Use of adaptive hybrid filtering process in Crohn's disease lesion detection from real capsule endoscopy videos

Vasileios S. Charisis, Leontios J. Hadjileontiadis ✉

Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 54636, Greece
✉ *E-mail: leontios@auth.gr*

Published in Healthcare Technology Letters; Received on 26th November 2015; Revised on 9th February 2016; Accepted on 16th February 2016

The aim of this Letter is to present a new capsule endoscopy (CE) image analysis scheme for the detection of small bowel ulcers that relate to Crohn's disease. More specifically, this scheme is based on: (i) a hybrid adaptive filtering (HAF) process, that utilises genetic algorithms to the curvelet-based representation of images for efficient extraction of the lesion-related morphological characteristics, (ii) differential lacunarity (DL) analysis for texture feature extraction from the HAF-filtered images and (iii) support vector machines for robust classification performance. For the training of the proposed scheme, namely HAF-DL, an 800-image database was used and the evaluation was based on ten 30-second long endoscopic videos. Experimental results, along with comparison with other related efforts, have shown that the HAF-DL approach evidently outperforms the latter in the field of CE image analysis for automated lesion detection, providing higher classification results. The promising performance of HAF-DL paves the way for a complete computer-aided diagnosis system that could support the physicians' clinical practice.

1. Introduction: The advent of capsule endoscopy (CE) [1] marked a revolution in the field of gastroenterology and medical imaging as it permitted, for the first time, non-invasive and successful visual inspection of the entire length of small bowel (SB). A disposable miniature capsule is swallowed by the patient and, along its journey through the digestive track (DT), it captures and wirelessly transmits two frames per second at a wearable data recorder. At the end of the examination, these images are transferred from the recorder to a personal computer and reviewed by the physician. CE is beneficial in evaluating various disorders, such as bleeding, polyps, ulcers and Crohn's disease (CD). CD is a chronic inflammatory bowel disease that causes internal lesions, which evolve to open sores and ulcers, in the DT and especially in SB. Despite the revolutionary benefits introduced by CE, there are still challenging issues to deal with. A CE examination produces 60.000 images, approximately, and the burdensome task of reviewing them is highly time-consuming as it may cost up to 2 h for an experienced clinician. Moreover, abnormalities may appear only in a couple of frames, i.e. milliseconds on the computer screen, and be easily missed. These issues have motivated researchers to develop computer-assisted diagnosis systems in order to reduce the labour of the clinician and the possibility of omitting a lesion.

In the recent literature, various approaches have been reported towards automatic DT content interpretation. However, the majority of them deal with the detection of bleeding, polyps and tumours and only a small proportion targets ulcer and CD-related lesions recognition [2]. The detection of such kind of erosion is rather challenging due to its great diversity in appearance (size, shape, colour and texture) and probably this is the reason for its small popularity among the researchers' community. The most promising proposed efforts utilise various image analysis techniques and texture and colour features extracted from various colour spaces. More specifically, ulcer detection is achieved using: the intensity of pixels at the three colour channels [3], curvelet transform (CT)-based uniform rotation invariant local binary patterns (LBP) classified by multilayer perceptron [4], bag-of-words-based features [scale-invariant feature transform (SIFT) and LBP] [5] and wavelet-based second-order statistical features [6]. In the same direction, the potential of structural features extracted from images presented in the empirical mode decomposition [7] and CT space [8] is investigated. Regarding CD-related lesion detection, the most competent works are based on mean-shift algorithm, colour histogram statistics and Haralick features [9, 10] and a fusion of MPEG-7 descriptors and support vector machines (SVM) classifiers [11]. The above approaches, although they exhibit promising results, they suffer from the shortcomings of LBP, SIFT and MPEG-7 tools and engage quite limited databases.

In this Letter, an innovative CE image analysis system is introduced for the efficient detection of CD-related lesions. This system combines, for the first time, the space-frequency of curvature structures domain [8] with the spatial morphology distribution domain [7] that both have been proven to be in line with the basic characteristics of the inflammatory tissue recognition problem and provide with useful information towards this direction, when engaged individually. The space-frequency domain is defined by the curvelet-based sub-images and is used in order to extract the lesion-related structural characteristics and reconstruct more informative images via a filtering process called hybrid adaptive filtering (HAF). The spatial morphology distribution domain is defined by the differential lacunarity (DL) analysis that is applied on the HAF-processed images in order to extract texture features. The classification step is implemented by SVM and the whole process takes place in the Cr channel of YCbCr space.

2. Background

2.1. Curvelet transform: CT constitutes the central part of HAF section of the proposed approach that targets to reveal multi-resolution and multi-scale curved structural components of CD lesions towards efficient feature extraction. The motive for using multi-resolution and multi-scale analysis is that CD lesions are characterised by great diversity in terms of size, shape, orientation and illumination and the background variation is intense. Thus, a robust tool is required that is capable of revealing structural and morphological information in multiple scales and directions. In addition, since the lesions are not characterised by a specific shape, the analysis has to capture efficiently curved structures as they accommodate the majority of informative content.

A popular tool that has been used for multi-resolution image analysis is wavelet transform (WT). WT is very efficient in describing

one-dimension (1D), piecewise smooth signals and capturing point singularities. However, 2D signals exhibit edges (1D singularities), being typically smooth curves that separate smooth regions, that cannot be described appropriately by wavelets. The reason is that 2D wavelets are constructed by a tensor product of 1D wavelets and follow an isotropic scaling law. These characteristics lead to low adaptation to the geometry of the curve as well as to isolation of strongly directional (horizontal, vertical and diagonal) structures.

In an attempt to overcome these weaknesses, CT has been introduced [12]. The main idea of CT is to engage a superposition of anisotropic functions of multiple widths and lengths, that follow a parabolic scaling law (width $\approx$ length^2), in order to represent a curve. For the definition of continuous CT, two windows are used, an angular $V(\theta)$ and a radial $W(r)$, that are real-valued, non-negative and smooth. The product of V and W is a parabolic wedge U_j that constitutes the Fourier transform (FT) of a mother curvelet $\varphi_j(x)$. All the other curvelets at scale 2^{-j}, orientation θ_l and position $x_k^{(j,l)}$ are obtained by rotations, scaling and translations of φ_j. A curvelet coefficient is then defined as the inner product between an element $M \varepsilon R^2$ and a curvelet $\varphi_{j,l,k}$. The needle-shaped elements of CT are highly sensitive in the direction and their number depends on the analysis scale (finer scales contain more curvelets than coarser scales); hence, providing better analysis capabilities that WT counterparts [12].

The continuous CT can be extended to the digital space via either unequispaced fast FT or wrapping. Both techniques have the same complexity, however, the wrapping algorithm is somewhat simpler and, thus, more popular [12].

2.2. DL analysis: DL is a robust tool for multi-scale and translation invariant analysis of spatial patterns of dispersion, able to identify slight or sharp changes in pixel neighbourhoods with no direction selectivity, useful in the case of CE images.

2.2.1. Lacunarity: Lacunarity [13], meaning *gappiness* and deriving from the Latin term *lacuna*, is a fractal property, counterpart to fractal dimension (FD) that was proposed in order to differentiate surfaces that exhibit the same FD, but vary in appearance. More specifically, lacunarity describes the texture of a fractal by analysing the distribution of gap sizes in the data set across multiple scales. The more gaps with a broad range of sizes a set contains, the higher the lacunarity index. In a similar context, lacunarity is defined as a measure of heterogeneity and translational invariance. Sets with almost uniform distribution of gaps can be considered homogenous and are characterised by lower lacunarity than translationally and rotationally variant (heterogeneous) sets. However, homogeneity depends on the analysis scale. In other words, sets that appear homogeneous at large scales can be rather heterogeneous at smaller scales and vice versa. From this perspective, lacunarity is regarded as a scale-dependent measure of texture of an object [13]. Various techniques have been proposed to calculate lacunarity, but the simplest and most popular is the gliding box algorithm (GBA) [14] that is functional on 1D binary data.

2.2.2. Differential lacunarity: DL is another version of lacunarity that was introduced by Dong [15] in order to enable multi-scale lacunarity-based texture analysis of grayscale images. The extension of GBA in 2D is quite straightforward, but the requirement of binary data, even if it can be accomplished by thresholding, constitutes a serious limitation. The concept behind DL is the same as in lacunarity, but its calculation is based on a differential box counting method that engages a gliding window W ($w \times w$ pixels) and a gliding box R ($r \times r$ pixels). W scans the entire image and determines the area in which the 'mass' M will be calculated. R is an auxiliary box that helps to determine M at every position of W. In this context, w and r determine the analysis scale. If $Q(M, r, w)$ is the probability function of the distribution of M across the image, DL at scale r and w is defined as

$$\Lambda(w, r) = \frac{\sum_M M^2 Q(M, w, r)}{\left[\sum_M MQ(M, w, r)\right]^2}. \quad (1)$$

To achieve multi-scale analysis, it is common practice to calculate DL for a variety of scales, resulting in a DL–w curve, which characterises the specific space-filling pattern.

3. Proposed approach

3.1. Overview: The aim of the proposed HAF-DL system is that given a region of interest (ROI) J within a CE image, detect if J depicts CD lesion or normal mucosa. The overall structure of HAF-DL scheme is presented in Fig. 1. At first, the image is converted from RGB to YCbCr space and the Cr channel is isolated since it has been shown [8] that the majority of lesion-related information lie on the red–green difference plane. This conclusion was attributed to the fact that the reflected green light depends on the blood volume and the existence of lesion disturbs the microstructure of blood vessels and the thickness of mucosa. YCbCr space was selected because it is a perceptually uniform colour space that separates colour from brightness information and overcomes the disadvantage of high correlation between the RGB channels [16]. Then, the Cr channel is inputted to HAF section of HAF-DL scheme. The role of this filtering procedure is to identify the CD lesion-related characteristics, facilitating the feature extraction procedure that follows. In this perspective, HAF entails a genetic algorithm (GA) that acts upon the representation of CE images on the CT space. In the latter, the image (Cr channel) is decomposed into a series of curvelet-based sub-images of various scales and orientations. Then, GA, by engaging a fitness function (FF) that is based on the distribution of spatial relationships between pixels, selects the optimum sub-images that relate the most with the CD lesion-associated characteristics. The selected sub-images are combined through a reconstruction process, i.e. inverse CT, to produce a refined image. The latter enters the DL section of the HAF-DL scheme, where DL-based analysis takes place, resulting in powerful colour-texture information that after light processing produce an efficient feature vector (FV). At last, SVM-based classification is performed.

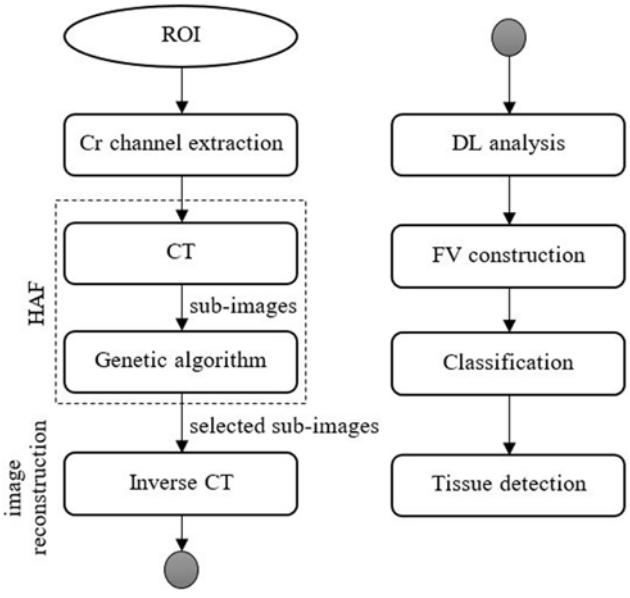

Fig. 1 *Proposed HAF-DL scheme*

doi: 10.1049/htl.2015.0055

3.2. Hybrid adaptive filtering: The aim of HAF process is to extract the CE image characteristics and isolate the ones that mostly relate to the CD lesion information. As denoted by the word 'hybrid', HAF is composed of two analysis tools, i.e. CT and a GA optimisation concept, in order to implement an adapted to the signal filtering procedure. CT can be considered as a filter bank that provides a pseudo-spectral representation since it decomposes an image into components (sub-images) at multiple scales and orientations. In order to take full advantage of the preceding capability, a GA-based technique was engaged for the identification and selection of the most optimum sub-images that correspond to informative structures within the image. The motive for such a concept was the observation that sub-images at specific scales and directions present higher CD-lesion detection capabilities than others [8].

One primary component of HAF is the FF of the GA, since it defines the selection criterion on which the filtering process is structured. In this Letter, a FF was applied that is based on the textural and morphological characteristics of the sub-images, as expressed by the DL–w curve. As noted above, DL is a scale-dependent measure of textural patterns and distributions. To this end, an image with uniform structures and patterns exhibits lower DL values than an image that is composed of arbitrary patterns. The DL–w curve can be regarded as a multi-scale representation of texture and its decay can disclose the existence of specific structures in the image. For instance, the presence of micro-patterns with moderate differentiation for a variety of analysis scales leads to a DL–w curve with smaller gradient than that of a DL–w curve extracted from either highly irregular and diverse or almost uniform structures. The rationale of applying such a FF is to capture the variations in morphological and textural characteristics of CE images. Images that produce DL–w curves with small gradient may account for informative mucosa, either normal or eroded, whereas images with DL–w curves with steep slope may correspond to worthless content, such as bubbles and folds. From this perspective, the DL-based filtering would suppress the components of the initial image that do not relate to useful texture information. This hypothesis is justified by the observations made, based on the results of [8], where the efficient sub-images corresponded to DL–w curves with smaller gradient than that of the less efficient sub-images.

3.3. Feature vector: HAF process is followed by DL analysis that targets the calculation of efficient colour-texture features. As noted before, the appearance of CD-related lesions varies a lot; hence, the texture analysis tool should be capable to achieve multi-scale, rotation invariant feature extraction. Such a tool is DL due to, not only its simple calculation, but also its precision in identifying either slight or sharp changes in pixel neighbourhoods at various scales and without direction selectivity, necessary in the case of CE data. The multi-scale analysis property of DL is triggered by calculating the DL value for a variety of window sizes (w) and a specific box size (r) instead of a single set of parameters. The reason for manipulating the analysis scale through w instead of r is that w determines the region on which the mass M is going to be calculated, and in line with [17], the bigger the area of M calculation, the coarser the analysis turns. On the other hand, r determines only to a certain degree the analysis scale as it mainly affects the granularity of intensity variation recognition. Consequently, in the current approach, by selecting a constant and small value for r and varying values for w, identification of slight variations in neighbouring pixels and multi-scale analysis is achieved. Last but not least, the DL–w curves are normalised to the DL value that corresponds to the smallest w in order to secure an identical reference level that has been shown to be essential [18]. To this end, the normalised DL–w curve ($\Lambda^N(w)$) may form the FV. However, the observation that the DL–w curves resemble the behaviour of hyperbola, as well as the need to reduce the FV dimension, highlights the essence of modelling the normalised DL–w curves with the function of hyperbola that is defined as

$$L(w) = \frac{b}{w^a} + c,\ w = [w_{\min}, w_{\max}], \tag{2}$$

where a represents the convergence of $L(w)$, b is the concavity of hyperbola and c is the translational term. The best interpretation of $\Lambda^N(w)$ by the model $L(w)$ is computed as the solution of a least squares problem, where parameters a, b and c are the independent variables [7]. Another way to extract useful features from the DL analysis is to use six statistical measures that are calculated on the DL–w curve [5, 19]. The six common statistical features extracted from $\Lambda^N(w)$ curve are mean, standard deviation, entropy, energy, skewness and kurtosis. It has been shown [20] that the utilisation of the aforementioned statistical measures along with the [a, b, c] parameters, that both express the global behaviour of the DL–w curve, but from a different perspective, and the three first values of DL–w curve ($\Lambda^N(w_{\min}+1)$, $\Lambda^N(w_{\min}+2)$ and $\Lambda^N(w_{\min}+3)$; the value $\Lambda^N(w_{\min})$ is not used as it is always equal to one), that express the local behaviour of the DL–w curve, constitutes the most efficient FV (FV_1). For comparison purposes, two other FVs will be examined: FV_2 contains all the values from the normalised DL–w curve and FV_3 contains the six statistical measures along with the parameters [a, b, c].

4. Experimental dataset and implementation

4.1. Training and testing dataset: To develop a competent and robust CD-related lesion detection methodology, apart from employing suitable mathematical tools, a sufficiently rich database is required, on which the proposed scheme will be trained and tested. Unfortunately, the majority of related reported efforts in the literature (59%) utilise databases that consist of no more than 500 images [2]. Using a limited training and testing dataset, or an expanded dataset but with highly correlated entries may cause unrealistic improved performance and not allow safe generalisation of the conclusions.

In this Letter, in order to construct the training dataset, 950 CE frames were used. More specifically, 500 frames depicting lesion free tissue and 450 frames presenting CD-related lesions and erosion were acquired from 13 patients who undertook a CE examination in NIMTS Gastroenterology Clinic, Athens, Greece. It should be highlighted that, in order to achieve the lowest possible similarity, each lesion frame comes from a different lesion event, which means that no two frames depict the same lesion region, even if it is rotated or zoomed in/out. Moreover, the lesions depicted are of various severities and sizes, based on the Lewis scores [21]. In the same direction, the non-lesion frames depict, not only normal mucosa, but also misleading content, such as bubbles, folds, villus, intestinal juices and so on, in order to create realistic conditions that hamper the detection problem. Afterwards, two experts, upon mutual agreement, cropped each image to a ROI, since the detection algorithms cannot be applied to the entire CE frame due to its rich content. Each lesion ROI contains the entire or part of the eroded region, while in cases of extended damage, the whole ROIs lie within the lesion region. Then, two other clinicians rated these ROIs, twice, and only the ones that were classified all four times in the same class were included in the dataset. This procedure allowed to assess the inter-/intra-rater variability and acquire a highly confident dataset that consists of 400 normal and 400 lesion ROIs. Some exemplary CE frames along with the corresponding ROIs are given in Fig. 2. It should be commented that the size of the ROIs varies from 40×40 to 120×120 pixels, since the extracted features are minimally affected by the resolution of ROI, and can be located anywhere in the frame. Deterministic ROI size and position in the frame was not an option in order to achieve as much generalisation as possible and produce a robust system.

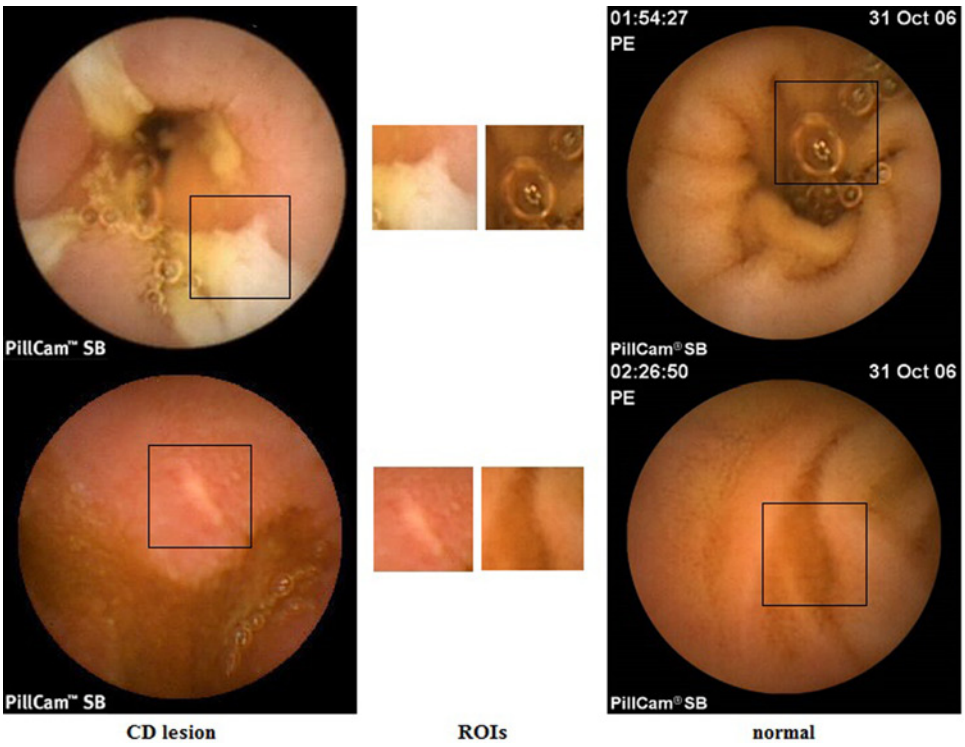

Fig. 2 *CE frames from the adopted training dataset and corresponding ROIs*

As far as the testing procedure is concerned, we did not use the above database with some sort of cross validation. On the contrary, ten 30 s (sec) long CE videos were employed that came from five new patients. From each video, 60 frames were extracted, considering the speed of the camera of the endoscopic capsule. Then each frame was segmented automatically in 12 ROIs (60 × 60 pixel), as presented in Fig. 3, since the proposed scheme was trained in such ROIs and could not provide efficient performance in case of entire CE frames. Each video contained at least one lesion, while each frame depicted no, one or more abnormal regions. The size of the lesions varied and in some cases it was big enough to fit in more than one ROI. The ROIs were labelled by two experts, upon mutual agreement.

4.2. HAF and DL implementation: Regarding CT implementation, the number of analysis scales and the number of analysis angles at the second scale had to be set. Related applications tend to apply three or four scales [4]; hence, in this approach, after exhaustive experiments, we opted for four [8]. As far as the number of angular resolution is concerned, it has been shown [8] that the optimum number of angles at the second scale is eight (minimum eligible value), as a trade-off between data redundancy and complexity. Other approaches use 16 angles [3], however, a p-test showed that differences between the angles 1–5, 2–6, 3–7 and 4–8 (in the case of eight angles) were statistically significant at 3.2%, while the statistical difference between the angles 1–9, 2–10, …, 8–16 (in the case of 16 angles) were statistically significant at 9.8% on average [20].

Considering the implementation of DL analysis, $\Lambda(w)$ was calculated for gliding box size $r = 3$ pixels and gliding window size $w = 4$–30 pixels, so as to be sufficiently lower than the size of ROI. Moreover, pilot experiments showed that the DL value for $w > 30$ is almost constant. The value of r was selected after exhaustive trials.

The GA used in HAF was based on [22]. The initial population was set to 20 chromosomes and the number of generations was set to 50. The mutation probabilities were $Pm(0\rightarrow1) = 0.001$ and $Pm(1\rightarrow0) = 0.01$, for eliminating tendencies to select the majority of sub-images.

For the HAF-based curvelet sub-image selection, the 25% of the training dataset was used. In other words, 100 normal and 100 abnormal samples were randomly selected. For each generation of GA, the FF value was calculated accordingly to the whole dataset

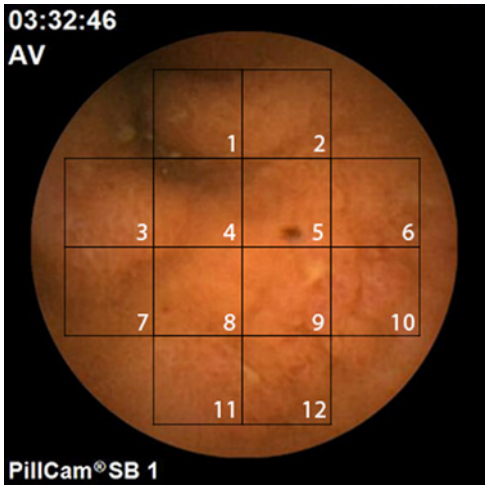

Fig. 3 *Predetermined CE image segmentation in 12 regions (60 × 60 pixel) on which the HAH-DL scheme will be applied for the testing procedure*

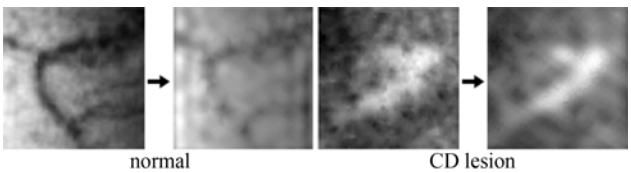

Fig. 4 *Cr channel of a normal and a CD lesion WCE image prior to (left) and after (right) the application of HAF*

doi: 10.1049/htl.2015.0055

Table 1 CD lesion versus normal features of HAF-DL_1 accompanied by *t*-test analysis. The format #±# corresponds to mean ± standard deviation

Feature	CD lesion	Normal	*p*-value
mean	0.8511 ± 0.0622	0.7740 ± 0.1158	2.8869×10^{-9}
standard deviation	0.0415 ± 0.0191	0.0595 ± 0.0367	4.2577×10^{-6}
energy	0.7302 ± 0.0979	0.6171 ± 0.1722	7.6905×10^{-8}
entropy	3.3989 ± 0.4665	3.8563 ± 0.6381	0.0486
skewness	1.4354 ± 0.4229	1.8545 ± 0.6195	3.9515×10^{-5}
kurtosis	4.2490 ± 1.4056	6.1426 ± 2.6988	5.7685×10^{-11}
a	2.0688 ± 0.6692	2.5933 ± 0.4849	0.0491
b	-1.2947 ± 0.1582	-1.1752 ± 0.0671	0.0427
c	0.7764 ± 0.1796	0.6333 ± 0.6948	0.0255
$\Lambda^N(w_{\min}+1)$	0.9622 ± 0.0177	0.9563 ± 0.0196	0.0199
$\Lambda^N(w_{\min}+2)$	0.9321 ± 0.0297	0.9195 ± 0.0366	0.0057
$\Lambda^N(w_{\min}+3)$	0.9092 ± 0.0379	0.8902 ± 0.0508	0.0018

of the 200 images. The selected sub-images were found to be (scale/angle): 1/(1), 2/(6), 3/(1, 4, 8, 12), 4/(1, 5, 8, 9, 13, 16) and they coincide to a certain degree with the most efficient sub-images as derived from [8]. Fig. 4 presents the Cr channel of a normal and CD-lesion image, prior (left) and after (right) the application of HAF. It is clear that HAF processing smooth's the image and accentuates the lesion region while diminishing the misleading attribute (deep red vein) of the normal image.

4.3. Classification setup: The classification process of the proposed scheme is based on a SVM classifier with radial basis function [23] which has been used extensively in related applications [5, 8, 11], showing superior performance. The classification performance is evaluated via the measures of sensitivity and specificity under three scenarios that assess the efficient of lesion detection in: ROI level (ROI scenario), frame level (frame scenario) and event level (event scenario). Regarding the frame scenario, one correctly predicted ROI in the frame is sufficient to characterise the frame as positive. On the other hand, the event scenario assumes that if a single lesion event appears in more than one consecutive frames, then, the correct characterisation of just one frame is enough to consider that the specific lesion event has been detected successfully.

4.4. Baseline approaches (BAs): For more effective evaluation of the performance of the proposed HAF-DL scheme, comparison with three of the most competent methodologies reported in the literature, that constitute the BAs in the current work, is performed, i.e. BA1 [3], BA2 [5] and BA3 [9].

5. Results and discussion: The performance of HAF-DL scheme is evaluated through the experimental results derived from the application of the proposed CD lesion detection technique to the ten endoscopic videos. To this end, the recognition results for each classification scenario (ROI, frame and event) are presented.

Prior to presenting the classification results, Table 1 presents the mean value ± standard deviation of the CD lesion versus normal features engaged in HAF-DL_1, as calculated from the training dataset, along with the t-test probability [24] of each feature. From the t-test results, tabulated in Table 1, it can be concluded that all the features included in the proposed FV, namely HAF-DL_1, exhibit significant difference and contribute to the differentiation between abnormal and normal regions.

5.1. ROI scenario: Table 2 presents the classification results in terms of true positives (TP) and true negatives (TN) for each video separately for the three FVs of HAF-DL scheme (HAF-DL_1, HAF-DL_2 and HAF-DL_3) and the three BAs. In the third row, P and N are the actual number of positives and negatives, respectively. The format #–# corresponds to TP–TN.

As far as the three FVs of the proposed HAF-DL scheme are concerned, the results in Table 2 make it clear that the raw data from DL–w curve (FV_2) are less efficient in discriminating lesions from normal tissue than the DL–w synopsis from the statistical measures and the hyperbola parameters (FV_3), in terms of both sensitivity and specificity. More specifically, FV_2 (HAF-DL_2) achieves 40.0–70.0% sensitivity (61.0% mean value) and 63.3–95.9% specificity (79.0% mean value), while FV_3 demonstrates 60.0–75.5% sensitivity (70.0% mean value) and 74.3–91.3% specificity (84.0% mean value). On the other hand, the combination of features that describe the local and the global behaviour of the curve (FV_1) provide with the best performance, achieving 75.0–86.7% sensitivity and 79.0–91.3% specificity, with mean values 81.0 and 87.0%, respectively. The inferior results derived from HAF-DL_2 may be attributed to the increased length of FV_2 that causes the curse of dimensionality effect.

By comparing the results from HAF-DL scheme with those from the BAs, it can be concluded that the advanced nature of HAF as well as the powerful texture features extracted from DL are far more competent in detecting CD lesions. BA_1, which attempts pixel wise erosion detection based on within pixel channel intensity relationships, proves to be unusable since the sensitivity for two videos is 0%, and for the rest ranges around the randomness level (44.4–60.0%). Thus, it can be concluded that exclusive utilisation of colour features is inappropriate for CD lesions analysis. To continue, BA_2, by using CT-based LBP features, achieves slightly improved performance, compared to BA_1, 50.0–70.0% (60.0% mean value) sensitivity and 69.4–87.6% (76.0% mean value) specificity, but the DL-based features prove to be far more efficient. At last, BA_3, by applying MPEG-7 descriptors, although it succeeds the highest results between BAs (60.0–76.7% sensitivity and 78.6–92.0% specificity), is less competent that HAF-DL_1, mainly in detecting the abnormal regions, as mean sensitivity is 10% lower than that of the proposed approach.

5.2. Frame scenario: The classification results at a frame level are presented in Table 3 in a similar way as the results in ROI

Table 2 Classification results at ROI scenario. The format #–# corresponds to TP–TN and $P(N)$ is the actual number of positive (negative) ROIs

Methodology	Number of video									
	1 *P*: 60 *N*: 660	2 *P*: 150 *N*: 570	3 *P*: 5 *N*: 715	4 *P*: 6 *N*: 714	5 *P*: 30 *N*: 690	6 *P*: 20 *N*: 700	7 *P*: 50 *N*: 670	8 *P*: 10 *N*: 710	9 *P*: 83 *N*: 637	10 *P*: 18 *N*: 702
HAF-DL_1	45–592	119–461	4–648	5–602	26–574	17–639	43–529	8–645	68–551	14–620
HAF-DL_2	34–520	91–398	2–659	3–550	19–501	14–572	31–496	5–681	54–403	12–572
HAF-DL_3	40–561	106–430	3–621	3–652	22–529	15–605	36–562	6–637	60–473	12–599
BA_1	31–407	73–351	0–472	0–519	17–453	11–483	28–451	6–582	42–461	8–429
BA_2	36–458	88–401	0–539	3–555	21–499	12–510	33–540	6–622	51–488	9–500
BA_3	39–541	107–473	3–590	4–615	23–542	15–597	36–586	6–653	58–509	12–571

Table 3 Classification results at frame scenario. The format #–# corresponds to TP–TN and *P*(*N*) is the actual number of positive(negative) images

Methodology	Number of video									
	1 *P*:12 *N*:48	2 *P*:45 *N*:15	3 *P*:3 *N*:57	4 *P*:6 *N*:54	5 *P*:14 *N*:46	6 *P*:5 *N*:55	7 *P*:25 *N*:35	8 *P*:10 *N*:50	9 *P*:31 *N*:29	10 *P*:7 *N*:53
HAF-DL_1	10–43	37-14	3–44	5–44	12–43	5–43	22–33	8–44	25-15	6–47
HAF-DL_2	6–35	30-11	1–45	3–40	9–35	3–37	16–31	5–46	20-11	5–43
HAF-DL_3	8–38	32-13	2–42	3–48	10–38	3–40	18–33	6–45	22-13	5–44
BA_1	6–26	26-10	0–35	0–35	9–30	3–29	14–27	6–40	15-12	3–32
BA_2	6–28	30-11	0–38	3–38	10–35	3–32	16–30	6–41	20-11	3–37
BA_3	8–39	32-15	2–40	4–44	11–40	4–40	18–33	6–44	22-12	5–43

scenario. The format #–# corresponds to TP–TN images, while *P* (*N*) is the number of actual positive (negative) images in each video sequence. It should be highlighted that these results are more informative than the previous ones (ROI scenario), since it is more important to evaluate the ability of a scheme to correctly identify positive and negative frames than areas inside the frames. Regarding the FVs of the proposed approach, the first one (FV_1) is the most efficient, as expected, delivering 80.0–100% (84.2% mean value) sensitivity and 51.7–94.3% (83.7% mean value) specificity, while FV_2 and FV_3 achieve 22.2%/8.1% and 15.2%/3.6% lower mean sensitivity/specificity, respectively. The BAs, as in the ROI scenario, exhibit rather inferior performance compared to HAF-DL_1. More specifically, BA_1, BA_2 and BA_3 achieve 0.0–64.3% (51.9% mean value), 0.0–71.4% (61.4% mean value) and 60.0–80.0% (70.9% mean value) sensitivity and 41.4–80.0% (62.4% mean value), 37.9–85.7% (69.8% mean value) and 41.4–100% (79.2% mean value) specificity, respectively. It can be concluded that the proposed methodology is considerably more efficient in correctly classifying CD lesion frames as it delivers 13.3% higher mean sensitivity than the second most potent approach (BA_3). Regarding specificity, BA_3 manages to classify properly all normal frames in video #2; however, HAF-DL_1 achieves the overall highest results. It is noteworthy that the specificity rates of video #9 range from 37.9 to 51.7% due to excessive misleading intestinal content (because of poor patient pre-CE preparation). In case of ignoring these results, the overall specificity of HAF-DL_1 reaches 86.0%.

5.3. Event scenario: Although the above results seem rather satisfactory, from a medical expert's point of view they might seem quite inadequate, especially when it comes to sensitivity, since almost one out five positive frames is omitted. However, these rates correspond to individual frame detection and the latter notion is somewhat unfair. It is not rear that a single lesion appears in more than one consecutive frames; hence, leading to a better chance of detection. Table 4 presents the classification results of event scenario, where the format #/# corresponds to the number of unique lesions detected/the total number of unique lesions, in other words, these fractions correspond to the sensitivity rate. In this Letter, it is observed that HAF-DL_1 scheme identified correctly all lesion events for eight out of ten CE videos, while for the other two videos (#2 and #7) the success rate was 85.7 and 83.3%, respectively, leading to 93.5% mean sensitivity. On the other hand, BA1, BA2 and BA3 achieved 54.8, 64.5 and 77.4% mean sensitivity, respectively. Hence, the proposed scheme, by employing a CT-based adaptive filtering techniques and DL-based texture features is far more efficient than other similar techniques in the literature.

Table 4 Classification results at event scenario. The format #/# corresponds to number of unique lesions detected/total number of unique lesions

Methodology	Percentage of video									
	1	2	3	4	5	6	7	8	9	10
HAF-DL_1	3/3	6/7	1/1	1/1	5/5	2/2	5/6	2/2	3/3	1/1
HAF-DL_2	2/3	5/7	1/1	1/1	3/5	1/2	4/6	1/2	2/3	1/1
HAF-DL_3	2/3	5/7	1/1	1/1	3/5	1/2	4/6	1/2	2/3	1/1
BA_1	2/3	4/7	0/1	0/1	3/5	1/2	3/6	1/2	2/3	1/1
BA_2	2/3	5/7	0/1	1/1	3/5	1/2	4/6	1/2	2/3	1/1
BA_3	2/3	5/7	1/1	1/1	4/5	2/2	4/6	2/2	2/3	1/1

5.4. Computational cost: Considering the computational cost, the unoptimised Matlab (The Mathworks, USA) implementation of HAF-DL scheme requires 32.2 s to process each 30 s video on a 4-core, 2.67 GHz desktop computer, something quite disappointing in case of a real-time application. Focusing on even more efficient realisations, other programming languages (such as C++), and multithreading programming should be considered.

6. Conclusion: A novel methodology, namely HAF-DL, for CD-related lesion detection using only the Cr channel of CE images in YCbCr colour model was presented. The proposed approach coupled HAF, an adaptive CT-/GA-based filtering technique for isolating the hidden lesion information, with DL analysis, a robust texture feature extraction tool, resulting in three different FVs with increased classification potential. Two lengthy datasets, one with 800 frames and another with ten 30-s-long CE videos, were used for the training and the evaluation of the proposed scheme, respectively. Three classification scenarios were implemented, concerning the level of detection (ROI, frame and event), that evidenced the increased effectiveness and consistency of HAD-DL scheme versus other relevant methods. The promising performance of HAF-DL approach paves the way for a holistic computer-aided diagnosis system on the service of clinicians.

7. Funding and declaration of interests: Conflict of interest: none declared.

8 References

[1] Iddan G., Meron G., Glukhovsky A., *ET AL.*: 'Wireless capsule endoscopy', *Nature*, 2000, **405**, (6785), pp. 405–417

[2] Liedlgruber M., Uhl A.: 'Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review', *IEEE Rev. Biomed. Eng.*, 2011, **4**, pp. 73–88

[3] Iakovidis D., Koulaouzidis A.: 'Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software', *Gastrointest. Endosc.*, 2014, **80**, (5), pp. 877–883

[4] Li B., Meng M.: 'Texture analysis for ulcer detection in capsule endoscopy images', *Image Vis. Comput.*, 2009, **27**, pp. 36–42

doi: 10.1049/htl.2015.0055

[5] Yu L., Yuen P.C., Lai J.: 'Ulcer detection in wireless capsule endoscopy images'. Proc. IEEE Int. Conf. on Pattern Recognition, Tsukuba Int. Congress Center, Tsukuba, Japan, November 2012, pp. 45–48
[6] Liu X., Gu J., Xie Y., *ET AL.*: 'A new approach to detecting ulcer and bleeding in wireless capsule endoscopy images'. Proc. IEEE-EMBS Int. Conf. on Biomedical and Health Informatics, Hong Kong, January 2012, pp. 737–740
[7] Charisis V., Hadjileontiadis L., Liatsos C., *ET AL.*: 'Capsule endoscopy image analysis using texture information from various colour models', *Comput. Methods Programs Biomed.*, 2012, **107**, (1), pp. 61–74
[8] Eid A., Charisis V., Hadjileontiadis L., *ET AL.*: 'A curvelet-based lacunarity approach for ulcer detection from wireless capsule endoscopy images'. Proc. IEEE Int. Symp. on Computer-Based Medical Systems, University of Porto, Porto, Portugal, June 2013, pp. 273–278
[9] Girgis H., Mitchell B., Dassopoulos T., *ET AL.*: 'An intelligent system to detect Chrohn's disease inflammation in wireless capsule endoscopy videos'. Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, Congress Center De Doelen, Rotterdam, Netherlands, April 2010, pp. 1373–1376
[10] Jebarani W., Daisy V.J.: 'Assessment of Crohn's disease lesions in wireless capsule endoscopy images using SVM based classification'. Proc. IEEE Int. Conf. on Signal Processing Image Processing & Pattern Recognition, Karunya University, Coimbatore, India, February 2013, pp. 303–307
[11] Kumar R., Zhao Q., Seshamani S., *ET AL.*: 'Assessment of Crohn's disease lesions in wireless capsule endoscopy images', *IEEE Trans. Biomed. Eng.*, 2012, **59**, (2), pp. 355–362
[12] Candes E.J., Donoho D.L.: 'Curvelets – A surprisingly effective nonadaptive representation for objects with edges', in Fitting A., Rabut C.C., and Schumaker L.L., (Eds). 'Curves and Surfaces' (Vanderbilt University Press, Nashville, TN, 1999)
[13] Mandelbrot B.B.: 'The fractal geometry of nature' (Freeman, NY, 1983)
[14] Allain C., Coitre M.: 'Characterizing the lacunarity of random deterministic fractal sets', *Phys. Rev. E*, 1991, **44**, pp. 3552–8
[15] Dong P.: 'Test of a new lacunarity estimation method for image texture analysis', *Int. J. Remote Sens.*, 2000, **21**, (17), pp. 3369–3373
[16] Tkalcic M., Tasic J.F.: 'Colour spaces: perceptual historical and applicational background'. Proc. IEEE Int. Conf. on Computer as a Tool, Ljubljana, Slovenia, September 2003, pp. 304–308
[17] Plotnick R.E., Gardner R.H., Hargrove W.W., *ET AL.*: 'Lacunarity analysis: a general technique for the analysis of spatial patterns', *Phys. Rev. A*, 1996, **53**, (5), pp. 5461–8
[18] Hadjileontiadis L.J.: 'A texture-based classification of crackles and squawks using lacunarity', *IEEE Trans. Biomed. Eng.*, 2009, **56**, (3), pp. 718–732
[19] Haralick R.M.: 'Statistical and structural approaches to texture', *Proc. IEEE*, 1979, **67**, (5), pp. 786–804
[20] Charisis V.: 'Application of advanced signal and image processing techniques in the field of digestive endoscopy', PhD thesis, 2015, Aristotle University of Thessaloniki, Greece
[21] Lewis S.: 'Expanding role of capsule endoscopy in inflammatory bowel disease', *World J. Gastroenterol.*, 2008, **14**, (26), pp. 4137–41
[22] Goldberg E., Holland H.: 'Genetic algorithms and machine learning', *Mach. Learn.*, 1988, **3**, (2), pp. 95–99
[23] Cristianini N., Shawe-Taylor J.: 'An introduction to support vector machines and other kernel-based learning methods' (Cambridge University Press, Cambridge, UK, 2000)
[24] Mankiewicz R.: 'The story of mathematics' (Princeton University Press, Princeton, NJ, 2004)