Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

In-vivo Barrett's esophagus digital pathology stage classification through feature enhancement of confocal laser endomicroscopy

Noha Ghatwary
Amr Ahmed
Enrico Grisan
Hamid Jalab
Luc Bidaut
Xujiong Ye

SPIE.

Noha Ghatwary, Amr Ahmed, Enrico Grisan, Hamid Jalab, Luc Bidaut, Xujiong Ye, "*In-vivo* Barrett's esophagus digital pathology stage classification through feature enhancement of confocal laser endomicroscopy," *J. Med. Imag.* **6**(1), 014502 (2019), doi: 10.1117/1.JMI.6.1.014502.

In-vivo Barrett's esophagus digital pathology stage classification through feature enhancement of confocal laser endomicroscopy

Noha Ghatwary,[a,b,*] Amr Ahmed,[c] Enrico Grisan,[d] Hamid Jalab,[e] Luc Bidaut,[a] and Xujiong Ye[a]
[a]University of Lincoln, Computer Science Department, Brayford Pool, Lincoln, United Kingdom
[b]Arab Academy for Science and Technology, Computer Engineering Department, Alexandria, Egypt
[c]University of Nottingham, Computer Science Department, Semenyih, Selangor, Malaysia
[d]University of Padova, Department of Information Engineering, Padova, Italy
[e]University of Malaya, Department of Computer System and Technology, Kuala Lumpur, Malaysia

Abstract. Barrett's esophagus (BE) is a premalignant condition that has an increased risk to turn into esophageal adenocarcinoma. Classification and staging of the different changes (BE in particular) in the esophageal mucosa are challenging since they have a very similar appearance. Confocal laser endomicroscopy (CLE) is one of the newest endoscopy tools that is commonly used to identify the pathology type of the suspected area of the esophageal mucosa. However, it requires a well-trained physician to classify the image obtained from CLE. An automatic stage classification of esophageal mucosa is presented. The proposed model enhances the internal features of CLE images using an image filter that combines fractional integration with differentiation. Various features are then extracted on a multiscale level, to classify the mucosal tissue into one of its four types: normal squamous (NS), gastric metaplasia (GM), intestinal metaplasia (IM or BE), and neoplasia. These sets of features are used to train two conventional classifiers: support vector machine (SVM) and random forest. The proposed method was evaluated on a dataset of 96 patients with 557 images of different histopathology types. The SVM classifier achieved the best performance with 96.05% accuracy based on a leave-one-patient-out cross-validation. Additionally, the dataset was divided into 60% training and 40% testing; the model achieved an accuracy of 93.72% for the testing data using the SVM. The presented model showed superior performance when compared with four state-of-the-art methods. Accurate classification is essential for the intestinal metaplasia grade, which most likely develops into esophageal cancer. Not only does our method come to the aid of physicians for more accurate diagnosis by acting as a second opinion, but it also acts as a training method for junior physicians who need practice in using CLE. Consequently, this work contributes to an automatic classification that facilitates early intervention and decreases samples of required biopsy.
[DOI: 10.1117/1.JMI.6.1.014502]

Keywords: Barret's esophagus; confocal laser endomicroscopy; enhancement; fractional differentiation and integration; grade classification.

Paper 18193RR received Sep. 5, 2018; accepted for publication Feb. 5, 2019; published online Mar. 5, 2019.

1 Introduction

In recent years, the rate of esophageal cancer has been increasing significantly in industrial countries.[1,2] Barrett's esophagus (BE) and gastroesophageal reflux disease are considered the primary cause of reaching malignancy of the esophagus columnar.[3,4] As shown in Fig. 1, BE is the deformation (metaplasia) of the healthy cells above the lower esophageal sphincter.[6,7] Even though it is ranked the 18th most common cancer in most countries, it is considered one of the deadliest diseases with only 18.8% chance of a 5-year survival rate.[8]

BE should be diagnosed in its early stages, and the patient has to have regular follow-ups to avoid evolving into a later stage.[9] It starts to appear when "normal squamous" (NS) epithelium transforms into a metaplastic mucosa containing gastric or intestinal mucosa.[6] Figure 2 demonstrates the transformation stages from normal cells until reaching the cancerous stage. To detect BE and classify its histopathology grade, both endoscopy and biopsy evidences are required.[11]

According to Mainz confocal Barrett's classification,[12] the transformation of the cells in the esophagus tube has a different vessel appearance and cell structure. The stages can be categorized into four histopathology grades: NS is the normal stage, where the patients have no disease, "gastric metaplasia (GM)" is the first stage of cell deformation accompanied with mucus, "intestinal metaplasia (IM)" is the main precancerous stage—often considered as proper BE—with dysplasia in the esophageal path,[13] patients who have GM can also have IM[14] and finally, "neoplasia mucosa (NPL)" is the later stage that might be cancerous.

Confocal laser endomicroscopy (CLE) is an *in-vivo* endoscopy technology, which provides an instant characterization of esophageal histology[15] and has been available for examination since 2005.[16] It captures highly magnified images of the mucosal layer in the gastrointestinal (GI) tract,[17] and it can reduce the need for biopsy by providing a virtual slide.[18] To benefit from the CLE capabilities, the physicians need to be very well trained to have the ability to differentiate between the small changes of the different pathology stages.[19]

*Address all correspondence to Noha Ghatwary, E-mail: nghatwary@lincoln.ac.uk; noha.ghatwary@aast.edu

2329-4302/2019/$25.00 © 2019 SPIE

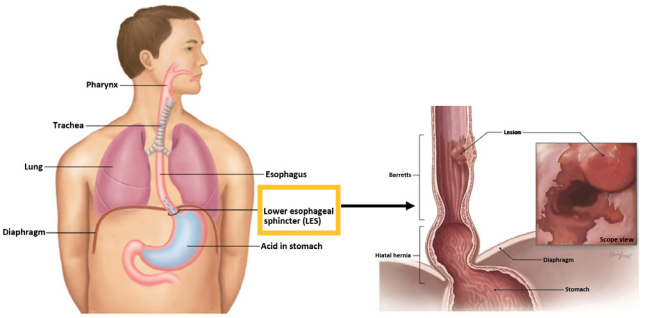

Fig. 1 Location of Barrett's inside human body and BE lesion view.[5]

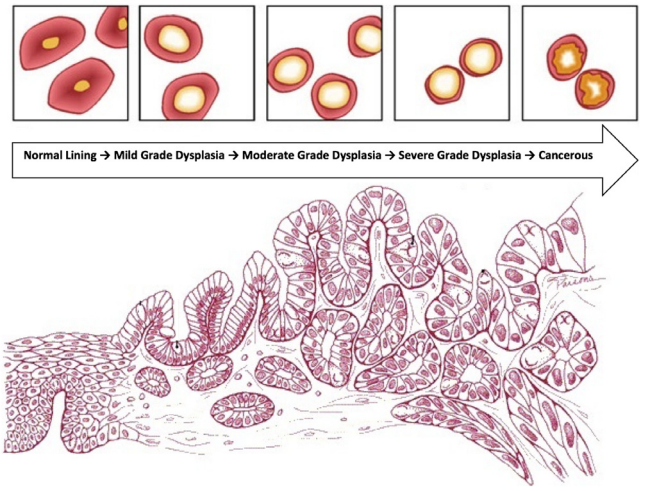

Fig. 2 Cell transformation stages from normal to dysplasia (mild, moderate, and severe) to cancer in esophagus lining.[10]

The main aim of this study is to develop a system that can automatically and accurately classify the IM (precancerous) and NPL (cancerous) stages from the other types of cell deformation in the esophagus tube as they are considered important stages in the detection of BE. The model also can serve as a second opinion for physicians and will support a decrease in the number of biopsy samples needed for each patient.

To summarize, our aim is to develop an automatic BE classification system by focusing on:

- Enhancing the CLE images by improving the feature details using a combination of "fractional integration (FI)" and differentiation for a facilitated computerized classification and an improved visualization for physicians.
- Analyzing the cell architecture and vessel properties of each stage to extract an effective combination of features for the classification process.
- Developing a unified framework that can automatically classify the captured CLE image in a real-time manner.

The paper is organized as follows: Sec. 2 gives a brief description of the previous study in the literature, followed by Sec. 3, which represents a detailed description of the system modules. Section 4 illustrates the extensive experiments compared with the state-of-the-art methods to validate the model efficiency. Finally, Sec. 5 concludes the paper by recommending the next steps.

2 Related Background

In this section, we provide a brief description of endoscopy technologies used to detect/classify BE focusing on CLE, followed by an overview of the previous studies on BE classification using CLE.

2.1 *Endoscopy Used for BE Classification*

Endoscopy is a nonsurgical process that examines the different cavities within a human body.[20] Several endoscopic technologies have been developed for a couple of decades to examine the upper GI tract and proved their efficiency in the examination process, such as standard white light endoscopy (WLE), high-definition WLE (HD-WLE), chromoendoscopy, wireless capsule endoscopy, narrow band imaging, optical coherence tomography, and CLE.

The WLE/HD-WLE endoscopes are considered the main endoscopy tools used for detection process.[21] It was reported that an appropriate diagnosis using CLE when added with the WLE/HD-WLE might lead to a less number of biopsy samples taken from the patient.[22]

CLE can give a magnified microscopy image (up to 1000×) for the subsurface of tissue analysis.[23] It is adequate for real-time examination for diagnosis to reduce the number of samples performed through traditional biopsies.[24] After injecting an intravenous contrast agent (typically fluorescein) to stain the intercellular gap to outline cell borders, a blue laser is used to scan the mucosal surface focusing at a determined depth. The CLE is applicable for clinical use to diagnose BE.[25] It is also considered to be an important field that will highlight attention for more research in the field of automatic classification.[26] More details about endoscopes used for Barrett's detection and classification can be found at Refs. 5, 27–29.

2.2 *Related Work of BE Classification Methods*

There exists an amount of research available in the literature for automatic BE stage classification using different endoscopes that are discussed in Ref. 30. However, in this review, we will be mainly focusing on methods utilizing the CLE only.

A patch-based classification method was suggested by Grisan et al.[31] to distinguish between the IM and GM grades by extracting rotation invariant local binary patterns (RLBP) and contrast features from CLE images. The result showed 98.85% sensitivity and 65.22% specificity for detecting the IM class, which was considered efficient in classifying specific regions inside the image.

Grisan et al.[32] introduced a computer diagnosis method for classification between IM, GM, and NPL with an overall accuracy of 84%. In this system, the features are extracted based on an image-based approach and processed on a two-stage classifier. In the first phase, images were classified as either NPL or not. Images from the non-NPL class, from the first stage, are passed to the next level, where they are classified as either IM or GMP, based on a proposed leakage pattern extraction. The evaluation of the classification performance was based on the leave-one-out cross-validation on 336 CLE images.

Veronese et al.[14] employed a technique to classify the different stage IM, GM, and NPL. In this model, a combination between patch-based and image-based feature extraction has been established. In the first stage, a patch-based classifier is used to extract intensity distribution values, geometric characteristics, and RLBP to find if the image is IM or not. Utilizing

a voting scheme, if the number of the positive sub-blocks in one image is higher than a certain threshold, then the images are categorized as IM if not then a studied leakage pattern method extracts a different feature from the whole image and is used for classification in next stage to GM and NPL. This model was able to achieve an overall accuracy of 96%.

In our previous work,[33] we proposed a classification model that discriminated between the IM, GM, and NPL on CLE images. The model was divided into three phases: the first phase was improving the image details by using an ad hoc filter and the second phase was extracting different features, such as intensity features, wavelet features, GLCM, fractal dimension, and fuzzy LBP. These sets of features were divided into three batches. Each feature_vector was used in one of the levels in the final multistage classification phase. The model was able to achieve overall accuracy of 90.458%, sensitivity of 0.93, and specificity of 0.86 for classifying the IM class with an average of 12.1 s for an image.

A deep learning method to identify IM, GM, and NPL was put forward by Hong et al.,[34] designing a convolutional neural network (CNN) composed of four CNN layers with two max-pooling layers in between and two fully connected networks at the end. The overall accuracy of the system based on the testing images was only 80.77%. This model has a limitation of running the model properly due to the limited size of the available dataset during the training phase even after applying augmentation. The dataset before augmentation is composed of 235 classified as 155 IM, 26 GM, and 55 NPL. The small amount of GM images trained led the proposed model to fail in categorizing any of the GM testing images. Additionally, the results of testing the network are based on a very small imbalanced data sample that consists of 26 images only (17 IM, 4 GM, and 5 NPL), which does not imply the efficiency of the proposed network.

Even though there exist studies for categorizing the BE, there are also some limitations among the literature, which still needs to be investigated to improve the performance. As a recall, BE is a type of disease that requires regular endoscopy surveillance after diagnosis to monitoring the cell deformation progress. A good automatic classification system must have a better accuracy rate compatible with the checkups that may lead to fewer biopsy samples.

To reach this, more investigation about an automated and accurate classification of the BE grades is needed. This can be done by studying the internal features of each class closely to build an automatic classification model effectively.

3 Proposed Model

Automated digital pathology classification of a CLE image is considered to be a challenging process for several reasons. Although each stage has its own histopathological characteristics, the transformation between each stage is considered visually small and difficult to identify easily. Moreover, the doctor examining the patients needs to be trained in the CLE imaging modality and is required to have background knowledge of histopathology. Figure 3 represents samples from the four different stages captured using the CLE.

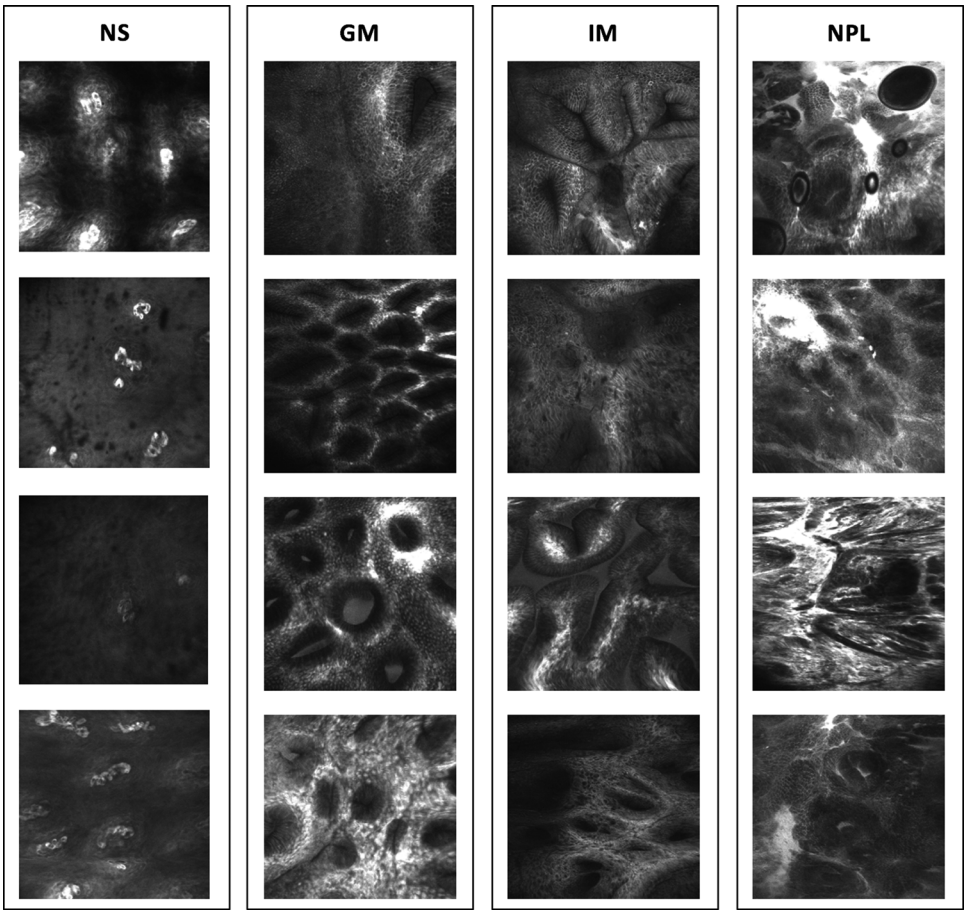

Fig. 3 Example of the CLE images from our dataset for the four types: NS, GM, IM, and NPL.

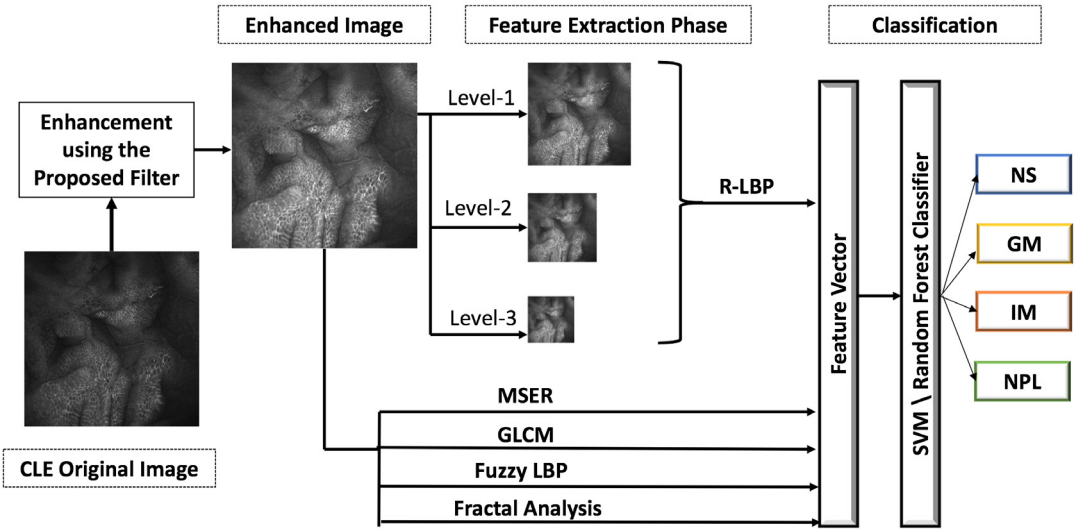

Fig. 4 The framework of the proposed classification model. The input image is first enhanced through the proposed filter. Then, different features are extracted to classify the class of the image.

The proposed approach consists of three steps, which are illustrated in Fig. 4. First, the CLE image is enhanced using a filter, and then different sets of features are extracted to discriminate between the four stages. Finally, SVM and random forest (RF) classifiers are trained separately to classify the pathology grade. Each of these steps will be described in details in the following section.

3.1 *Enhancement Phase*

In the first phase of the proposed model, the CLE image is enhanced by applying a digital filter that utilizes the fractional differentiation (FD) and FI in the wavelet sub-bands. As shown in Fig. 5, the proposed filter first decomposes the image into its discrete wavelet transform (DWT), dividing it into four sub-bands (LL, LH, HL, and HH). Then, FI is applied to the diagonal sub-bands (LH-HL) to remove the noise, while the FD is applied to the HH sub-band to improve selected texture features. The improved image is reconstructed by applying the inverse DWT, and then the FD filter is reapplied on the whole reconstructed image to improve the overall texture. In the following subsections, each phase will be explained in detail.

3.1.1 *Discrete wavelet transform*

DWT decomposes the image into four different frequency sub-bands holding the majority of the data position and emphasizing the features. These sub-bands correspond to approximate, horizontal, vertical, and diagonal features, respectively. The LL sub-band is approximately located at half the original image, while the HH sub-band contains the high-frequency details of the image. On the other hand, the HL-LH holds the changes to an image. The DWT is generally used to improve the features,[35] since it allows selective and separate suppression of coefficients in the different sub-bands, thus differently affecting low-frequency, high frequency, and directional features. We empirically chose Daubechies (db2)[36] as the mother wavelet of the DWT analysis at level 1 decomposition.

3.1.2 *Fractional differential and fractional integration*

FD and FI are mathematical operations related to the field of fractional calculus that deals with noninteger values.[37] FD has proven in the literature to provide better performance in improving the texture of images than other methods[38] while FI has shown to be an effective image noise removal method maintaining image feature details.[39]

In medical image processing, the texture is one of the key features that can improve the performance of classification. Texture can help in describing the positioning and local spatial variation of pixel intensity.[40] Applying integer-order differentiation arising from the discrete nature of the image may result in the disruption of the fine textural details that we need to capture. Therefore, using the FD is an efficient method to deal with the texture like problems. In our model, we apply the FD twice: first, it is applied to the HH sub-band to improve the high-frequency details of the image. Second, it is applied to the overall image after reconstruction from the DWT to enhance the overall texture details of the image. The FD is

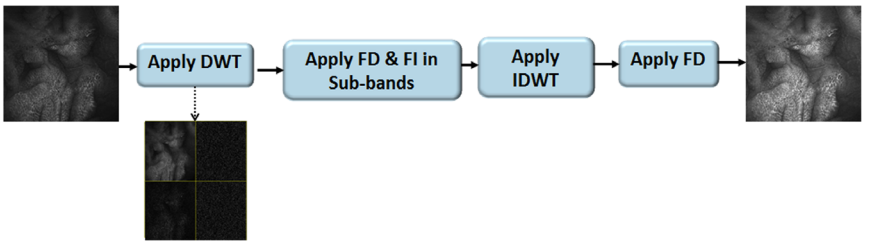

Fig. 5 Proposed enhancement filter to improve the features of the input image for accurate classification.

implemented using a mask filter based on Eq. (1) inspired from Ref. 41:

$$\frac{ds(x,y)}{dx} = -\frac{1}{[2\cos(2\Pi\alpha)h^{\alpha}]}\sum_{d=0}^{n} M_d s(x-kh,y) \quad 0<\alpha<1, \tag{1}$$

$$M_0 = -\frac{\Gamma(1-\alpha/2)}{\alpha\Gamma(1-\alpha/2)\Gamma(-\alpha)}, \tag{2}$$

$$M_d = \frac{(-1)^{d+1}\Gamma(\alpha/2)\Gamma(1-\alpha/2)}{\Gamma(\alpha/2-d+1)\Gamma(\alpha/2+d+1)\Gamma(-\alpha)} \quad d=\pm1,\pm2,\ldots, \tag{3}$$

where α is the derivative order of the FD that takes a noninteger value ranging for $0<\alpha<1$, M is the applied mask with window size of $[n*n]$ and is calculated based on Eqs. (2) and (3), d is the direction where the masked is applied.

On the other hand, denoising is important to remove the noise from the image while preserving the quality of its features. So, the FI is applied at the LH-HL sub-bands, where they hold the changes of images or edges along vertical and horizontal directions, respectively. The FI can remove noise and sustain the texture and edge features in an image.[42] In the proposed model, we utilize the FI mask suggested by Ref. 43 for the enhancement phase.

3.2 *Feature extractions*

Features are calculated based on the properties of the histopathology stage. Each grade (NS, GM, IM, and NPL) has a particular internal structure. The computation complexity is also taken into consideration as the CLE is an *in-vivo* technology; so, the automatic classification process needs to be performed in a real-time manner. An aggregation of texture and intensity features is calculated as below.

3.2.1 *Multiscale pyramid with rotation invariant LBP*

A multiscale feature named multiscale pyramid with rotation invariant local binary pattern (MP-RLBP) is proposed. It extracts the RLBP from different levels of Gaussian pyramid images. Gaussian pyramid samples the image down into smaller groups of pixels by calculating the average weight of neighbored pixels. It provides useful information for texture mapping. RLBP[44] is an extension of LBP that was first represented in Ref. 45. The RLBP is calculated by shifting the output binary code circularly according to its neighbors. It has two essential elements, P the corresponding pixel count and R the radius length. The image is first decomposed into N-levels using a Gaussian pyramid.[46] The RLBP is extracted from each scaled image, as shown in Fig. 6, to measure the relationship between a pixel and its neighbor as a descriptor. The N-level of Gaussian pyramid in the proposed model is adjusted to level-3 while the parameters of RLPB were set to $R=4$ with $P=8$.

3.2.2 *Maximally stable extremal regions*

Maximally stable extremal region (MSER) is known as a shape descriptor that was first introduced by Matas et al.[47] It can detect regions having different properties by evaluating the stability of extremal regions, which represent the high- and low-intensity regions compared to all pixels of the outer boundaries.[48] MSER is considered a fast region detector with a good performance for the homogeneous regions with distinctive boundaries in an image. It has four main parameters: [threshold (t), minimum ($\min R$) and maximum ($\max R$) size of each region and maximum stability function q], defining q as follows:

$$q(R_t) = \frac{A(R_t)}{\frac{\partial}{\partial t}A(R_t)}, \tag{4}$$

where A expresses the area of region R at threshold t. The image features detected by MSER are the stable regions that are mapped into a global high-dimensional feature vector of size 64. As previously explained, each BE grade type has a certain deformation of the cell properties based on the stage. Therefore, extracting MSER features help improve the accuracy of classification. In the proposed model, we empirically set the variables of the MSER to $t=2$, $\min R=30$, $\max R=1400$, and $q<0.25$.

3.2.3 *Gray level co-occurrence matrices*

Gray level co-occurrence matrix (GLCM) captures the second-order statistical features for texture. Fourteen textural features were defined by Haralick,[49] computing different properties to obtain texture feature details. One of the dysplasia properties is that it usually has a high entropy value. Moreover, low contrast and homogeneity of pixel pairs helps to differentiate the degree of dysplasia. For that reason, the following GLCM features (entropy, contrast, and homogeneity) are utilized in our model:

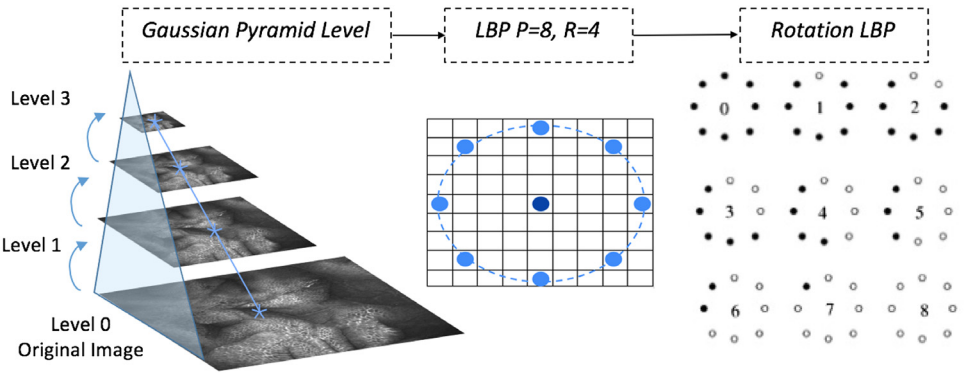

Fig. 6 Example of MP-RLBP extraction from CLE multiscale pyramid image.

$$\text{Entropy} = \sum_{i,j=0}^{N-1} -\ln(P_{ij})P_{ij}, \tag{5}$$

$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2, \tag{6}$$

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}, \tag{7}$$

where P_{ij} is the element of normalization between two pixels i and j, and N is the number of gray levels in the image.

3.2.4 *Fractal texture features*

The calculated set of features includes the fractal texture features, as presented by Ref. 50. This feature measures the fractal dimension using the box-counting method, mean gray level and size (pixel count) from set of binary images. The binary images are created using a two-threshold decomposition that characterizes the texture patterns of the CLE input image. Additionally, we compute the "Lacunarity", which measure the spatial distribution of the fractal gaps. The low lacunarity represents a homogenous texture, where all gaps represent the same size, on the other hand, high lacunarity provides heterogeneous texture. Our model will benefit from this feature to identify the gaps caused by the vessel appearance and compare between high complex details in similar stages (i.e., GM and IM).

3.2.5 *Fuzzy local binary pattern*

Fuzzy local binary pattern (FLBP) is another extension of LBP. The LBP measures the relationship between a pixel intensity and its neighboring intensities. The fuzzy logic deals with the uncertainty of the LBP and improves the textures classification by employing a set of fuzzy rules.[35] The FLBP is described in Ref. 51, where two membership functions were implemented to generate the rules in order to extract the texture descriptor. Since the cell texture representation is very challenging and, in each consecutive stage, there exists a very high similarity in their properties, a feature, such as FLBP, will then have the ability to measure texture information.

3.3 *Classification*

In this work, we evaluate two classifiers that are mostly employed by various CAD systems, briefly explained in the following section:

3.3.1 *Support vector machine*

SVM is a learning algorithm that is originally introduced by Ref. 52 and successfully extended by some researchers. In the classification process, SVM utilizes a maximal margin hyperplane to separate a set of binary labeled data. In the case of nonlinear separation, the SVM requires a kernel method (polynomial kernel, radial basis function, and hyperbolic tangent) to automatically realize a nonlinear mapping of the feature space to maximize the margin hyperplane. The chosen kernel is defined as $K(x,y)$. In the proposed model, the SVM classifier with a polynomial kernel is used. The polynomial kernel function can be expressed as follows:

$$K(x,y) = (x.y+c)^d d > 0, \tag{8}$$

where c is a parameter that trades off between the impact of the high-order parameter against the lower order ones and d represents the degree of the polynomial that relates to the sum of the supported variables.

3.3.2 *Random forest*

RF is a classification method that deploys an ensemble decision trees first introduced in Ref. 53. It is composed of a selection of tree classifiers, where each classifier randomly selects a subset of the input vector and each tree votes for the highest selected class to categorize the input. Different parameters are initialized in a RF classifier, such as depth of tree (D_tree) and random seed point (rp), to be discussed in Sec. 4.2.

4 Experimental Results and Discussion

Experiments are carried out to measure the effectiveness of the proposed automatic BE stage classification. We have compared our method to four state-of-the-art methods. The results along with quantitative analysis and discussion are presented in this section.

4.1 *Experimental Dataset*

A CLE dataset consisting of 557 images of four different histopathology grades from 96 patients was used to test the efficiency of the proposed model (IM 58 patients of 402 images, GM 10 patients of 41 images, NPL 16 patients of 68 images, NS 12 patients of 45 images). Endomicroscopy was performed by two experienced endoscopists at the European Oncological Institute (IEO, Milan, Italy) and Veneto Institute of Oncology (IOV, Padova, Italy) during routine clinical surveillance endoscopy in patients with BE, using a confocal laser endoscope (EC-3870CIFK; Pentax, Tokyo, Japan), allowing simultaneous videoendoscopy and endomicroscopy. Preparation of the patients includes conscious sedation. The confocal images were obtained after injection of 10% fluorescein sodium. The resolution of each image is 1024×1024 (corresponding to 500×500 μm) that was obtained at a scan rate of 0.8 frames/s using an optical slice thickness of 7 μm and stored digitally. The range of the z-axis was 0 to 250 μm below the surface layer.

4.2 *Evaluation Measures and Experimental Setting*

4.2.1 *Classification evaluation measures*

To evaluate the performance of the proposed model in classifying BE different stages, we employ the standard performance metrics generally adopted in medical image classification (accuracy, sensitivity, specificity, precision, and F-measure) defined as follows:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \tag{9}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{10}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}}, \tag{11}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \tag{12}$$

$$F\text{-Measure} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}, \tag{13}$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

4.2.2 *Enhancement filter evaluation measures*

To assess the performance of the enhancement filter objectively, we utilize two well-known image quality quantitative measures: contrast improvement index (CII) and Tenengrad measure.

- CII measures the improvement of the contrast between the enhanced and original image:

$$\text{CII} = \frac{A_E}{A_I} \tag{14}$$

  where A_E is the average values of local contrast C from the enhanced image and A_I from the original image. The local contrast C is calculated from a window size of 3×3 as $\frac{\text{maximum}-\text{minimum}}{\text{maximum}+\text{minimum}}$. The increase of the CII values indicates an improvement in the contrast of the enhanced image.
- Tenengrad measure is used to examine whether structural information in the enhanced image has been improved or not, therefore, it is one of the most accurate and robust measures for image quality evaluation. For each enhanced image E, the gradient $\Delta E(i,j)$ at each pixel location (i,j) is used to calculate the Tenengrad value, where the partial derivatives are acquired through a high-pass filter using Sobel operator, with the convolution kernels e_i and e_j. The gradient magnitude is defined as follows:

$$S(i,j) = \sqrt{[e_i \times E(i,j)]^2 + [e_j \times E(i,j)]^2}, \tag{15}$$

  and the Tenengrad value (T) for an image is calculated as follows:

$$T = \sum_i \sum_j S(i,j)^2 \quad S(i,j)^2 > t, \tag{16}$$

  where t is a threshold. A larger Tenengrad value implies a higher quality of an image.

4.2.3 *Experimental setup*

For implementation and experimental evaluation, the Matlab_R2016a has been used on a 2.9 GHz dual-core Intel Core i5 with 8.0 GB SDRAM. Moreover, we evaluated and tested different α values for the enhancement filter in several directions with different window size. In our study, the window size of the enhancement filter mask (M) is adjusted to 5×5 and applied in $d = 8$ directions. The α value in FD process can range from $0 < \alpha < 1$, as explained earlier, when applying FD in the HH sub-band α is set to 0.6 while when applied to the whole image, it is set to $\alpha = 0.4$.

Additionally, we compared the performance of two classification methods expressed in Sec. 3.3. The variables of the SVM polynomial kernel have been set to $c = 1$ and $d = 2$ [in Eq. (9)] while the parameters of the RF classifier were set to $D\_\text{tree} = 100$ and $rp = 1$. Furthermore, the extracted features were concatenated to form a feature vector of size (620×1) made up of MP-RLBP (286×1), MSER (64×1), GLCM (3×1), fractal texture features (11×1), FLBP (256×1) to be used for classification.

4.3 *Evaluation Results and Discussion*

In order to support the doctors with a valid second opinion, the proposed model is concerned with the accuracy of automatically classifying each histopathology grade, specifically the precancerous stage IM and later NPL stage.

The confusion matrix in Table 1 illustrates the performance of the proposed model using the leave-one-patient-out cross-validation (LOPO-CV) with the SVM classifier that showed a better performance than the RF as will be discussed. Since each patient might have more than one image, this type of validation is more efficient to measure the confidence of the model and avoid any bias classification. Based on this validation method, the model was able to achieve an overall accuracy of 96.05% with a sensitivity of 0.97 for IM, 0.90 for GM, 0.94 for NPL, and 1.00 for NS. The results show that the misclassified images are mostly classified incorrectly as a higher grade. Therefore, the system is considered better than misclassifying any true positives that need to be examined.

Moreover, experiments have been applied to the dataset without using the filter to evaluate the efficiency of the proposed enhancement filter on the classification results, we extracted the suggested features from the original image (without enhancement) and classified using both the SVM and RF to evaluate the effect of the filter and the classifiers on the proposed model performance. The results of each classifier are compared together in Table 2, and the sensitivity for each class and overall accuracy are illustrated in Fig. 7. Starting with the IM class, the SVM classifier was able to detect more IM images accurately with less false positives when compared to the RF (with or without filter) showing highest values throughout the table for all the evaluation measures. Moreover, applying the enhancement filter to the images increases the sensitivity from 95% to 97%, specificity from to 79% to 96%, precision from 94% to 96%, and F-measure from 96% to 97% when using the SVM classifier. While in the case of using the RF, results using enhancement increased the sensitivity from 86% to 90%, specificity from to 88% to 91%, precision from 95% to 96%, and F-measure from 91% to 94%.

Table 1 Proposed model confusion matrix using leave-one-patient-out cross-validation (LOPO-CV) on the 96 patients with SVM classifier.

	IM	GM	NPL	NS	Sensitivity	*F*-measure
IM	389	1	12	0	0.97	0.97
GM	3	37	1	0	0.90	0.92
NPL	3	1	64	0	0.94	0.88
NS	0	0	0	45	1.00	1.00
Specificity	0.96	0.99	0.97	1.00	Accuracy = 96.05%	
Precision	0.98	0.94	0.83	1.00		

Table 2 Evaluation of the model with and without (W/O) using the proposed enhancement filter using SVM and RF for model validation. The experiments are tested using LOPO-CV on the 96 patients.

	Image with filter using SVM classifier				Image W/O filter using SVM classifier				Image with filter using RF classifier				Image W/O filter using RF classifier			
Grade	IM	GM	NPL	NS	IM	GM	NPL	NS	IM	GM	NPL	NS	IM	GM	NPL	NS
Sensitivity	**0.97**	0.90	**0.94**	**1.00**	0.95	0.67	0.85	1.00	0.90	**1.00**	0.86	1.00	0.86	0.88	0.76	0.96
Specificity	**0.96**	**0.99**	0.97	**1.00**	0.79	0.98	0.97	0.99	0.91	0.97	**0.99**	1.00	0.88	0.95	0.92	0.99
Precision	**0.98**	**0.94**	0.83	**1.00**	0.94	0.81	**0.88**	0.97	0.96	0.67	0.55	1.00	0.95	0.62	0.60	0.91
F-Measure	**0.97**	**0.92**	**0.88**	**1.00**	0.96	0.73	0.84	0.98	0.94	0.80	0.67	1.00	0.91	0.73	0.86	0.93
Accuracy	**96.05%**				92.01%				91.00%				86.00%			

Note: The bold values represent the highest values obtained for each category based on the method.

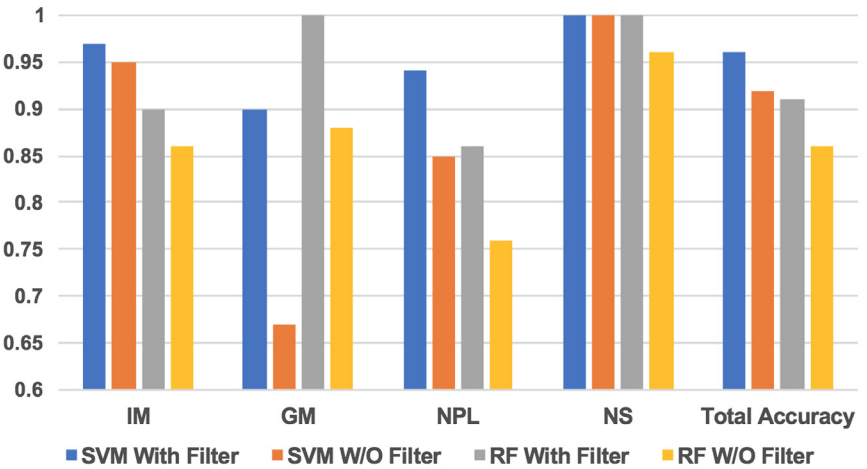

Fig. 7 Comparison between the accuracy of classifying each grade separately and the overall model with and without applying the enhancement filter to the CLE image using both the SVM and RF classifier.

Followed by the GM class, using the RF classifier in enhanced images outperformed with a result of 100% while a better result for the specificity was shown when using the SVM classifier indicating that SVM was able to decrease the number of false positives for this class. As shown in the table, using the filter for both classifiers showed a significant increase in the results of all evaluation measures.

Pursuing the results of NPL class, the sensitivity and *F*-measure with results of 94% and 88% using the SVM on enhanced images surpassed the results from RF (with and without filter) and SVM without the filter. The specificity for the RF classifier on the enhanced images was better than the other three values. On the other hand, the precision value (88%) of the SVM without the filter was the best in this case. This was the only incidence, where the experimental results for nonenhanced images showed a better performance than enhanced images throughout the table.

Finally, the NS class results were improved when using the filer for both classifiers resulting in an accuracy of 100% without allowing any other classes to be misclassified as NS. As a conclusion from this comparison, the SVM classifier on enhanced images was more efficient compared to RF for classifying the four pathology stages. Therefore, it will be used for the rest of the evaluations made in this section.

Furthermore, to evaluate the performance of the proposed enhancement filter on improving the quality of the image, we employ different quantitative measures, as discussed in the previous section. Table 3 illustrates the performance measure values obtained after applying the proposed filter in comparison with different standard enhancement techniques: histogram equalization (HE), adaptive HE, median filter, Wiener filter, and Gaussian filter. From the table, it can be seen that the proposed enhancement filter gives a higher CII value compared to the standard enhancement methods showing that the filter is able to provide better contrast within the image. Additionally, throughout the table, the Tenegard value outperforms against the other conventional filters, therefore, we can conclude that the structural information has been improved, which leads to an improved classification result. In addition to the quantitative evaluation results, we also demonstrate some qualitative results in Fig. 8 that represent an example of different samples from CLE images before and after applying the enhancement filter.

Additional experiments are tested by evaluating the model on an individual dataset. The patients' images were split into 60% training and 40% testing. As shown in Table 4, the model was able to maintain high performance by achieving an overall 93.72%, misclassifying 5 IM as NPL, 4 GM as IM, 3 GM as NPL, and 2 NPL as IM.

Table 3 Performance measure values obtained after applying different enhancement techniques on the CLE image.

	Proposed filter	HE	Adaptive HE	Median	Wiener	Gaussian
CII	**3.283**	2.458	2.801	0.648	0.656	0.827
Tenengrad ($\times 10^3$)	**14.712**	13.572	11.903	13.340	11.883	12.849

Note: The bold values represent the highest values obtained for each category based on the method.

Table 4 Confusion matrix of the proposed model on an individual dataset. The training set of 60% (58 patients) and testing set of 40% (33 patients).

	IM	GM	NPL	NS	Sensitivity	*F*-measure
IM	157	0	5	0	0.96	0.97
GM	4	11	3	0	0.61	0.78
NPL	2	0	19	0	0.90	0.79
NS	0	0	0	22	1.00	1.00
Specificity	0.90	1.00	0.96	1.00	Accuracy = 93.72%	
Precision	0.96	1.00	0.70	1.00		

As a further study, a comparison of the results for the presented model with other state-of-the-art models is demonstrated in Table 5. We employ the publicly available dataset provided by the ISBI'16 challenge[54] used by both Ghatwary et al.[33] and the deep learning method by Hong et al.[34] By comparing the proposed model with Ghatwary et al.,[33] our model surpassed the overall accuracy by 7%. Moreover, by evaluating each class separately, a significant improvement was observed in both sensitivity and specificity for the three categories. For the method proposed by Hong et al.,[34] we could not compare the results of our model with it. The results illustrated by Hong et al.[34] in Table 5 were based on only a total of 26 images (a small subset from the dataset provided by ISBI'16 challenge[54]) from 262 images that are used by our model and Ghatwary et al.;[33] therefore, it was going to be an unfair comparison. Meanwhile, when evaluating the results by Hong et al.,[34] the accuracy showed a low performance of 80.77%. Moreover, the results indicate that images from GM and NPL were misclassified as IM.

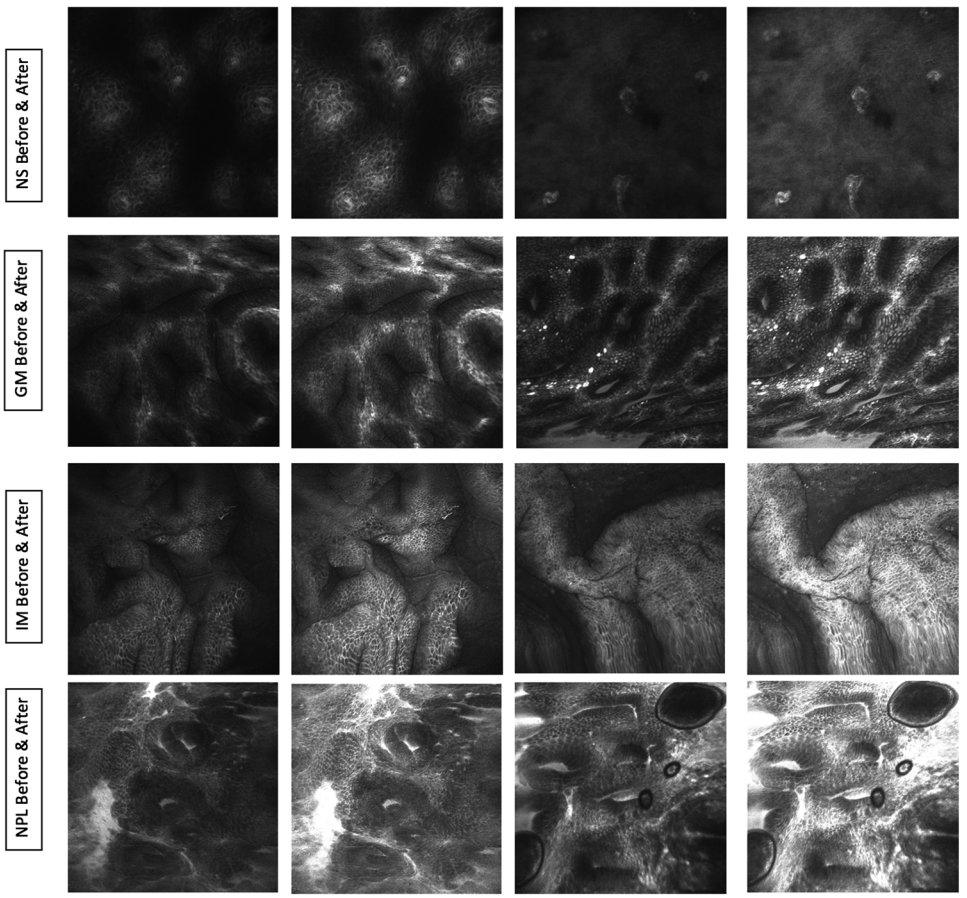

Fig. 8 Example of different sample of CLE images before and after using the enhancement filter.

Table 5 Comparison between proposed model, Ghatwary et al.,[33] and Hong et al.[34] using leave-one-image-out cross-validation approach on 262 images of different stages.

	Proposed model	Ghatwary et al.[33]	Hong et al.[34]
		Total accuracy	
	97.71%	90.46%	80.77%
		Sensitivity	
IM	0.98	0.94	**1.00**
GM	**0.83**	0.70	0.00
NPL	**0.97**	0.90	0.80
		Specificity	
IM	**0.93**	0.88	0.44
GM	**1.00**	0.96	1.00
NPL	0.96	0.97	**1.00**

Note: The bold values represent the highest values obtained for each category based on the method.

Another comparison assessment is shown in Table 6 to illustrate the evaluation of the proposed model against the most recent state-of-the-art methods Veronese et al.[14] and Grisan et al.,[32] using the same dataset of 337 images with three different classes only (GM, IM, and NPL). Moreover, using the same evaluation method of a leave-one-image-out cross-validation. As shown, the proposed model exceeded the overall accuracy by 2.97% and 18.16%, respectively. Also, each class was evaluated separately, beginning with the IM class—the main precancerous stage—which is considered the primary target, since its detection through the classification stage is critical to the therapeutic plan. By evaluating the sensitivity, the proposed model not only surpassed[14] by 4% but also outperformed[32] by 22%. However, the specificity of the proposed model falls short by 1% compared to Ref. 14. The main reason behind this fall is that one image from NPL class was classified as IM. On the other hand, in Ref. 14, the IM was misclassified as another class; hence, their IM specificity was not affected.

Table 6 Comparison between proposed model, Veronese et al.,[14] and Grisan et al.[32] using leave-one-image-out cross-validation approach on 337 images of different stages.

	Proposed model	Veronese et al.[14]	Grisan et al.[32]
		Total accuracy	
	99.11%	96.14%	80.95%
		Sensitivity	
IM	**0.99**	0.95	0.77
GM	**1.00**	0.96	0.78
NPL	0.98	**1.00**	1.00
		Specificity	
IM	0.99	**1.00**	0.97
GM	**1.00**	0.99	0.94
NPL	**0.99**	0.96	0.84

Note: The bold values represent the highest values obtained for each category based on the method.

GM class is the smallest dataset among the three categories in this experimental evaluation. Thus, misclassification of an image leads to an obvious impact on the results. Both sensitivity and specificity of the current model were able to maintain the highest performance by correctly classifying all the GM images with no false positives while Refs. 14 and 32 missed 5 and 1 images, respectively.

Finally, by evaluating the sensitivity of NPL stage, both Refs. 14 and 32 achieved a 100% for this class. However, the proposed model did not experience a significant decline as only a single image was misclassified. On the other hand, the specificity of our model outperformed[14,32] indicating the improvement in decreasing the classification of the other two classes as NPL.

Further investigation has been made since the CLE is an *in-vivo* process, it requires the classification to be done on a real-time basis. Therefore, it is essential to take into account the computation time. The execution time for each phase (image enhancement and feature extraction) was measured separately. The average processing time for enhancement of an image required 2 to 3 s while the feature extraction process required 1.9 to 2.5 s per image. Therefore, the total average processing time required by the proposed model to classify an image is an average of 3.9 to 6.5 s.

Moreover, in Table 7, we compare the computation time required to classify a single image using our method with other state-of-the-art methods. As shown, the proposed model was able to classify the stage of the abnormality in less time than the other methods. The average time of the model was faster than Ghatwary et al.[33] by 7.8 s, Veronese et al.[14] by 2.95 s, and Grisan et al.[32] by 3.3 s.

One of the main reasons that the methods proposed by Ghatwary et al.,[33] Veronese et al.,[14] and Grisan et al.[32] take more time is that they are multistage classification models, where the preprocessing, feature extraction, and classification is done on several stages based on the abnormality type. Additionally, our models (Refs. 32 and 33) are image-based feature extraction systems. Therefore, the processing time toward the feature extraction phase can be considered similar. However, the model in Ref. 14 was divided into two phases: a patch-based phase and an image-based phase, which requires more time to divide a single image into patches and select the suitable ones for feature extraction.

The time by the proposed model is considered reasonable and convenient for the examination process since the mean

Table 7 Comparison of the computation time (in seconds) between proposed model, Ghatwary et al.,[33] Veronese et al.,[14] and Grisan et al.[32] for image classification.

Proposed model	Ghatwary et al.[33]	Veronese et al.[14]	Grisan et al.[32]
3.9 to 6.5	9 to 17	7.1 to 9.2	6.7 to 10.3

inspection time of the CLE is around 22 min. A patient needs between 9 and 45 min to be examined, and the CLE image is captured with a rate of 0.8 frames/s at a resolution of 1024×1024.[12]

5 Conclusion

This paper presents an automatic classification for Barrett's esophagus stages focusing on improving the classification of IM category to support the physician's opinion. This will help decrease the required biopsy samples and monitor the dysplasia before turning into cancer. Preprocessing steps are first applied to enrich the CLE input image for feature extraction. The enhancement is a vital part of the contribution to the proposed system, carried out by implementing a filter that employs the FI and FD in the sub-bands of the DWT. Subsequently, the FD is applied to the whole image after it regains its original form. Afterward, a multiscale feature MP-LBP, GLCM, fractal analysis, fuzzy LBP, and MSER are calculated and fed into a classifier based on SVM (polynomial kernel). The proposed system shows promising results with an overall accuracy of 96.05% when it is validated on a leave-one-out patient validation. Moreover, the model achieved 93.7% overall accuracy when splitting the dataset into 60% training and 40% testing. In conclusion, higher accuracy results are obtained—particularly in the IM class (97%)—represented by the proposed automatic Barrett's classification model when using the SVM classifier on enhanced images. In future studies, we will focus on developing a complete CAD system that supports the full diagnosis of BE, by automatically detecting and classifying the abnormalities related to BE in the esophagus.

Disclosures

The authors declare that they have no conflict of interest with regard to the work presented.

Acknowledgments

We want to thank European Oncological Institute (IEO, Milan, Italy) and Veneto Institute of Oncology (IOV, Padova, Italy) for providing the dataset used in this work.

References

1. H. G. Coleman et al., "Increasing incidence of Barrett's oesophagus: a population-based study," *Eur. J. Epidemiol.* **26**(9), 739–745 (2011).
2. F. van der Sommen et al., "Computer-aided detection of early neoplastic lesions in Barrett's esophagus," *Endoscopy* **48**(7), 617–624 (2016).
3. S. Rajendra and P. Sharma, "Barrett esophagus and intramucosal esophageal adenocarcinoma," *Hematol. Oncol. Clin. North Am.* **31**(3), 409–426 (2017).
4. J. Flejou, "Barrett's oesophagus: from metaplasia to dysplasia and cancer," *Gut* **54**(suppl 1), i6–i12 (2005).
5. B. V. Naini et al., "Barrett's oesophagus diagnostic criteria: endoscopy and histology," *Best Pract. Res. Clin. Gastroenterol.* **29**(1), 77–96 (2015).
6. H. G. Coleman et al., "Symptoms and endoscopic features at Barrett's esophagus diagnosis: implications for neoplastic progression risk," *Am. J. Gastroenterol.* **109**(4), 527–534 (2014).
7. D. Boschetto, G. Gambaretto, and E. Grisan, "Automatic classification of endoscopic images for premalignant conditions of the esophagus," *Proc. SPIE* **9788**, 978808 (2016).
8. "Cancer of the esophagus - cancer stat facts," https://seer.cancer.gov/statfacts/html/esoph.html (2018).
9. E. Bird-Lieberman and R. Fitzgerald, "Early diagnosis of oesophageal cancer," *Br. J. Cancer* **101**(1), 1–6 (2009).
10. Johns Hopkins Department of Pathology, "Barrett's esophagus," http://pathology.jhu.edu/beweb/understanding.cfm.
11. R. E. Sampliner, "Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus," *Am. J. Gastroenterol.* **97**(8), 1888–1895 (2002).
12. R. Kiesslich et al., "In vivo histology of Barrett's esophagus and associated neoplasia by confocal laser endomicroscopy," *Clin. Gastroenterol. Hepatol.* **4**(8), 979–987 (2006).
13. K. K. Wang and R. E. Sampliner, "Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus," *Am. J. Gastroenterol.* **103**(3), 788–797 (2008).
14. E. Veronese et al., "Hybrid patch-based and image-wide classification of confocal laser endomicroscopy images in Barrett's esophagus surveillance," in *IEEE 10th Int. Symp. Biomed. Imag. (ISBI)*, IEEE, pp. 362–365 (2013).
15. P. Sharma et al., "White paper Aga: advanced imaging in Barrett's esophagus," *Clin. Gastroenterol. Hepatol.* **13**(13), 2209–2218 (2015).
16. R. Kiesslich et al., "Confocal laser endoscopy for diagnosing intraepithelial neoplasias and colorectal cancer in vivo," *Gastroenterology* **127**(3), 706–713 (2004).
17. R. Kiesslich et al., "Confocal laser endomicroscopy," *Gastrointest. Endoscopy Clin. North Am.* **15**(4), 715–731 (2005).
18. M. Wallace et al., "Miami classification for probe-based confocal laser endomicroscopy," *Endoscopy* **43**(10), 882–891 (2011).
19. P. Rajan et al., "Automated diagnosis of Barrett's esophagus with endoscopic images," in *World Congr. Med. Phys. Biomed. Eng.*, Munich, Germany, Springer, pp. 2189–2192 (2009).
20. T. Gotoda, "Endoscopic resection of early gastric cancer," *Gastric Cancer* **10**(1), 1–11 (2007).
21. N. Ghatwary, A. Ahmed, and X. Ye, "Automated detection of Barrett's esophagus using endoscopic images: a survey," in *Annu. Conf. Med. Image Understanding and Anal.*, Springer, pp. 897–908 (2017).
22. M. I. Canto et al., "*In vivo* endomicroscopy improves detection of Barrett's esophagus–related neoplasia: a multicenter international randomized controlled trial (with video)," *Gastrointest. Endoscopy* **79**(2), 211–221 (2014).
23. M. Aubreville et al., "Patch-based carcinoma detection on confocal laser endomicroscopy images - a cross-site robustness assessment," arXiv:1707.08149 (2017).
24. N. D. Parikh et al., "Confocal laser endomicroscopy features of sessile serrated adenomas/polyps," *U. Eur. Gastroenterol. J.* **4**(4), 599–603 (2016).
25. S. Beg and K. Ragunath, "Image-enhanced endoscopy technology in the gastrointestinal tract: what is available?" *Best Pract. Res. Clin. Gastroenterol.* **29**(4), 627–638 (2015).
26. M. Liedlgruber and A. Uhl, "Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review," *IEEE Rev. Biomed. Eng.* **4**, 73–88 (2011).
27. R. S. Gill and R. Singh, "Endoscopic imaging in Barrett's esophagus: current practice and future applications," *Ann. Gastroenterol.* **25**(2), 89–95 (2012).
28. M. P. Clermont, A. M. Gamboa, and F. F. Willingham, "The role of endoscopy in the diagnosis, staging, and management of esophageal cancer," in *Esophageal Cancer*, pp. 123–148, Springer, Cham (2015).
29. R. Singh and S. P. Yeap, "Endoscopic imaging in Barrett's esophagus," *Expert Rev. Gastroenterol. Hepatol.* **9**(4), 475–485 (2015).
30. L. A. de Souza et al., "A survey on Barrett's esophagus analysis using machine learning," *Comput. Biol. Med.* **96**, 203–213 (2018).
31. E. Grisan et al., "239 computer aided diagnosis of Barrett's esophagus using confocal laser endomicroscopy: preliminary data," *Gastrointest. Endoscopy* **75**(4), AB126 (2012).
32. E. Grisan, E. Veronese, and G. Diamantis, "Computer aided diagnosis of Barrett's esophagus using confocal laser endomicroscopy: premlinary data," *Dig. Liver Dis.* **44**, S147–S148 (2012).
33. N. Ghatwary et al., "Automatic grade classification of Barrett's esophagus through feature enhancement," *Proc. SPIE* **10134**, 1013433 (2017).
34. J. Hong, B.-Y. Park, and H. Park, "Convolutional neural network classifier for distinguishing Barrett's esophagus and neoplasia endomicroscopy images," in *39th Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc. (EMBC)*, IEEE, pp. 2892–2895 (2017).

35. S. M. Youssef, A. A. ElFarag, and N. M. Ghatwary, "Adaptive video watermarking integrating a fuzzy wavelet-based human visual system perceptual model," *Multimedia Tools Appl.* **73**(3), 1545–1573 (2014).
36. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pennsylvania (1992).
37. N. Ghatwary, A. Ahmed, and H. Jalab, "Liver CT enhancement using fractional differentiation and integration," in *Int. Conf. Signal and Image Eng.*, Vol. **1** (2016).
38. Y.-F. Pu, J.-L. Zhou, and X. Yuan, "Fractional differential mask: a fractional differential-based approach for multiscale texture enhancement," *IEEE Trans. Image Process.* **19**(2), 491–511 (2010).
39. H. A. Jalab and R. W. Ibrahim, "Denoising algorithm based on generalized fractional integral operator with two parameters," *Discrete Dyn. Nat. Soc.* **2012**, 1–14 (2012).
40. G. Castellano et al., "Texture analysis of medical images," *Clin. Radiol.* **59**(12), 1061–1069 (2004).
41. Q. Yu et al., "The use of a riesz fractional differential-based approach for texture enhancement in image processing," *ANZIAM J.* **54**, 590–607 (2013).
42. H. Guo et al., "Image denoising using fractional integral," in *IEEE Int. Conf. Comput. Sci. and Autom. Eng. (CSAE)*, IEEE, Vol. **2**, pp. 107–112 (2012).
43. H. A. Jalab and R. W. Ibrahim, "Texture enhancement based on the Savitzky-Golay fractional differential operator," *Math. Prob. Eng.* **2013**, 1–8 (2013).
44. R. Mehta and K. Egiazarian, "Rotated local binary pattern (RLBP): rotation invariant texture descriptor," in *Int. Conf. Pattern Recognit. Appl. and Methods*, Institute of Electrical and Electronics Engineers IEEE, pp. 497–502 (2013).
45. T. Mäenpää, "The local binary pattern approach to texture analysis: extenxions and applications," Dissertation, Oulun Yliopisto (2003).
46. E. H. Adelson et al., "Pyramid methods in image processing," *RCA Eng.* **29**(6), 33–41 (1984).
47. J. Matas et al., "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vision Comput.* **22**(10), 761–767 (2004).
48. M. Li and Z. Yin, "Cell segmentation using stable extremal regions in multi-exposure microscopy images," in *IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, IEEE, pp. 526–530 (2016).
49. R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE* **67**(5), 786–804 (1979).
50. A. F. Costa, G. Humpire-Mamani, and A. J. M. Traina, "An efficient algorithm for fractal analysis of textures," in *25th SIBGRAPI Conf. Graphics, Patterns and Images (SIBGRAPI)*, IEEE, pp. 39–46 (2012).
51. D. K. Iakovidis, E. G. Keramidas, and D. Maroulis, "Fuzzy local binary patterns for ultrasound texture characterization," *Lect. Notes Comput. Sci.* **5112**, 750–759 (2008).
52. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**, 273–297 (1995).
53. T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Doc. Anal. and Recognit.*, IEEE, Vol. **1**, pp. 278–282 (1995).
54. "aidasub-clebarrett - home," https://aidasub-clebarrett.grand-challenge.org (2015).

Noha Ghatwary is currently pursuing her PhD in computer science in the Department of Computer Science, University of Lincoln, United Kingdom. She is currently a teaching assistant in the Computer Engineering Department, Arab Academy for Science and Technology, Alexandria, Egypt. Her research interests include medical image analysis, computer vision and machine learning.

Amr Ahmed is an associate professor of computer vision and image analysis in the School of Computer Science. Before joining the University of Nottingham, he held several research and academic positions (since 2005) in various UK universities (including Surrey and Lincoln). He also worked in industry for several years, including at Sharp Labs of Europe (SLE), Oxford United Kingdom, as a research scientist, and other engineering consultants companies abroad.

Enrico Grisan graduated in electrical engineering at the University of Padova in 2001, and he was then research consultant for the University of Padova and Nidek Technologies. In 2005, he earned a PhD in bioengineering jointly from the University of Padova and City University London. After being an intern at Siemens Corporate Research in 2005 and then postdoc fellow in Padova from 2005 to 2008, he has been appointed (tenured) as an assistant professor in bioengineering since 2008. His main research activities involve the automatic analysis of medical images and the discovery of clinical biomarkers from them. He has been working on retinal images for the quantification or retinopathy-related changes, on confocal endomicroscopy images for virtual histology, on multispectral MRI for quantifying cortical lesion burden, on perfusion patterns in contrast-enhanced ultrasound, and finally on prenatal ultrasound.

Hamid Jalab received his bachelor's degree in electrical engineering from the University of Technology, Baghdad, Iraq, and MSc and PhD degrees in computer systems from Odessa National Polytechnic University. He is currently a senior lecturer (research track) at the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. His current research interests include digital video processing, digital image processing, digital image forensics, image retrieval, information security, and artificial neural networks.

Luc Bidaut has worked with and on most aspects of biomedical imaging and technology in highly multidisciplinary research, clinical and translational international environments, always in direct collaboration with all relevant stakeholders from scientific, technical and medical disciplines. His active involvement in the development, implementation and actual deployment of related technologies and applications was and remains primarily focused on maximizing the utility and actionability of the information collected through imaging modalities and other sensors, at various stages of the translational pipeline or clinical workflow.

Xujiong Ye is a professor of medical imaging and computer vision in the School of Computer Science, University of Lincoln, UK. She has over 20 years of research and development experiences in medical imaging and computer vision from both academia and industry. Her main research is to develop computational models using advanced medical image analysis, computer vision and artificial intelligence to support clinicians in decision-making.