regions whereas dark red gives a clue of a bleeding spot. Since RGB models do not have an explicit lightness component where color integrants are dissociated from luminance components and they struggle to give information related to luminosity [21]. The  $L$  channel in CIE-Lab color space is employed to extract the lightness component (luminosity) by converting the image from RGB to CIE-Lab. It may also be mentioned that the  $L$  channel is also close to human vision. Since  $a$  and  $b$  channels of CIE-Lab color space also describe the color, they are ignored.

In the second step, the set of extracted block features is split into train and test subsets. Similar to the selection of the most suited DBC method for feature extraction, a classifier is selected by evaluating the

performance of six most frequently used classification methods on the WCE dataset. The best performing classifier is observed to be RF classifier (see Tables 3–6). The proposed ML model is trained, tested, and finally deployed for WCE image analysis.

We now proceed to discuss the proposed feature extraction algorithm in detail. Algorithm 1 takes as an input, an RGB image  $I_{N \times N}$  and produces a feature vector  $F_{1 \times v}$ , where  $v$  is the total number of blocks obtained from the  $R$ ,  $G$ ,  $B$  and  $L$  channels of the image. Each channel block is of size  $g \times g$ . The DBC or number of boxes  $b_s$  in each block is computed by Eq. (3). Finally, all the features obtained from each channel are concatenated to form a feature vector  $F$  ( $F_{1 \times v}$ ), which is fed into ML algorithms.

**Algorithm 1.** Extraction of fractal features (EFF) algorithm.

### 3.3. Synthetic minority oversampling technique

The WCE datasets considered in this work are highly imbalanced and biased towards normal images. Therefore, if we apply ML algorithms on an imbalanced dataset, classification results will be inclined towards the majority class. So, in this case, the classifier will classify abnormal images as normal and will falsely result in overall high accuracy. This problem can be solved by either oversampling minority class or by under-sampling majority class in-order to balance the number of samples in each class. SMOTE [22] uses an oversampling technique by generating auxiliary samples of minority class from real instances through a linear combination of two similar samples using the nearest neighbor and structural similarity index. So, we can prevent over-fitting and generalize the results. In this work, the SMOTE considers SVM with linear kernel and 5 nearest neighbours while augmenting data of the minority class.

### 3.4. Machine learning algorithms

Generally, it is a tedious task to select an ML algorithm, which best suits the data. In this study, six ML algorithms namely KNN [53], SVM [54], ELM [55,56], NB [57], DT [58], RF [59–61] are adopted and their performances are evaluated on WCE image datasets. Each classifier is tuned in such a way that it can give its best performance. The KNN algorithm depends on a user defined number  $k$  and finding a suitable value of  $k$  is a difficult task if it is not known a-priori. In this study, the value of  $k$  is determined experimentally. SVM with linear and cubic polynomial

**Table 2**  
DBC methods and corresponding parameters.

Methods	Block size	Box height	Box-count in $(i,j)^{th}$ grid
DBC [44]	$\frac{N}{2}$	$\frac{gQ}{N}$	$\left\lceil \frac{q_{max}}{h} \right\rceil - \left\lceil \frac{q_{min}}{h} \right\rceil + 1$
RDBC [45]	$\left\lceil \frac{N}{g} \right\rceil + 1 \leq \left\lceil \frac{N}{g-1} \right\rceil$	$\frac{gQ}{N}$	$\left\lceil \frac{q_{max} - q_{min}}{h} \right\rceil$
SDBC [46]	$\frac{N}{2}$	S	$\left\lceil \frac{q_{max} - q_{min} + 1}{h} \right\rceil$
DBC, Li [47]	$\frac{N}{2}$	$\frac{g-1}{0.5(I_{max} - I_{min})}$	$\begin{cases} \left\lceil \frac{q_{max} - q_{min}}{h} \right\rceil & \text{if } q_{max} \neq q_{min} \\ 1 & \text{if } q_{max} = q_{min} \end{cases}$
DBC, Li [48]	$\frac{g-1}{2}$	$\frac{gQ}{N}$	$\left\lceil \frac{q_{max} - q_{min}}{h} \right\rceil$
DBC, Liu [49]	$\frac{N}{2}$	$\frac{gQ}{N}$	$\begin{cases} \left\lceil \frac{q_{max} - q_{min} + 1}{h} \right\rceil & \text{if } q_{max} \neq q_{min} \\ 1 & \text{if } q_{max} = q_{min} \end{cases}$
DBC, Lai [50]	$\frac{N}{2}$	$\frac{g(I_{max} - I_{min} + 1)}{N}$	$\begin{cases} \left\lceil \frac{q_{max} - q_{min} + 1}{h} \right\rceil & \text{if } q_{max} \neq q_{min} \\ 1 & \text{if } q_{max} = q_{min} \neq 0 \\ 0 & \text{if } q_{max} = q_{min} = 0 \end{cases}$
DBC, Panigrahy [43]	$\frac{N}{2}$	$\frac{2g}{N}$	$\left\lceil \frac{qt_{max} - qt_{min} + 1}{h} \right\rceil \frac{i_p}{g(g+1)/2}$ , where, $i_p$ is number of pixels grid

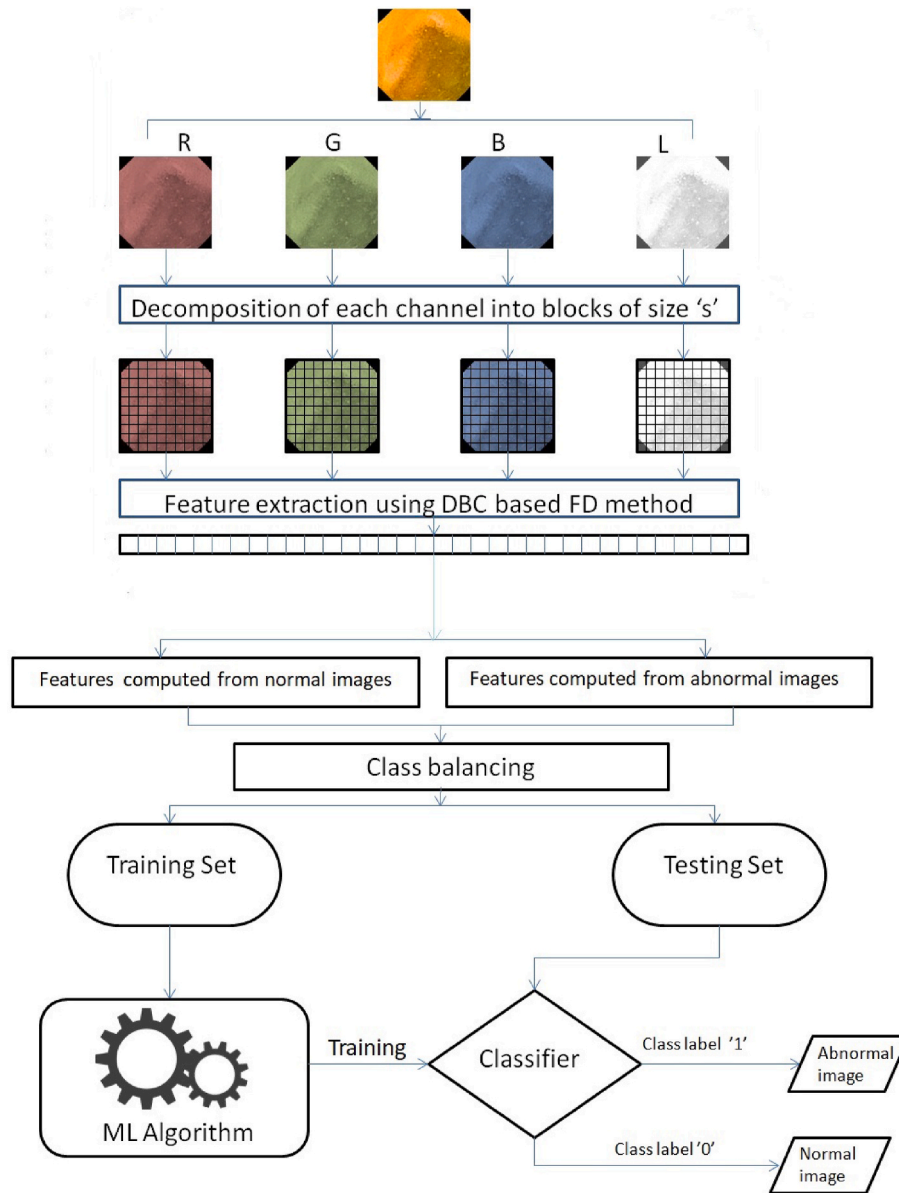


Fig. 2. A Schematic block diagram representing the proposed methodology for WCE image classification.

Table 3

Accuracy scored in different DBC methods using various classifiers for abnormal WCE image detection.

Methods	Dataset-I						Dataset-II					
	ELM	KNN	NB	DT	SVM	RF	ELM	KNN	NB	DT	SVM	RF
DBC [44]	0.52	0.58	0.52	0.76	0.80	0.83	0.47	0.68	0.70	0.89	0.98	0.98
RDBC [45]	0.57	0.60	0.54	0.76	0.80	0.84	0.48	0.69	0.70	0.90	0.98	0.98
SDBC [46]	0.58	0.60	0.68	0.71	0.80	0.83	0.50	0.71	0.70	0.90	0.98	0.97
DBC, Li [47]	0.58	0.59	0.63	0.71	0.79	0.84	0.49	0.68	0.73	0.91	0.97	0.99
DBC, Li [48]	0.55	0.48	0.64	0.72	0.79	0.83	0.50	0.69	0.73	0.90	0.98	0.98
DBC, Liu [49]	0.57	0.59	0.64	0.70	0.79	0.83	0.47	0.69	0.72	0.95	0.98	0.99
DBC, Lai [50]	0.54	0.61	0.64	0.71	0.76	0.81	0.49	0.68	0.70	0.91	0.98	0.98
DBC, Panigrahy [43]	0.60	0.58	0.58	0.75	0.80	0.85	0.50	0.70	0.70	0.91	0.98	0.99

kernels were initially evaluated and the cubic SVM is selected due to its better performance. The regularization parameter  $c$ , in the SVM is set to 1 with a squared  $L2$  penalty term in order to avoid over-fitting. Generally, a lower  $c$  value makes the decision surface smoother while a higher  $c$  value helps in better classification. The ELM is implemented with 10 neurons in its hidden layer and  $\tanh$  activation function with a logistic

regressor exploited for classification. NB classifiers differ mainly by the assumptions that they make regarding the distribution of features. In this study, the Gaussian NB classifier with normal distribution is used after transforming the features into a normal distribution. In this work, for implementing the DT, Gini [62] index is adopted as a measure of impurity for classification. The Gini impurity gives nearly the same

**Table 4**

Precision scored in different DBC methods through various classifiers in abnormal WCE image detection.

Methods	Dataset-I						Dataset-II					
	ELM	KNN	NB	DT	SVM	RF	ELM	KNN	NB	DT	SVM	RF
DBC [44]	0.51	0.70	0.68	0.77	0.82	0.83	0.48	0.80	0.72	0.89	0.98	0.98
RDBC [45]	0.58	0.73	0.63	0.76	0.82	0.85	0.48	0.80	0.73	0.91	0.98	0.98
SDBC [46]	0.60	0.72	0.68	0.70	0.82	0.83	0.50	0.80	0.71	0.90	0.98	0.97
DBC, Li [47]	0.58	0.71	0.63	0.72	0.81	0.83	0.49	0.80	0.75	0.90	0.98	0.99
DBC, Li [48]	0.54	0.69	0.64	0.73	0.81	0.84	0.50	0.80	0.75	0.90	0.98	0.98
DBC, Liu [49]	0.57	0.72	0.66	0.71	0.81	0.84	0.47	0.80	0.74	0.91	0.98	0.99
DBC, Lai [50]	0.54	0.72	0.66	0.71	0.78	0.82	0.49	0.80	0.72	0.91	0.98	0.98
DBC, Panigrahy [43]	0.61	0.71	0.63	0.75	0.82	0.86	0.66	0.81	0.75	0.91	0.98	0.99

**Table 5**

Recall scored in different DBC methods with various classifiers in abnormal WCE image detection.

Methods	Dataset-I						Dataset-II					
	ELM	KNN	NB	DT	SVM	RF	ELM	KNN	NB	DT	SVM	RF
DBC [44]	0.51	0.58	0.52	0.76	0.79	0.83	0.48	0.68	0.70	0.89	0.98	0.98
RDBC [45]	0.56	0.61	0.54	0.76	0.80	0.84	0.48	0.69	0.70	0.90	0.98	0.98
SDBC [46]	0.58	0.60	0.67	0.71	0.79	0.83	0.50	0.70	0.70	0.90	0.98	0.97
DBC, Li [47]	0.58	0.58	0.63	0.71	0.79	0.84	0.50	0.68	0.73	0.90	0.97	0.99
DBC, Li [48]	0.57	0.58	0.64	0.74	0.81	0.82	0.51	0.69	0.73	0.90	0.98	0.98
DBC, Liu [49]	0.56	0.59	0.64	0.71	0.80	0.83	0.47	0.68	0.72	0.91	0.98	0.99
DBC, Lai [50]	0.54	0.61	0.64	0.71	0.76	0.81	0.49	0.68	0.71	0.91	0.98	0.98
DBC, Panigrahy [43]	0.60	0.58	0.57	0.75	0.80	0.85	0.49	0.70	0.70	0.91	0.98	0.99

**Table 6**

F1-score obtained from different DBC methods on various classifiers in abnormal WCE image detection.

Methods	Dataset-I						Dataset-II					
	ELM	KNN	NB	DT	SVM	RF	ELM	KNN	NB	DT	SVM	RF
DBC [44]	0.51	0.63	0.42	0.77	0.80	0.83	0.48	0.68	0.69	0.89	0.98	0.98
RDBC [45]	0.56	0.66	0.44	0.76	0.81	0.84	0.48	0.69	0.69	0.90	0.98	0.98
SDBC [46]	0.56	0.53	0.68	0.70	0.80	0.83	0.50	0.66	0.69	0.90	0.98	0.97
DBC, Li [47]	0.58	0.52	0.63	0.71	0.79	0.83	0.50	0.65	0.72	0.90	0.98	0.99
DBC, Li [48]	0.55	0.53	0.65	0.73	0.81	0.83	0.50	0.65	0.72	0.90	0.98	0.98
DBC, Liu [49]	0.56	0.52	0.64	0.71	0.80	0.83	0.46	0.65	0.71	0.91	0.98	0.99
DBC, Lai [50]	0.53	0.54	0.63	0.71	0.76	0.81	0.48	0.64	0.70	0.91	0.98	0.98
DBC, Panigrahy [43]	0.60	0.50	0.54	0.75	0.80	0.85	0.48	0.66	0.69	0.91	0.98	0.99

results as entropy gives, however, Gini is faster than entropy. The maximum depth of tree is taken as 50 which is determined experimentally that best suits the given dataset. Since the RF uses DTs, the Gini index is again employed for all its material advantages. The maximum depth of each DT is not fixed and maximum number estimators (DT) are taken as 500, which are decided by trial and error method.

## 4. Experimental results and discussion

### 4.1. Environment setting

The proposed work is implemented in Python language using the Anaconda environment installed on Windows 10. The system is coupled with 8 GB of RAM and Intel Core i7 4th generation CPU.

### 4.2. Dataset description

In this study, one publicly available benchmark dataset [30] and one private dataset [17] are adopted to evaluate the proposed method. Both the datasets are referred to as Dataset-I and Dataset-II. A brief description of each of these datasets is as follows:

#### 4.2.1. Dataset-I

Dataset-I is also known as KID and it consists of images and videos. In the present work, only images are considered. In this dataset, all the

images are of size  $360 \times 360$  pixels, acquired by MicroCam sensor [10] in RGB color space. The Dataset-I is further divided into KID Dataset 1 [15] and KID Dataset 2 [10]. On the one hand, KID Dataset 1 contains 77 images covering abnormalities like angioectasias, aphthae, chylous cysts, polypoid lesions, villous oedema, bleeding, lymphangiectasias, ulcers, and stenoses. While, KID Dataset 2 [10], consists of 2371 images of miscellaneous findings of polypoid, vascular and, inflammatory lesions, and normal images of esophagus, stomach, small bowel, and colon. Therefore, a total of 670 abnormal and 1778 normal images are considered in this study. After removing unwanted black borders the size of images reduces to  $320 \times 320$  pixels.

#### 4.2.2. Dataset-II

Dataset-II consists of 37 WCE videos in which 21 videos are of patients with abnormal findings whereas 16 videos are related to the subjects with normal findings evaluated by an experienced physician. Initially, frames are separated from all the videos to create a WCE image dataset. Out of these frames, 309 images are selected and labeled as abnormal depending upon the ground truths available with every video file whereas 2063 images are labeled as normal. Each image is an RGB color image of size  $288 \times 288$  pixels. The WCE apparatus employed while creating this dataset is EndoCapsule developed by Olympus which produces frames at the rate of 2fps [17]. Anomalies such as lipoma (9), xanthoma (3), erosion (26), ulcer (149), aphthous lesions (23), erythema (61), villous (29), and bleeding (9) are covered in the dataset

---

```

1: procedure EFF( $I_{N \times N}$ )           ▷ Computing fractal features on image  $I$ .
2:    $X \leftarrow 255$                  ▷ The  $X$  is total number of gray levels.
3:    $g \leftarrow [2, N/2]$            ▷ The value of  $g$  depends on the DBC method used for
                                     feature extraction mentioned in Table 2.
4:    $h \leftarrow gR/N$                ▷ The  $h$  is the height of a box computed using Eq. 2.
5:   Let  $F$  be an empty feature vector of size  $1 \times v$  and  $C$  be the channels of
                                     image  $I$ .
6:   while  $c \in C$  do
7:     Divide  $I$  into grids of size  $g \times g$ 
8:     while  $j \in \text{grids}$  do
9:       Compute  $b_s$  for each  $j$  by Eq. 3
10:      Concatenate the value of  $b_s$  to  $F$ 
11:    end while
12:  end while
13:  return  $F$ 
14: end procedure

```

---

where the number inside parenthesis denotes the number of images of a particular anomaly.

All the experiments are performed on both the datasets separately. Each dataset is randomly divided into train and test sets in a ratio of 80 : 20. The datasets are biased towards normal images as the count of normal images are more than abnormal images. Thus, class balancing is needed otherwise, results will get affected more by the class with normal images. Therefore features of the abnormal class are over-sampled to match with the number of features of normal images using SMOTE [22].

#### 4.3. Evaluation metrics

In this work, the classification performance of the proposed method is done based on a standard metric named accuracy [63–65]. Accuracy measures the ratio of correct predictions over all predictions. Mathematically, it is defined by Eq. (5).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (5)$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative predictions respectively. However, accuracy is not enough due to the accuracy paradox [63]. So, other metrics namely, precision, recall, and F1-score are also considered in this study, which are computed by Eqs. (6)–(8) respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

Precision quantifies the fraction of actual over predicted positive images whereas recall is the fraction of actual abnormal images over correctly classified abnormal images. Moreover, F1-score provides a

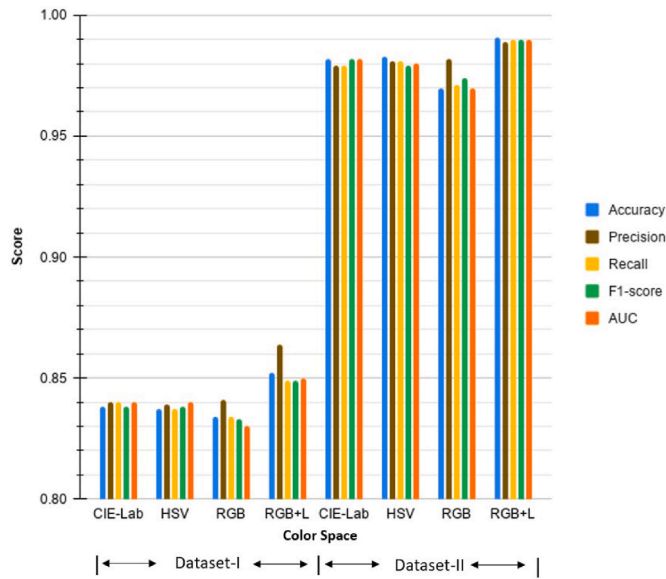
single score that balances both the concerns of precision and recall. The classification performance is also tested using area under the receiver operating characteristic curves (AUC). The receiver operating characteristic (ROC) curve depicts the relationship between specificity and sensitivity and also we can find the relative trade-offs between true positive rates (TPR) and false positive rates (FPR). AUC is scale-invariant and it does not consider the classification threshold. Since, AUC gives a probabilistic value that a random positive sample is ranked higher than a negative sample, therefore, higher AUC is always desirable.

#### 4.4. Results and comparison

In this study, a total of 7 experiments are conducted. In the first experiment, all the DBC methods mentioned in Table 2 are implemented on the datasets for extracting features, which are fed into the ML algorithms namely, KNN, SVM, ELM, NB, DT, and RF separately. The performances of state-of-the-art DBC methods are computed using four evaluation metrics namely, accuracy, precision, recall, and F1-score discussed in Section 4.3. Values of these metrics obtained by applying various DBC methods are reported in Tables 3–6 respectively. It is clear from Tables 3–6 that the improved DBC method devised by Panigrahy et al. [43] with RF classifier yields the highest accuracy of 85% and 99% when compared with other DBC methods in both the datasets. Hence, this method is preferred for the extraction of fractal features in the proposed method. Moreover, it can be deduced from the results that the RF outperforms all the other classifiers considered in the study whereas the SVM performs marginally lower than RF. In this study, RF is the only ensemble classifier that is considered. It utilizes unpruned DTs as weak classifiers. The DTs take data samples as inputs and produce the results independently. Then, the mode of outputs of each of these DTs is considered as final. Since, DTs lag in giving good results due to over-fitting raised during training, RF mitigates this problem by averaging the result produced by DT. Probably this might be the reason for better performance of RF.

In the second experiment, the proposed method paired with the RF





**Fig. 3.** Performance of the proposed method on different color spaces and suggested a combination of *RGB* with *L* on the Dataset-I and the Dataset-II. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

classifier is also tested on RGB, HSV, CIE-Lab, and the adopted hybrid color spaces and the obtained results are shown in Fig. 3. It is observed from Fig. 3 that converting the WCE image originally in RGB plane to HSV and CIE-Lab spaces are giving better results, however, the performance is even better when the *L* channel is integrated with RGB color space achieving the highest AUC of 85% and 99% on Dataset-I and Dataset-II respectively. When the HSV color space is considered, 83% and 97% AUCs are reached on the Dataset-I and Dataset-II respectively. However, the results obtained using HSV and CIE-Lab color spaces are almost identical.

As features are extracted in block-wise fashion, it is necessary to find the block size for which extracted features give the best classification results. In the third experiment, features are computed by applying the proposed technique on different block sizes varying from 3 to 13 with a step size of 2 while keeping an eye on the performance in terms of accuracy. Block size greater than 13 does not seem to be lucrative since the contribution of pixels belonging to abnormal regions of smaller size may be ignored in the larger block size. The performance of the proposed method with respect to the block size is shown in Table 7. It is clear from the table that features extracted with the block size of 7 give the best result in both the datasets, although marginally.

Table 8 shows the size of each image in pixels, the length of a feature vector of each channel in an image, the total length of a feature vector of four channels, and the number of normal and abnormal images in both the datasets. Table 9 reports the size of the feature matrices of both the datasets before and after data augmentation using SMOTE. To test the

**Table 7**

Effect of block size on accuracy score of the proposed method in Dataset-I and Dataset-II.

Block size	Dataset-I	Dataset-II
3 × 3	0.838	0.991
5 × 5	0.826	0.985
7 × 7	<b>0.850</b>	<b>0.992</b>
9 × 9	0.833	0.985
11 × 11	0.841	0.990
13 × 13	0.845	0.987

The best classification results are highlighted by the bold characters.

**Table 8**

The length of the feature vector of both the datasets before applying SMOTE.

Dataset	Image size in pixels	Length of a feature vector of each channel	Length of a feature vector of four channels	Number of normal/abnormal WCE images
Dataset-I	320 × 320 × 3	2025	8100	1776/670
Dataset-II	288 × 288 × 3	1681	6724	2063/309

**Table 9**

The size of the feature matrices of both the datasets before and after data augmentation using SMOTE.

	Dataset-I		Dataset-II	
	Feature matrix of normal images	Feature matrix of abnormal images	Feature matrix of normal images	Feature matrix of abnormal images
Before applying SMOTE	1776 × 8100	670 × 8100	1776 × 8100	1776 × 8100
After applying SMOTE	2063 × 6724	309 × 6724	2063 × 6724	2063 × 6724

**Table 10**

Comparing performance of the proposed technique with and without applying SMOTE on computed features.

Metrics	With SMOTE		Without SMOTE	
	Dataset-I	Dataset-II	Dataset-I	Dataset-II
Accuracy	0.85	0.99	0.74	0.92
Precision	0.86	0.99	0.68	0.93
Recall	0.85	0.99	0.74	0.92
F1-score	0.85	0.99	0.66	0.90
AUC	0.85	0.99	0.52	0.69

effect of SMOTE, the proposed method is evaluated with and without applying SMOTE on the extracted features. The results reported in Table 10 show that the performance of the proposed model is better in both the datasets when the SMOTE is utilized. This is the fourth experiment in the series of experiments performed in the present study.

In the fifth experiment, the proposed method is compared with six state-of-the-art methods namely, Range Ratios Color [8], Raw Histogram [24], CHOBS [14], Color histogram [26], and weak CNN [10]. The performances of these methods are reported in Table 11. 5-fold cross-validation is performed to ensure that the results are not biased with respect to the choice of samples in training and testing sets. The

**Table 11**

Classification results of the proposed method with other state-of-the-art methods.

Features	Dataset-I			Dataset-II		
	Accuracy	F1-score	AUC	Accuracy	F1-score	AUC
Range Ratios Color [8]	0.65	0.42	0.53	0.85	0.79	0.80
Raw histogram [24]	0.73	0.61	0.50	0.92	0.90	0.72
Bag of Words [9]	0.80	0.79	0.79	0.91	0.89	0.85
CHOBS [14]	0.73	0.72	0.66	0.97	0.96	0.92
Weak CNN [10]	0.82	<b>0.87</b>	<b>0.85</b>	0.97	0.96	0.97
<b>Proposed Method</b>	<b>0.85</b>	0.84	<b>0.85</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>

The best classification results are highlighted by the bold characters.

proposed method demonstrates superior performance in terms of accuracy when compared with all state-of-the-art methods. However, the weak CNN, introduced by Iakovidis et al. [10] performs better than the proposed method on Dataset-I, in terms of F1-score and AUC. The Weak CNN introduced by Iakovidis et al. [10] contains five convolution layers such that the first four layers are followed by a max-pool layer each and the last one is followed by three fully connected layers. The last layer is the classification layer that uses a softmax classifier for identification of abnormal images. Nonetheless, the proposed method is observed to be better than Weak CNN [10] on Dataset-II with respect to all the three measures and on Dataset-I with respect to accuracy. To check if the difference in the performance is significant, Student's *t*-Test is performed which is detailed as follows.

In the sixth experiment, the Student's *t*-Test [66] is performed on the accuracy scores of the proposed method and the Weak CNN introduced in Ref. [10] to find out if the difference in the accuracy scores of the two methods is significant. Accuracy is recorded on ten rounds of experiments using the test datasets and statistical values like mean, variance, and the standard deviation of accuracy scores are computed, which are listed in Table 12. After the application of the *t*-test, the *p*-value comes out to be close to 0 in Dataset-I whereas it is near to 0.005 in Dataset-II as listed in Table 13. This indicates that there is 0% and less than 0.5% percent chance of getting the same results from both the methods on Dataset-I and Dataset-II respectively.

Accurate identification of texture patterns plays an important role in detection of abnormalities in WCE images. Some of the existing textural descriptors like Haralick features [67], LBP [68] and KAZE features [69] have demonstrated their capabilities in many computer vision problems including those in biomedical imaging. In view of this, a comparison of the proposed method is made with the above three feature descriptors keeping RF as the classifier. Results of this seventh experiment are presented in Table 14. This shows that the proposed DBC based FD features performs better on both the datasets.

#### 4.5. Discussion

The results presented in the previous section demonstrate that the proposed method performs better than some of the existing methods on both the datasets. Fig. 4 shows the visualization of features representing abnormalities in images using the extracted features with the blocks size of  $7 \times 7$ . Here the example image A from the Dataset-II is of size  $288 \times 288 \times 3$ . A is transformed into a  $41 \times 41$  image named B which is the visualization of feature vector representing the differential box counts of R channel in A. It may be noted that distinct colors are assigned to different pixel values. It may be observed that the regions where a pathology exists in image A, can be distinctly seen in B which are represented as pixels having values equal to the box-count of a block in the corresponding region of the original image. It is observed that abnormal regions normally lead to greater box height producing larger box-count values. This distinctive capability of the feature extractor is possibly a reason for better performance of the proposed method.

Although classification accuracy of the proposed method is high close to 99% in Dataset-II, there is always a possibility of getting FP and

**Table 12**

Statistical values computed on accuracy scores recorded on 10 rounds of WCNN [10] and the proposed method.

Metrics	WCNN [10]			Proposed Method		
	Mean	Variance	Standard Deviation	Mean	Variance	Standard Deviation
Dataset-I	0.78	$7.6 \times 10^{-5}$	0.009	0.84	$1.6 \times 10^{-5}$	0.004
Dataset-II	0.97	$6.5 \times 10^{-5}$	0.004	0.98	$2.5 \times 10^{-5}$	0.005

**Table 13**

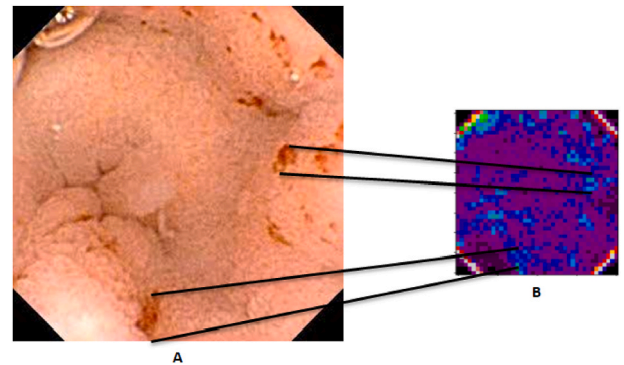
Student's *t*-Test results on accuracy scores of the proposed method and Weak CNN [10].

Dataset	t-statistic	p-value
Dataset-I	- 18.77	$2.89 \times 10^{-13}$
Dataset-II	- 3.16	$5.39 \times 10^{-3}$

**Table 14**

Comparison of the proposed method with other textural feature descriptors.

Features	Dataset-I			Dataset-II		
	Accuracy	F1-score	AUC	Accuracy	F1-score	AUC
Haralick [67]	0.77	0.78	0.77	0.94	0.94	0.93
LBP [68]	0.82	0.82	0.83	0.96	0.97	0.96
KAZE [69]	0.72	0.60	0.65	0.86	0.80	0.85
<b>Proposed Method</b>	<b>0.85</b>	<b>0.85</b>	<b>0.85</b>	<b>0.99</b>	<b>0.98</b>	<b>0.99</b>

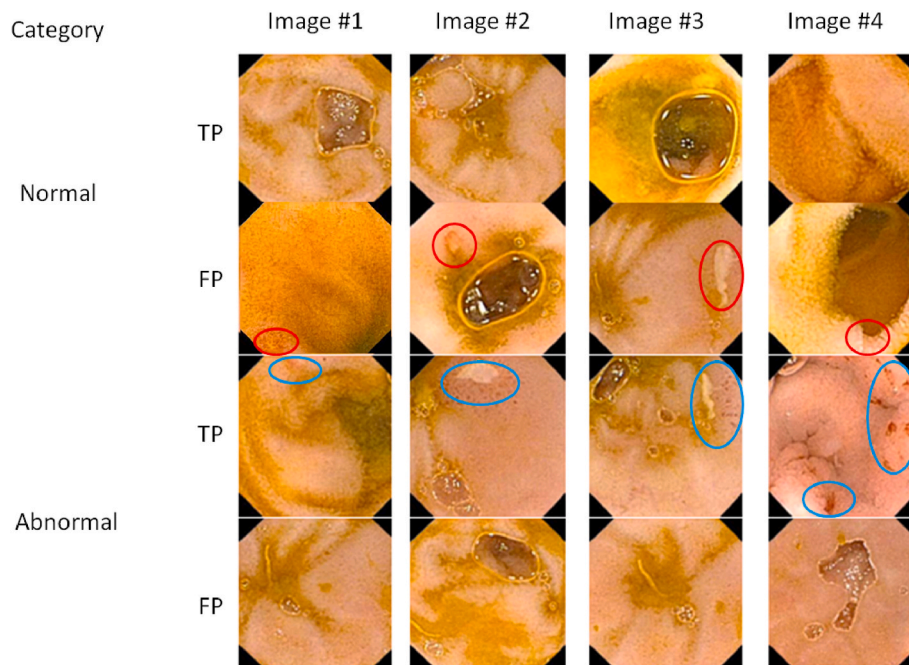


**Fig. 4.** An example of WCE image with its corresponding features. (A) A  $288 \times 288 \times 3$  WCE image containing abnormal findings. (B) A  $41 \times 41$  image representing features computed from DBC method on R channel of image A converted into an image for visualization. Bold lines showing correlation of abnormalities in A with pixels representing box-count in B highlighted with light blue color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

TN results. Fig. 5 gives FP and TP classification results of sixteen randomly selected samples from normal and abnormal images of Dataset-II, as produced by the proposed method. The FP results are investigated for the identification of the reason behind inaccurate results on the application of the proposed method. It is observed that some normal regions have a texture similar to abnormal regions therefore few normal images are falsely classified as abnormal under FP of a normal category. Since the size of an abnormality can be either very small especially when it is read inside  $7 \times 7$  windows, the contribution of the anomalous features will be negligible which falsely makes an abnormal image predicted as normal. As the image is divided into blocks, pixels containing abnormality may get distributed into more than one block and the blocks in the boundary of abnormality can generate features defining normal regions. Probably these reasons are responsible for the prediction of false positive results.

## 5. Conclusion

In this study, a relatively new approach for detection of abnormality in WCE images is discussed. Instead of considering global features, this work focuses on block-wise local features. The local features are extracted from four color channels of each image by the DBC based FD method introduced in Refs. [43]. The SMOTE is adopted to overcome the



**Fig. 5.** An illustration of correctly classified and mis-classified WCE images using the proposed method where blue ellipses show presence of abnormality in correctly classified abnormal images and red ellipses shows pathology in mis-classified abnormal images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

high imbalance problem. The experimental results illustrate that the proposed method outperforms some of the state-of-the-art methods in terms of accuracy, and F1-score when the block size is 7. These datasets are small and highly imbalanced, it would be good to test the proposed method on other datasets as well. As a future work, we aim to tackle localization of abnormality and not only detecting the presence of it.

#### Declaration of competing interest

The authors declare no conflict of interest.

#### Acknowledgment

The authors would like to express their sincere thanks to the anonymous reviewers whose comments and suggestions have helped in improving the quality of the paper. This work is partially supported by the project “Prediction of diseases through computer assisted diagnosis system using images captured by minimally-invasive and non-invasive modalities”, Computer Science and Engineering, PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur India (under ID: SPARC-MHRD-231). This work is also partially supported by the project IT4Neuro(degeneration), reg. nr. CZ.02.1.01/0.0/0.0/18\_069/0010054 and by the project “Smart Solutions in Ubiquitous Computing Environments”, Grant Agency of Excellence, University of Hradec Kralove, Faculty of Informatics and Management, Czech Republic (under ID: UHK-FIM-GE-2020).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbiomed.2020.104094>.

#### References

- [1] F. Gong, C. Swain, T. Mills, An endorobot for gastrointestinal endoscopy, *Gut* 35 (1994) S52.
- [2] G. Iddan, G. Meron, A. Glukhovskiy, P. Swain, Wireless capsule endoscopy, *Nature* 405 (6785) (2000), 417–417.
- [3] M. Mylonaki, A. Fritscher-Ravens, P. Swain, Wireless capsule endoscopy: a comparison with push enteroscopy in patients with gastroscopy and colonoscopy negative gastrointestinal bleeding, *Gut* 52 (8) (2003) 1122–1126.
- [4] I. Tachei, Kapslová Endoskopie, Nucleus HK, 2008.
- [5] A. Koulaouzidis, D.K. Iakovidis, A. Karargyris, J.N. Plevris, Optimizing lesion detection in small-bowel capsule endoscopy: from present problems to future solutions, *Expet Rev. Gastroenterol. Hepatol.* 9 (2) (2015) 217–235.
- [6] B. Li, M.Q.-H. Meng, Computer-aided detection of bleeding regions for capsule endoscopy images, *IEEE Trans. Biomed. Eng.* 56 (4) (2009) 1032–1039.
- [7] S.A. Karkanis, D.K. Iakovidis, D. Karras, D. Maroulis, Detection of lesions in endoscopic video using textural descriptors on wavelet domain supported by artificial neural network architectures, in: *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, vol. 2, IEEE, 2001, pp. 833–836.
- [8] A.K. Kundu, S.A. Fattah, M.N. Rizve, 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. Healthcare Eng.* (2018), 9423062.
- [9] Y. Yuan, B. Li, M.Q.-H. Meng, Bleeding frame and region detection in the wireless capsule endoscopy video, *IEEE J.Biomed. Health Inf.* 20 (2) (2015) 624–630.
- [10] D.K. Iakovidis, S.V. Georgakopoulos, M. Vasilakakis, A. Koulaouzidis, V. P. Plagianakos, Detecting and locating gastrointestinal anomalies using deep learning and iterative cluster unification, *IEEE Trans. Med. Imag.* 37 (10) (2018) 2196–2210.
- [11] Z. Xiao, L.N. Feng, A study on wireless capsule endoscopy for small intestinal lesions detection based on deep learning target detection, *IEEE Access* 8 (2020) 159017–159026.
- [12] M.A. Khan, S. Kadry, M. Alhaisoni, Y. Nam, Y. Zhang, V. Rajinikanth, M.S. Sarfraz, Computer-aided gastrointestinal diseases analysis from wireless capsule endoscopy: a framework of best features selection, *IEEE Access* 8 (2020) 132850–132859.
- [13] V.S. Sadasivan, C.S. Seelamantula, High accuracy patch-level classification of wireless capsule endoscopy images using a convolutional neural network, in: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, 2019, pp. 96–99.
- [14] T. Ghosh, S.A. Fattah, K.A. Wahid, Chobs: color histogram of block statistics for automatic bleeding detection in wireless capsule endoscopy video, *IEEE J.Transl. Eng. Health Med.* 6 (2018) 1–12.
- [15] D.K. Iakovidis, A. Koulaouzidis, Automatic lesion detection in wireless capsule endoscopy—a simple solution for a complex problem, in: *2014 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2014, pp. 2236–2240.
- [16] Y. Yuan, M.Q.-H. Meng, Deep learning for polyp recognition in wireless capsule endoscopy images, *Med. Phys.* 44 (4) (2017) 1379–1389.
- [17] A. Novozámský, J. Flusser, I. Tachei, L. Sulík, J. Bureš, O. Krejcar, Automatic blood detection in capsule endoscopy video, *J. Biomed. Optic.* 21 (12) (2016), 126007.
- [18] A. A. Al-Rahayfeh, A. A. Abuzneid, Detection of Bleeding in Wireless Capsule Endoscopy Images Using Range Ratio Color, *arXiv preprint arXiv:1005.5439*.



- [19] X. Jia, M.Q.-H. Meng, A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016, pp. 639–642.
- [20] A.K. Sekuboyina, S.T. Devarakonda, C.S. Seelamantula, A convolutional neural network approach for abnormality detection in wireless capsule endoscopy, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE, 2017, pp. 1057–1060.
- [21] C. Panigrahy, A. Seal, N.K. Mahato, Fractal dimension of synthesized and natural color images in lab space, *Pattern Anal. Appl.* 23 (2) (2020) 819–836.
- [22] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (2002) 321–357.
- [23] A.S. Ashour, N. Dey, W.S. Mohamed, J.G. Tromp, R.S. Sherratt, F. Shi, L. Moraru, Colored video analysis in wireless capsule endoscopy: a survey of state-of-the-art, *Curr. Med. Imag.* 16 (1) (2020).
- [24] J. Liu, X. Yuan, Obscure bleeding detection in endoscopy images using support vector machines, *Optim. Eng.* 10 (2) (2009) 289–299.
- [25] G. Lv, G. Yan, Z. Wang, Bleeding detection in wireless capsule endoscopy images based on color invariants and spatial pyramids using support vector machines, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 6643–6646.
- [26] S. Sainju, F.M. Bui, K. Wahid, Bleeding detection in wireless capsule endoscopy based on color features from histogram probability, in: 2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), IEEE, 2013, pp. 1–4.
- [27] Y. Yuan, B. Li, M.Q.-H. Meng, Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images, *IEEE Trans. Autom. Sci. Eng.* 13 (2) (2015) 529–535.
- [28] H. Chen, S. Wang, Y. Ding, D. Qian, Saliency-based bleeding localization for wireless capsule endoscopy diagnosis, *Int. J. Biomed. Imag.* 2017 (2017), 8147632.
- [29] Medtronic, Capsule Endoscopy Products. URL <http://capsuleendoscopy.org/>.
- [30] A. Koulaouzidis, D.K. Iakovidis, D.E. Yung, E. Rondonotti, U. Kopylov, J.N. Plevris, E. Toth, A. Eliakim, G.W. Johansson, W. Marlicz, et al., Kid project: an internet-based digital video atlas of capsule endoscopy for research purposes, *Endosc. Int. Open* 5 (6) (2017) E477.
- [31] X. Jia, M.Q.-H. Meng, 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, 2017, pp. 3154–3157.
- [32] Z. Li, K. Kamnitsas, B. Glocker, Overfitting of neural nets under class imbalance: analysis and improvements for segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 402–410.
- [33] Q. Xu, M. Zhang, Z. Gu, G. Pan, Overfitting remedy by sparsifying regularization on fully-connected layers of cnns, *Neurocomputing* 328 (2019) 69–74.
- [34] S. Liu, An improved differential box-counting approach to compute fractal dimension of gray-level image, in: 2008 International Symposium on Information Science and Engineering, vol. 1, IEEE, 2008, pp. 303–306.
- [35] B.B. Chaudhuri, N. Sarkar, Texture segmentation using fractal dimension, *IEEE Trans. Pattern Anal. Mach. Intell.* 17 (1) (1995) 72–77.
- [36] L. Yu, D. Zhang, K. Wang, W. Yang, Coarse iris classification using box-counting to estimate fractal dimensions, *Pattern Recogn.* 38 (11) (2005) 1791–1798.
- [37] A. Seal, C. Panigrahy, Human authentication based on fusion of thermal and visible face images, *Multimed. Tool. Appl.* 78 (21) (2019) 30373–30395.
- [38] A.N. Esgiar, R.N. Naguib, B.S. Sharif, M.K. Bennett, A. Murray, Fractal analysis in the detection of colonic cancer images, *IEEE Trans. Inf. Technol. Biomed.* 6 (1) (2002) 54–58.
- [39] H. Eguiraun Martínez, M. K. López de Ipiña Peña, M. Falarza, M. Iciar, Application of Entropy and Fractal Dimension Analyses to the Pattern Recognition of Contaminated Fish Responses in Aquaculture.
- [40] I. Jamaludin, M.C. Azemin, A. Sapuan, A. Zainuddin, R. Hassan, 2d and 3d complexity analysis on mri images using fractal dimension, *J. Telecommun. Electron. Comput. Eng.* 10 (1–8) (2018) 161–164.
- [41] C. Panigrahy, A. Seal, N. Mahato, Mri and spect image fusion using a weighted parameter adaptive dual channel pcnn, in: IEEE Signal Process. Lett. Pubmed Partial Author article title stitle stitle Volume Page, 27, 2020, pp. 690–694.
- [42] C. Panigrahy, A. Seal, N.K. Mahato, D. Bhattacharjee, Differential box counting methods for estimating fractal dimension of gray-scale images: a survey, *Chaos, Solitons & Fractals* 126 (2019) 178–202.
- [43] C. Panigrahy, A. Seal, N.K. Mahato, Quantitative texture measurement of gray-scale images: fractal dimension using an improved differential box counting method, *Measurement* 147 (2019), 106859.
- [44] N. Sarkar, B.B. Chaudhuri, An efficient differential box-counting approach to compute fractal dimension of image, *IEEE Trans. Syst. Man Cybern.* 24 (1) (1994) 115–120.
- [45] X. Jin, S. Ong, et al., A practical method for estimating fractal dimension, *Pattern Recogn. Lett.* 16 (5) (1995) 457–464.
- [46] W.-S. Chen, S.-Y. Yuan, C.-M. Hsieh, Two algorithms to estimate fractal dimension of gray-level images, *Opt. Eng.* 42 (8) (2003) 2452–2465.
- [47] J. Li, C. Sun, Q. Du, A new box-counting method for estimation of image fractal dimension, in: 2006 International Conference on Image Processing, IEEE, 2006, pp. 3029–3032.
- [48] J. Li, Q. Du, C. Sun, An improved box-counting method for image fractal dimension estimation, *Pattern Recogn.* 42 (11) (2009) 2460–2469.
- [49] Y. Liu, L. Chen, H. Wang, L. Jiang, Y. Zhang, J. Zhao, D. Wang, Y. Zhao, Y. Song, An improved differential box-counting method to estimate fractal dimensions of gray-level images, *J. Vis. Commun. Image Represent.* 25 (5) (2014) 1102–1111.
- [50] K. Lai, C. Li, T. He, L. Chen, K. Yu, W. Zhou, Study on an improved differential box-counting approach for gray-level variation of images, in: 2016 10th International Conference on Sensing Technology (ICST), IEEE, 2016, pp. 1–6.
- [51] M. Long, F. Peng, A box-counting method with adaptable box height for measuring the fractal feature of images, *Radioengineering* 22 (1) (2013) 208–213.
- [52] C. Panigrahy, A. Seal, N.K. Mahato, Fractal dimension of synthesized and natural color images in lab space, *Pattern Anal. Appl.* 23 (2) (2019) 819–836.
- [53] P. Cunningham, S. J. Delany, K-Nearest Neighbour Classifiers–, arXiv preprint arXiv:2004.04523.
- [54] J. Kim, B. Kim, S. Savarese, Comparing image classification methods: K-nearest-neighbor and support-vector-machines, in: Proceedings of the 6th WSEAS International Conference on Computer Engineering and Applications, and Proceedings of the 2012 American Conference on Applied Mathematics, vol. 1001, 2012, 48109–2122.
- [55] Z. Wang, G. Yu, Y. Kang, Y. Zhao, Q. Qu, Breast tumor detection in digital mammography based on extreme learning machine, *Neurocomputing* 128 (2014) 175–184.
- [56] A. Seal, P.P.N. Reddy, P. Chaithanya, A. Meghana, K. Jahnavi, O. Krejcar, R. Hudak, An eeg database and its initial benchmark emotion classification performance, *Comput. and Math. Meth. Med.* 2020 (2020), 8303465.
- [57] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, *Mach. Learn.* 29 (2–3) (1997) 131–163.
- [58] B.J. Erickson, P. Korfiatis, Z. Akkus, T.L. Kline, Machine learning for medical imaging, *Radiographics* 37 (2) (2017) 505–515.
- [59] A. Seal, D. Bhattacharjee, M. Nasipuri, D. Rodríguez-Esparragón, E. Menasalvas, C. Gonzalo-Martin, Pet-ct image fusion using random forest and à-trous wavelet transform, *Int. J. Numer. Meth. Biomed. Eng.* 34 (3) (2018), e2933.
- [60] A. Seal, D. Bhattacharjee, M. Nasipuri, Human face recognition using random forest based fusion of à-trous wavelet transform coefficients from thermal and visible images, *AEU-Int. J. Electron. Commun.* 70 (8) (2016) 1041–1049.
- [61] A. Seal, A. Garcia-Pedrero, D. Bhattacharjee, M. Nasipuri, M. Lillo-Saavedra, E. Menasalvas, C. Gonzalo-Martin, Multi-scale rois selection for classifying multi-spectral images, *Multidimens. Syst. Signal Process.* (2019) 1–25.
- [62] L.E. Raileanu, K. Stoffel, Theoretical comparison between the gini index and information gain criteria, *Ann. Math. Artif. Intell.* 41 (1) (2004) 77–93.
- [63] K.K. Sharma, A. Seal, Modeling uncertain data using Monte Carlo integration method for clustering, *Expert Syst. Appl.* 137 (2019) 100–116.
- [64] K.K. Sharma, A. Seal, Clustering analysis using an adaptive fused distance, *Eng. Appl. Artif. Intell.* 96 (2020) 103928.
- [65] K.K. Sharma, A. Seal, Multi-view spectral clustering for uncertain objects, *Inf. Sci.* 547 (8) (2020) 723–745.
- [66] L. Johnston, Student's t-test, *J. Qual. Technol.* 2 (4) (1970) 243–245.
- [67] A. Porebski, N. Vandenbroucke, L. Macaire, Haralick feature extraction from lbp images for color texture classification, in: 2008 First Workshops on Image Processing Theory, Tools and Applications, IEEE, 2008, pp. 1–8.
- [68] A.F. Constantinescu, M. Ionescu, I. Rogoveanu, M.E. Ciurea, C.T. Streba, V. F. Iovanescu, S.A. Artene, C.C. Vere, Analysis of wireless capsule endoscopy images using local binary patterns, *Appl. Med. Inf.* 36 (2) (2015) 31–42.
- [69] P.F. Alcantarilla, A. Bartoli, A.J. Davison, Kaze features, in: European Conference on Computer Vision, Springer, 2012, pp. 214–227.