Software for Automated Classification of probe-based Confocal Laser Endomicroscopy Videos of Colorectal Polyps

Barbara André, Tom Vercauteren, Anna M. Buchner, Murli Krishna, Nicholas Ayache, Michael B. Wallace

Title:

Software for Automated Classification of probe-based Confocal Laser Endomicroscopy Videos of Colorectal Polyps

Running title:

Automated pCLE Classification of Colonic Polyps

Authorship:

The authors meet the following conditions: (1) substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; and (3) final approval of the version to be published.

Institution:

- Barbara André, Image Computing Group, Mauna Kea Technologies, 75010 Paris, France;
- Barbara André, Asclepios Research Team, INRIA Sophia Antipolis, 06902 Sophia Antipolis, France;
- Tom Vercauteren, Image Computing Group, Mauna Kea Technologies, 75010 Paris, France;
- Anna M. Buchner, Division of Gastroenterology, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, United States;
- Murli Krishna, Laboratory Medicine and Pathology, Mayo Clinic Hospital in Jacksonville, Jacksonville, FL 3224, United States;
- Nicholas Ayache, Asclepios Research Team, INRIA Sophia Antipolis, 06902 Sophia Antipolis, France;
- Michael B. Wallace, Gastroenterology & Hepatology, Mayo Clinic Hospital in Jacksonville, Jacksonville, FL 3224, United States.

Author contributions:

- André B was involved in study concept and design, analysis and interpretation of data, drafting of the manuscript and statistical analysis;
- Vercauteren T was involved in study concept and design, analysis and interpretation of data and study supervision;
- Buchner AM was involved in acquisition of data and analysis and interpretation of data;
- Krishna M was involved in analysis and interpretation of data;
- Ayache N was involved in study concept and design, analysis and interpretation of data and study supervision;
- Wallace MB was involved in acquisition of data, analysis and interpretation of data and study supervision.

Supportive foundations:

The present study uses data from the following clinical trial: NCT 00874263.

Correspondence to:

Barbara André, PhD Student, Image Computing Group, Mauna Kea Technologies, 9 rue d'Enghien, 75010 Paris, France. barbara.andre@maunakeatech.com

Telephone and fax:

- André B, Telephone: +33-1-70080961, Fax: +33-1-48241218
- Vercauteren T, Telephone: +33-1-48241129, Fax: +33-1-48241218
- Buchner AM, Telephone: +1- 215-6151677, Fax: +1- 215-6151688
- Krishna M, Telephone: +1-904-9532000, Fax: +1-904-9537329
- Ayache N, Telephone: +33-4-92387661, Fax: +33-4-92387669
- Wallace MB, Telephone: +1-904-9537382, Fax: +1-904-9537329

ABSTRACT

AIM

To support probe-based confocal laser endomicroscopy (pCLE) diagnosis by designing a software for automated classification of colonic polyps.

MATERIALS AND METHODS

Intravenous fluorescein pCLE imaging of colorectal lesions was performed on patients undergoing screening and surveillance colonoscopies, followed by polypectomies. All resected specimens were reviewed by a reference gastrointestinal pathologist blinded to pCLE information. Histopathology was used as criterion standard for the differentiation between neoplastic and non-neoplastic lesions. The pCLE video sequences, recorded for each polyp, were analyzed off-line by 2 expert endoscopists who were blinded to the endoscopic characteristics and histopathology. These pCLE videos, along with their histopathology diagnosis, were used to train the automated classification software which is a Content-Based Image Retrieval (CBIR) technique followed by *k*-nearest neighbor classification. The performances of the off-line diagnosis of pCLE videos established by the 2 expert endoscopists were compared with those of automated pCLE software classification. All evaluations were performed using leave-one-patient-out cross-validation to avoid bias.

RESULTS

135 colorectal lesions were imaged in 71 patients. Based on histopathology, 93 of these 135 lesions were neoplastic and 42 were non-neoplastic. The study finds no statistical significance for the difference between the performance of automated pCLE software classification (accuracy 89.6%, sensitivity 92.5%, specificity 83.3%, using leave-one-patient-out cross-validation) and the performance of the off-line diagnosis of pCLE videos established by the 2 expert endoscopists (accuracy 89.6%, sensitivity 91.4%, specificity 85.7%). There is very low power (< 6%) to

detect the observed differences. The 95% confidence intervals for equivalence testing are: −0.073 to 0.073 for the accuracy, −0.068 to 0.089 for the sensitivity and −0.18 to 0.13 for the specificity. Besides, the classification software proposed in this study is not a “black box” but an informative tool based on the query by example model that produces, as intermediate results, visually similar annotated videos that are directly interpretable by the endoscopist.

CONCLUSION

The proposed software for automated classification of pCLE videos of colonic polyps achieves high performance, comparable to that of off-line diagnosis of pCLE videos established by expert endoscopists.

KEYWORDS

Colorectal neoplasia; probe-based Confocal Laser Endomicroscopy (pCLE); Computer-aided diagnosis; Nearest neighbor classification software; Content-based image retrieval.

1. INTRODUCTION

Colorectal cancer is the second leading cause of cancer-related death in the United States[1]. Its development includes several morphological stages, from benign to adenomatous polyps with low grade dysplasia to adenocarcinoma. Suspicious lesions are usually detected with standard colonoscopy by the endoscopists who either perform confirmatory biopsy, or if high certainty exists, perform immediate therapy such as resection or ablation of diseased tissue. Because standard endoscopic imaging can only diagnose disease states with moderate levels of certainty[2,3], histopathology remains the criterion standard for final diagnosis[4]. However, the requirement for *ex vivo* histology implies a large proportion of unnecessary polypectomies and often requires a separate endoscopic procedure to be performed for treatment. It also increases the cost of colorectal cancer screening.

Probe-based Confocal Laser Endomicroscopy (pCLE, Mauna Kea Technologies, France) enables the endoscopist to image the epithelial tissue *in vivo*, at the microscopic level with a confocal miniprobe, and in real-time during ongoing endoscopy. Preliminary findings by Meining *et al.* [5] demonstrated the applicability of pCLE in diagnosing colorectal neoplasia *in vivo* with high sensitivity and specificity (93% and 92% respectively) in 13 patients with colorectal lesions. Venkatesh *et al.*[6] and De Palma[7] pointed out that confocal endomicroscopy offers the ability to target biopsies much more precisely and thus to reduce the number of random biopsies. In a recent study including a large pool of 75 patients, Buchner *et al.*[8] compared off-line diagnosis of pCLE videos to virtual chromoendoscopy (Narrow-Band Imaging and Fujinon Intelligent Color Enhancement) and showed that off-line diagnosis of pCLE videos had higher sensitivity (91% versus 77%) with similar specificity (76%). As noted by Wallace and Fockens[9], endoscopists now have the challenging task to perform “optical biopsies” and diagnose pCLE video sequences *in vivo*.

In order to provide an objective support for pCLE diagnosis, we aim at designing a computer-based system for the automated classification of colonic polyps into neoplastic and non-neoplastic lesions. As the physicians typically rely on similarity-based reasoning to establish a diagnosis from image queries, we propose a Content-Based Image Retrieval (CBIR) approach to automatically estimate the pathology of a new pCLE video. Indeed, contrary to “black box” classification systems, a CBIR-based classification system extracts, from a training database, annotated pCLE videos that are visually similar to the video of interest and directly interpretable by the endoscopist. The pathology of the video query is then estimated from the histopathological votes of these already diagnosed videos. Another advantage of CBIR-based classification is that the extracted similar videos can be presented to the endoscopist in a second reader paradigm to better support pCLE diagnosis.

The main goal of this study is to compare, using the same database of colonic polyps, the clinical performances of our automated pCLE classification software with those of off-line diagnosis of pCLE videos established by endoscopists expert in pCLE, with histopathology remaining the criterion standard reference.

2. MATERIALS AND METHODS

2.1. Patients

The patients included in the study were enrolled between November 2007 and March 2009 for previous studies approved by Mayo Clinic Institutional Review Board, and from which we collected all available data to ensure an as large as possible sample size. These patients were enrolled for the study of Buchner *et al.*[8] and for further studies of the same Mayo Clinic group. Only the patients with complete diagnostic data are considered in our study. All study participants gave full written consent. Patients were enrolled if they were due for surveillance or screening colonoscopies, evaluation of known or suspected polyps on other imaging modalities, and endoscopic mucosal resection of larger flat colorectal neoplasia. Exclusion criteria were patients with non corrected coagulopathy, women who were pregnant or breast feeding, documented allergy to fluorescein, and patients with no colorectal lesions found during a study colonoscopy. Twenty-four hours before the procedure, patients were prepped with 2 - 4 L polyethylene glycol solution. Conscious sedation was performed with intravenous administration of midazolam and meperidine.

2.2. Endoscopy Equipment and Procedure

All procedures were performed by either MBW or AMB using a high-definition colonoscope (Fujinon EC450HL5 or 490 ZW, Fujinon, Ft Wayne, NJ; Olympus CFH180, Olympus, Center Valley, NY). The system was equipped with the EPX 4400 processor (Fujinon Inc) or CV 180 Exera (Olympus, Co). The primary screening method was white-light high-definition colonoscopy. Then, either Fujinon Intelligent Color Enhancement mode 4 with Fujinon colonoscope or Narrow-Band Imaging with Olympus 180 series scope was used to characterize lesions in all patients.

The surface pit pattern of the lesion was classified according to Kudo criteria. Anatomical site and morphological class of lesions were recorded in accordance with the Paris classification[10]. Fluorescein sodium 2.5 - 5.0 mL 10% (AK Fluor, Akorn Pharmaceutical, Lake Forest, IL) solution was administered intravenously after the first polyp was identified. Immediately after fluorescein injection, pCLE video sequences of the lesions were acquired and recorded. According to the visual examination of both endoscopic and pCLE images, real biopsies were targeted to the most suspicious parts of the polyp. Appropriate treatment procedures, ranging from simple polypectomies to complex endoscopic mucosal resection of lesions, were then performed.

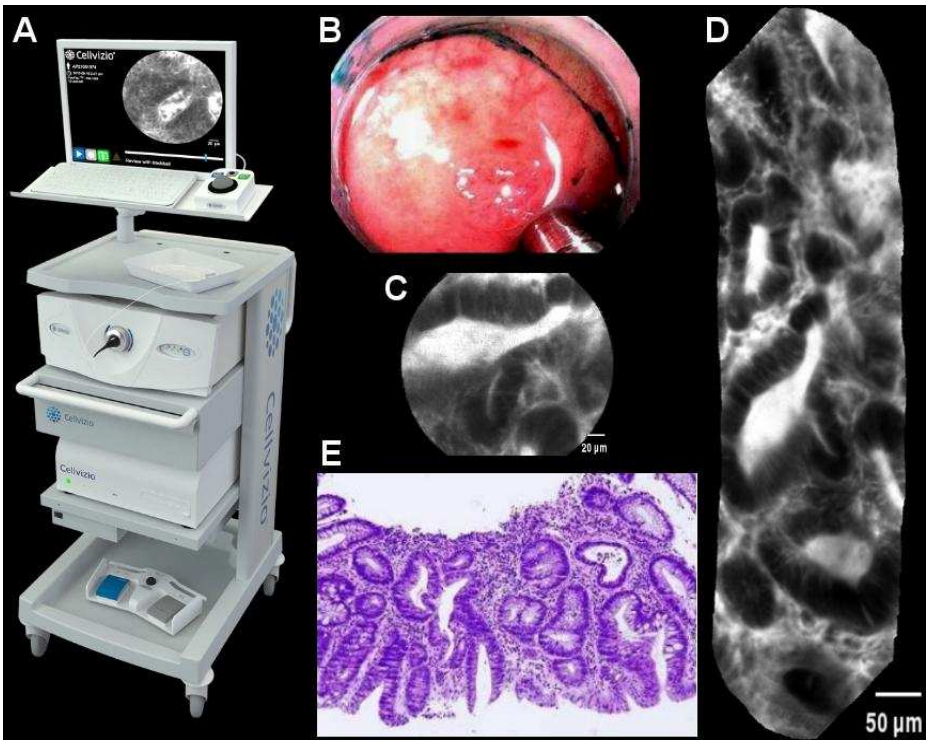

Figure 1: Imaging modalities of the colonic polyps. (A) Setup of pCLE imaging system (Cellvizio, Mauna Kea Technologies). (B) Endoscopic image of tubular adenoma, and the pCLE miniprobe. (C) An image of the pCLE video sequence. (D) A pCLE mosaic image built with the video mosaicing tool. (E) Histopathology image.

2.3. pCLE Acquisition Protocol

During a pCLE acquisition protocol, the endoscopist typically inserts, through the working channel of a standard endoscope, a confocal miniprobe (Coloflex UHD, Cellvizio GI) of external diameter 2.5 mm, which is made of 30, 000 optical fibers bundled together. The pCLE imaging setup, shown in Figure 1., allows to acquire pCLE images of field-of-view 240 µm at a rate of 9 to 12 frames per second. In stable pCLE video sequences the probe is in constant contact with the tissue. Representative endoscopic, pCLE, and histopathology images of tubular adenoma are shown in Figure 1.

Prior to pCLE evaluation of the study polyps, the 2 expert endoscopists (MBW, AMB) viewed extensive published material on pCLE and performed a self-calibration on training pCLE videos of 20 polyps of known pathology (10 neoplastic and 10 non-neoplastic). These “training” polyps were evaluated by a gastrointestinal pathologist (MK) and came from 9 patients not included in the study. Once acquired, the pCLE videos of the study lesions were evaluated off-line and in random order by the 2 expert endoscopists, who were blinded to histology diagnosis and endoscopic appearance of the lesion. The off-line diagnosis of pCLE videos was made based on the established modified Mainz criteria[11] for diagnosis of colorectal neoplasia, and according to pit pattern and overall crypt and vessel architecture. Of the whole pCLE video imaging a polyp, the sequence of the video containing the most malignant pCLE features was considered to represent the polyp.

2.4. Histopathology as Criterion Standard Diagnosis

All resected specimens were reviewed by a reference gastrointestinal pathologist (MK) blinded to the pCLE information. Only the size and anatomic location were provided, which is the routine clinical practice at the Mayo Clinic institution. Intraepithelial neoplasia was

defined using modified Vienna criteria[12,13]: benign polyps and hyperplastic polyps were classified as non-neoplastic lesions, while tubular adenoma, villous adenoma, tubulovillous adenoma and adenocarcinoma were classified as neoplastic lesions.

2.5. Standard Bag-of-Visual-Words Technique for Content-Based Image Retrieval

As the endoscopists use perceptual similarities between pCLE videos of known diagnosis to establish a diagnosis on a new pCLE video, we propose a content-based retrieval approach to design the automated pCLE video classification software. We revisited the standard Bag-of-Visual-Words (BoW) technique which has been successfully used in many content-based image retrieval applications in computer vision[14]. A thorough technical presentation of our methodology has been disclosed previously[15] but without detailed clinical evaluation.

Standard BoW technique for image retrieval can be decomposed into four steps: region detection on the image, description of the regions, discretization of the feature space and similarity measuring between images. The detection step extracts salient regions in the image using sparse detectors. During the description step, a descriptor computes for each salient region its description vector. Then, the discretization step uses the result of a clustering method that builds K clusters, i.e. K visual words, from the union of the description vector sets gathered across all the images of the training database. Each description vector counts for one visual word, so an image can be represented by a signature of size K which is the histogram of its visual words. By construction, image signatures are invariant by viewpoint changes (image translation, rotation and scaling) and affine illumination changes. Finally, the similarity measuring step defines the similarity distance between two images as an adequate distance between their signatures: the most similar training images to the image of interest are defined as being the closest ones in

terms of this distance.

2.6. Adjusting Bag-of-Visual-Words Technique for pCLE Video Retrieval

First, we observe that discriminative information is densely distributed in pCLE images. Second, we notice that several pCLE image patterns have the same shape but represent different objects characterized by their different size (e.g. mesoscopic crypts and microscopic goblet cells both have a rounded shape). So pCLE image description must not be invariant by scaling. To avoid scale invariance and to extract all the image information, we decide to apply, instead of standard sparse detectors, a dense detector that is made of overlapping disks having a fixed radius and localized on a dense regular grid. We maintain the invariance by in-plane translation and rotation, because the pCLE miniprobe translates and rotates along the tissue surface. Besides, as the diffusion rate of fluorescein administered before imaging procedure decreases through time, invariance by affine illumination changes is also preserved.

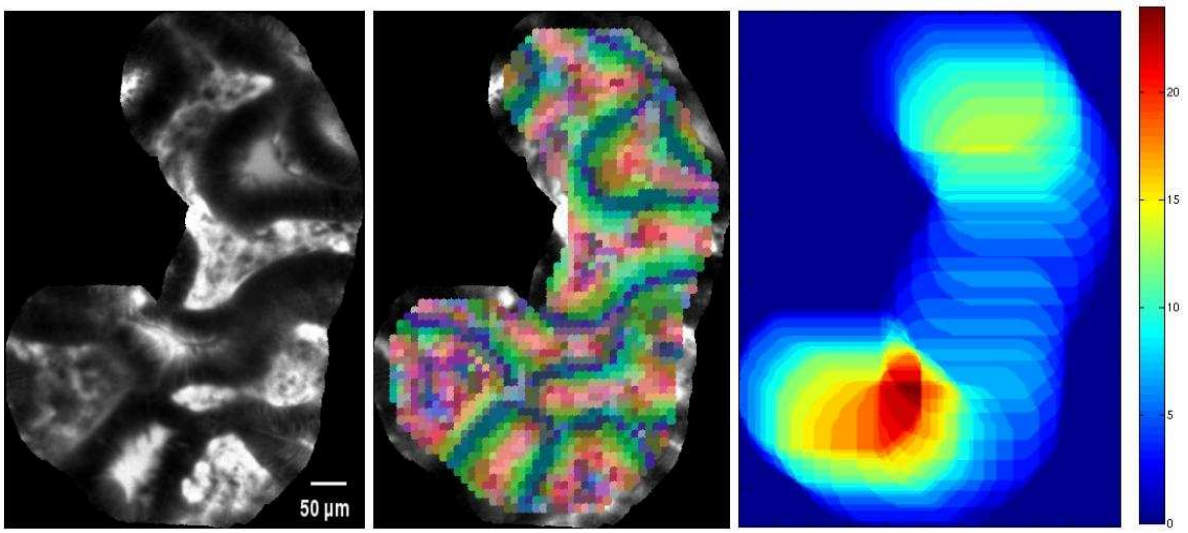

Figure 2: From left to right: Neoplastic pCLE mosaic obtained with non-rigid registration; Colored visual words mapped to the disk regions of radius 60 pixels in the mosaic image; Overlap scores of the local regions in the mosaic space, computed from the translation results of mosaicing.

Expert endoscopists pointed out that the field-of-view of single still images may not be large enough to make a robust diagnosis. So we decide to retrieve not single images but complete videos, by using the video mosaicing technique[16,17] (available in the Cellvizio software) to include spatial overlap between time-related images. Examples of mosaics built with the video mosaicing tool are shown in Figures 1, 4 and 5. To ensure online retrieval, we use the translation results of the real-time version of the video-mosaicing technique to weight the contribution of each local image region to its visual word, as illustrated in Figure 2. Then, we compute the video signatures with a histogram summation technique. Figure 3 presents the whole pipeline of our retrieval-based software classification framework, which can be run on line during ongoing colonoscopy.

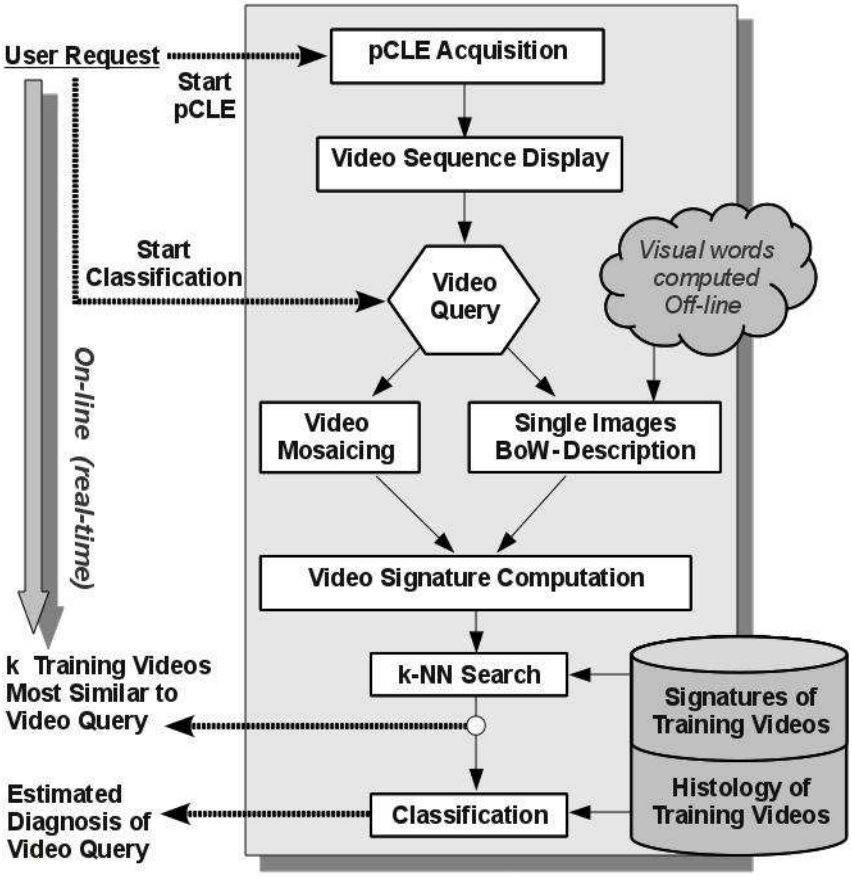

Figure 3: Pipeline of the pCLE retrieval-based software classification framework, from the acquisition of the pCLE video query by the Cellvizio system to the online automated diagnosis estimation.

2.7. Classification of pCLE Videos using Similarity Distance

Once the visual signature of the video query is computed, the k-Nearest Neighbor (k-NN) search step identifies the k closest training videos to the video query, by relying on the similarity distance between the video signatures. We then use the known histopathology diagnosis of these training videos to classify the query video, either as neoplastic or as non-neoplastic. Each of the k most similar training videos delivers a "histopathological" vote which is weighted by the inverse of its similarity distance to the video query.

Given the relatively small size of our pCLE database, we need to learn from as much data as possible. To avoid any bias while having a large training set, we employ cross-validation. As there are several videos acquired on the same patient, we perform a leave-one-patient-out cross-validation[18]: all videos from a given patient are excluded from the training set before being tested as queries of our retrieval and classification software. Cross-validation also allows us to find the optimal number of nearest neighbors, $k = 9$, which is the one that maximizes the accuracy of the retrieval-based software classification results.

2.8. Statistical Analysis

All the reported results of the automated software classification were obtained using leave-one-patient-out cross-validation. Statistical analysis was performed by BA.

To test for statistical difference between the two methods of interest, namely automated software classification and off-line classification by expert endoscopists, we use McNemar's tests[19] and show the corresponding power calculations with a type I error alpha=0.05. Two-sided P values < 0.05 were assumed to indicate statistical significance.

In order to assess statistical equivalence between the two methods, we use the two-sided *Z*-test between proportions[19,20] and compute 95% confidence intervals (CI). Because the 135 pCLE videos constitute a small sample size, we use a correction for continuity for the McNemar's test.

The statistics on overall accuracy are dependent on the relative fraction of non-neoplastic and neoplastic lesions examined, which in this study are 31.1% and 68.9%, respectively. Even though observations were made for more than one polyps in some patients, for the purposes of statistical analysis individual polyps (and their corresponding videos) were assumed to constitute independent observations. It is recognized that there was multiple testing of outcome data arising from individual polyps. Since the statistical tests were meant to highlight differences and since correction by Bonferroni's method would not have affected statistical significance in any of the comparisons, all *P* values are presented uncorrected for multiple testing.

3. RESULTS

3.1. Study Population and Colorectal Lesion Characteristics

Table 1 summarizes the demographic and general characteristics of the study population. None of the 71 patients experienced any endoscopic complications or adverse reaction to sodium fluorescein, with the exception of transient yellow discoloration of the skin and urine, which resolved by the time of discharge from the recovery room (skin) or within 24 hours (urine). Histopathology and morphological classification of the 135 analyzed colorectal lesions are also provided in Table 2.

Table 1. Study Population Characteristics

Study Population	Summary (n = 71)
Age	
median, (min, 25th, 75th, max)	75 (46, 68, 79, 93)
Gender, %	
Male	49
Female	51
History of colon cancer, %	9
Family history of colon cancer, %	10

Table 2. Colorectal Lesions Characteristics

Colorectal Lesions	Summary (n = 135)
Polyp size (mm)	
median, (min, 25th, 75th, max)	8 (1, 5, 20, 60)
Polyp location, %	
Cecum	24
Rectum	20
Ascending	18
Sigmoid	14.5
Transverse	15
Descending	5.5
Splenic flex	3
Histopathology diagnosis, %	
Hyperplastic	31
Tubular adenoma	52
Tubulovillous adenoma	11.5
Hyperplastic and adenomatous features	2.5
Adenocarcinoma	3
Neoplastic lesion, simplified histopathology, %	69
Paris classification, %	
1p	1
1s	57
2a	32
2b	5
2c	1
2a/c	4

3.2. Qualitative Results of Visual Similarities Between pCLE Videos

The pCLE database contains 135 pCLE videos representing each of the 135 polyps. The pCLE appearance of neoplastic lesions, compared to that of non-neoplastic lesions, included dilated irregular vessels, fluorescein leakage, cellular features of epithelial mucin depletion, and histological features of villiform crypts with increased optical density along epithelial border.

As the automated pCLE classification software is a similarity-based system that classified pCLE videos based on the votes of visually similar videos, its clinical relevance can be qualitatively evaluated by examining the intermediate results of video retrieval. Figure 4 shows 5 typical results of the automated pCLE retrieval software. We observe that, despite the high variability in appearance of a given histopathological class (neoplastic or non-neoplastic), the automatically retrieved videos called “neighbors” look quite similar to the video queries, respectively Q1, Q2, Q3 and Q4. Besides, we notice that the closer the neighbor is to the query, the more similar it is to it.

In terms of classification, the pathological class is estimated by the weighted votes of the 3 retrieved neighbors. In Figure 4, video queries Q1, Q2, Q3 and Q4 have been correctly classified with respect to histopathology, both by automated software classification and by expert endoscopists.

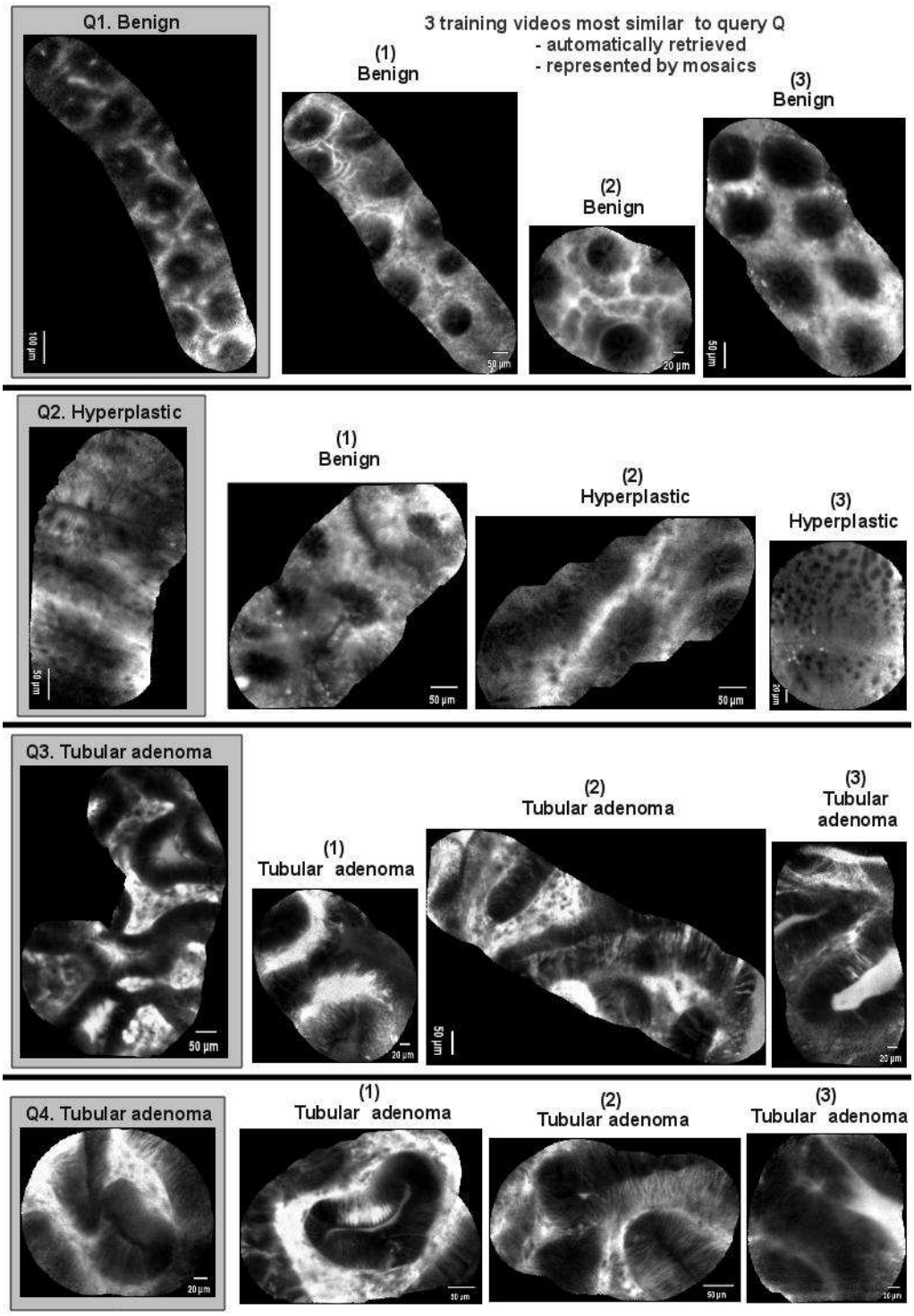

Figure 4: Typical results of automated pCLE video retrieval. The pCLE videos are represented by mosaic images; they are annotated with their histopathology diagnosis. Video queries are highlighted in gray and followed by their 3 most similar videos. Automated software classification (hyperplastic versus neoplastic) of query videos is based on the votes of the similar videos. With respect to histopathology, both the automated software classification and the pCLE diagnosis established by expert endoscopists are correct for these queries.

Figure 5 shows 3 other results that reveal some limitations of the automated pCLE retrieval software. Video query Q5 corresponds to a rare variety of hyperplastic polyp correctly classified as non-neoplastic by the expert endoscopists, but misclassified by the automated software classification because it is not represented in the training database for retrieval.

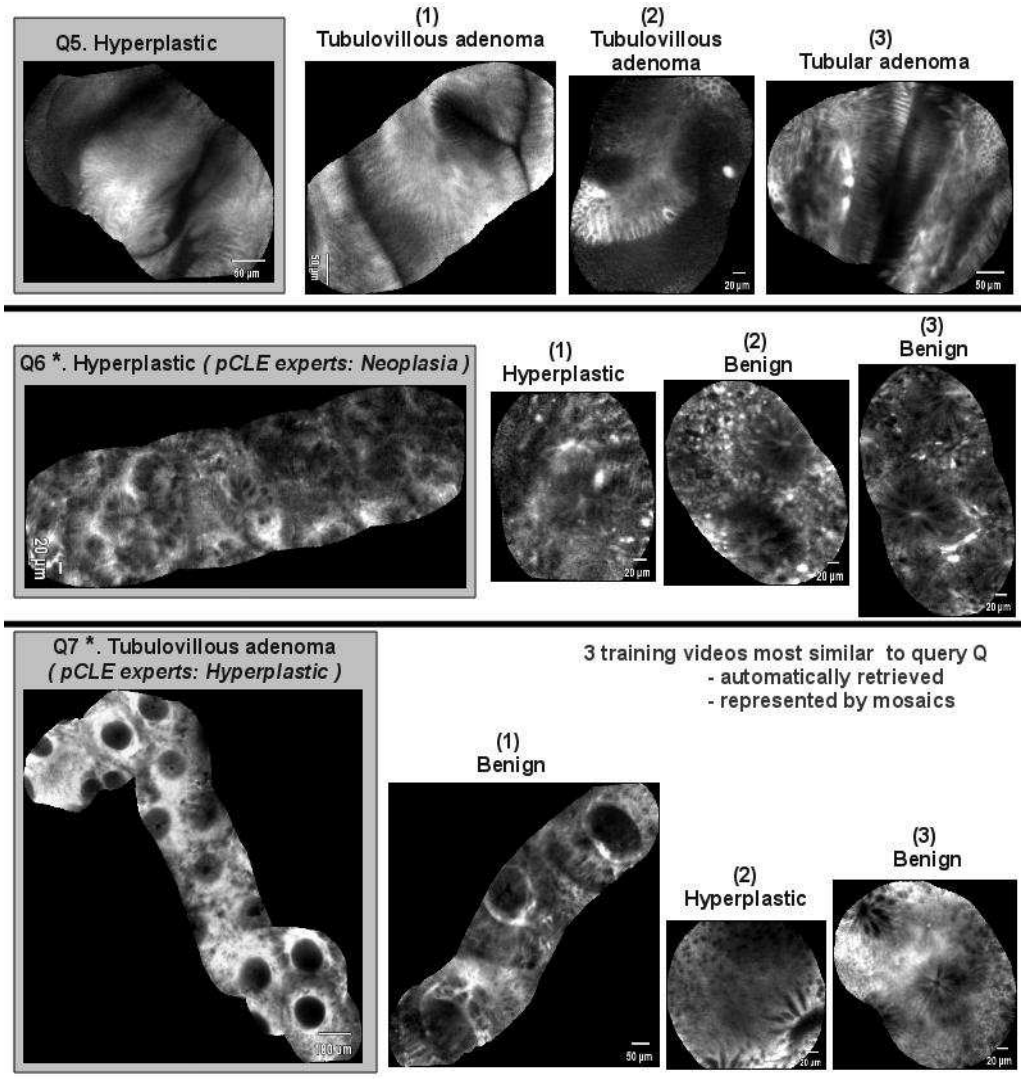

*Figure 5: Results of automated pCLE video retrieval represented as mosaics. With respect to histology: the automated software classification is correct for video query Q6 but incorrect for video queries Q5 and Q7, whereas the off-line diagnosis of pCLE videos established by the expert endoscopists is correct for video queries Q5 but incorrect for video queries Q6 and Q7 (for which this disagreement is marked by *).*

Video query Q6 corresponds to the ambiguous serrated adenoma case, correctly classified as non-neoplastic by the automated software classification, but misclassified by the expert endoscopists who consider serrated adenomas as malignant. Video query Q7 corresponds to a tubulovillous adenoma misclassified as non-neoplastic both by the expert endoscopists and by the automated software classification (this may be explained if a sampling error occurred and the corresponding biopsy was not performed exactly on the imaging spot).

3.3. Quantitative Results of Automated pCLE Classification Compared to Off-line Diagnosis of pCLE Videos established by Experts

Classification accuracy, sensitivity and specificity of the two methods, automated pCLE software classification (first method) and off-line diagnosis of pCLE videos established by the 2 expert endoscopists (second method), are listed in Table 3. Automated software classification reached a sensitivity of 92.5%, a specificity of 83.3% for a resulting accuracy of 89.6%. Expert review reached a sensitivity of 91.4%, a specificity of 85.7% and the same accuracy of 89.6%.

When testing for statistical difference, the *P* values provided by McNemar's tests show that the differences between the 2 methods are not statistically significant and that there is very low power (< 6%) to detect the observed differences.

When testing for statistical equivalence, the 95% confidence intervals provided by two-sided Z-tests between proportions are: - 0.073 to 0.073 for the accuracy, -0.068 to 0.089 for the sensitivity and -0.18 to 0.13 for the specificity. These intervals include zero and are sufficiently small to suggest that the methods are equivalent. In particular, the - 0.18 lower bound for the specificity is acceptable if the automated pCLE classification software is only taken as a second-reader tool to support pCLE diagnosis.

Table 3. Performance Comparison between Automated pCLE Classification and Off-line Expert Diagnosis of pCLE for the differentiation between neoplastic and non-neoplastic colonic polyps.

	(1) Automated pCLE Classification	(2) Off-line Expert Diagnosis of pCLE
Accuracy		
%	89.6	89.6
Fraction	121 / 135	121 / 135
Sensitivity		
%	92.5	91.4
Fraction	86 / 93	85 / 93
Specificity		
%	83.3	85.7
Fraction	35 / 42	36 / 42
Statistical significance between (1) and (2)		
McNemar's test, alpha=0.05		
Accuracy: (*P*, power)	(not significant, 2.5%)	
Sensitivity: (*P*, power)	(not significant, 6.5%)	
Specificity: (*P*, power)	(not significant, 5.2%)	
Statistical equivalence between (1) and (2)		
Two-sided *Z*-test		
95% CI for Accuracy	-0.073 to 0.073	
95% CI for Sensitivity	-0.068 to 0.089	
95% CI for Specificity	-0.18 to 0.13	

4. DISCUSSION

The present study demonstrates that, using a fairly representative database of colonic polyps, our automated software for the pCLE video classification has overall high accuracy, sensitivity and specificity, that are comparable to those of the off-line diagnosis of pCLE videos established by two endoscopists expert in pCLE. As the automated classification software can be run on line during ongoing colonoscopy, it could be used as a second-reader tool to support and improve not only off-line but also online diagnosis of pCLE established by endoscopists with various levels of expertise. In the majority of cases the second reader would agree with a moderately experienced endoscopist, who would be thus comforted in his/her diagnosis. For cases when they disagree, the endoscopist would have the opportunity to rethink his/her diagnosis and have more accurate *in vivo* interpretation. Besides, especially for small polyps, this second-reader tool could assist the endoscopist in adopting the “Diagnose, Resect and Discard Strategy” that dispenses with histopathological examination.

Gomez *et al.*[21] analyzed *in vivo* pCLE interpretation in distinguishing between neoplastic and non-neoplastic lesions among 3 expert endoscopists and estimated an average accuracy of 75% (sensitivity 76%, specificity 72%) with good to moderate interobserver agreement. Buchner *et al.*[22] demonstrated that accurate interpretation of pCLE images by 11 endoscopists, considered as non expert in pCLE, can be learned rapidly with a short 2 hour training session. The learning curve pattern of pCLE in predicting neoplastic lesions was demonstrated with improved accuracies in time from 63% to 86% as observers’ experience increased. Thus, prospectively, the automated classification software could be valuable not only for *in vivo* diagnosis support, but also for training support to improve the learning curve of the new endoscopists. Indeed, we have shown in a preliminary study[23] how interpretation difficulty can be automatically estimated by the software, in order to develop a self-training simulator for pCLE diagnosis with adjustable level of difficulty. For surgical skills,

evidences of the learning effect from the use of training tools have been provided in the thesis of Brydges[24], but further investigation is needed for the extension of learning effect analysis to diagnostic skills.

One of the advantages of our classification software is that it is not a “black box” but an informative tool based on the query by example model: it produces, as intermediate results, visually similar annotated videos that are directly interpretable by the endoscopist. From the qualitative observations of visual similarities between pCLE videos, we infer that the visually convincing results of the intermediate video retrieval step account for the relevance of the whole pCLE classification software. As few similar videos (less than 10) are necessary to classify a video query with a high accuracy, this visual information should be clinically useful for the endoscopist.

Further limitations of the classification software may include three main issues. First, a large training database is needed to be sufficiently representative of non-typical pCLE cases. This is even more challenging since the practice of pCLE is evolving and that new cases with atypical pCLE features may be still encountered. Second, the definition of “criterion standard” for colorectal cancer screening is debatable because expert endoscopists and pathologists do not always agree. This could be illustrated by many examples of hyperplastic polyps redefined later as sessile serrated lesions by GI pathologists, as in the study of Khalid *et al.*[25]. The third limitation is that an obtained biopsy may be acquired unintentionally from the area that does not correspond with the obtained pCLE imaging.

The task of the automated pCLE classification software is not to replace the endoscopist nor the pathologist but to assist the endoscopist in taking an informed decision. Before using the computer-based classification tool during an ongoing endoscopy procedure, more work is needed to improve its accuracy and to develop underlying tools that are both ergonomic and complementary. In particular, the online display of the retrieval outputs, for instance of the 3 most similar videos to the video

query, together with their histopathology and possible multimodal clinical data, may be a precious underlying indicator for diagnosis decision. Such a sophisticated “Smart Atlas” for pCLE would allow the endoscopists in different centers to share and enrich their pCLE knowledge during ongoing endoscopy. Further studies are warranted to evaluate the impact of using automated pCLE retrieval and classification software on the pCLE learning curve and on the diagnostic performance of the endoscopists.

COMMENTS

Background

Histopathology is the criterion standard for the diagnosis of colorectal cancers, but it implies a large proportion of unnecessary polypectomies and an inherent delay in diagnosis. Probe-based Confocal Laser Endomicroscopy (pCLE) is a recent technology that enables, during ongoing endoscopy, *in vivo* imaging of the epithelium at the microscopic level.

Research frontiers

Several studies have already demonstrated the applicability of pCLE in diagnosing colorectal neoplasia *in vivo* with high sensitivity and specificity. Because pCLE is a relatively recent imaging technology, the interpretation of pCLE videos of colonic polyps for diagnostic purposes is still challenging for many non-expert endoscopists.

Innovations and breakthroughs

This is believed to be the first study to propose, with the aim of supporting *in vivo* diagnosis of colorectal cancers, a CBIR-based classification software that automatically extracts visually similar annotated videos directly interpretable by the endoscopist. The extracted annotated videos can be presented to the endoscopist in a second reader paradigm to better support pCLE diagnosis. Furthermore, this study demonstrates that this novel software achieves a high diagnostic performance, which is statistically comparable to that of off-line diagnosis of pCLE videos established by expert endoscopists.

Applications

The classification software proposed in this study is an objective tool which has the potential to support the interpretation of pCLE videos of colonic polyps for diagnostic purposes. Further studies are warranted to

evaluate the impact of using the automated classification software on the pCLE learning curve and on the diagnostic performance of the endoscopists.

Terminology

Probe-based Confocal Laser Endomicroscopy (pCLE): an imaging system that allows the endoscopist to visualize the epithelium *in vivo*, at the microscopic level and in real-time during ongoing endoscopy.

Content-Based Image Retrieval (CBIR): a computer vision technique that automatically extracts, given a query image, several training images with the most similar appearance to the query.

Acknowledgments:

The authors thank Dr. Muhammad Waseem Shahid for his contribution to data collection and analysis. A preliminary report of this study was presented at the AGA Late Breaking Abstract Forum (Imaging and Advanced Technology - Confocal Endoscopy), Digestive Disease Week, 2010.

The present study uses data from the following clinical trial: NCT 00874263.

REFERENCES

1. **Hawk ET**, Levin B. Colorectal cancer prevention. *J Clin Oncol* 2005; **23**:378-391 [PMID: 15637400, DOI:10.1200/JCO.2005.08.097]

2. **Norfleet RG**, Ryan ME, Wyman JB. Adenomatous and hyperplastic polyps cannot be reliably distinguished by their appearance through the fiberoptic sigmoidoscope. *Dig Dis Sci* 1988; **33**:1175-1177 [PMID:3044716, DOI:10.1007/BF01535796]

3. **Rastogi A**, Keighley J, Singh V, Callahan P, Bansal A, Wani S, Sharma P. High accuracy of narrow band imaging without magnification for the real-time characterization of polyp histology and its comparison with high-definition white light colonoscopy: a prospective study. *Am J Gastroenterol* 2009; **104**: 2422-2430 [PMID:19584829, DOI:10.1038/ajg.2009.403]

4. **Winawer SJ**, Zauber AG, Fletcher RH, Stillman JS, O'Brien MJ, Levin B, Smith RA, Lieberman DA, Burt RW, Levin TR, Bond JH, Brooks D, Byers T, Hyman N, Kirk L, Thorson A, Simmang C, Johnson D, Rex DK. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. *CA Cancer J Clin* 2006; **56**:143-159; quiz 184-5 [PMID:16697750, DOI:10.3322/canjclin.56.3.143]

5. **Meining A**, Saur D, Bajbouj M, Becker V, Peltier E, Höfler H, von Weyhern CH, Schmid RM, Prinz C. In vivo histopathology for detection of gastrointestinal neoplasia with a portable, confocal miniprobe: an examiner blinded analysis. *Clin Gastroenterol Hepatol* 2007; **5**:1261-1267 [PMID:17689297, DOI:10.1016/j.cgh.2007.05.019]

6. **Venkatesh K**, Cohen M, Evans C, Delaney P, Thomas S, Taylor C, Abou-Taleb A, Kiesslich R, Thomson M. Feasibility of confocal endomicroscopy in

the diagnosis of pediatric gastrointestinal disorders. *World J Gastroenterol* 2009; **15**(18): 2214–2219 [PMID:19437560, DOI:10.3748/wjg.15.2214]

7. **De Palma GD**. Confocal laser endomicroscopy in the “in vivo” histological diagnosis of the gastrointestinal tract. *World J Gastroenterol* 2009; **15**(46):5770-5 [PMID:19998496, DOI:10.3748/wjg.15.5770]

8. **Buchner AM**, Shahid MW, Heckman MG, Krishna M, Ghabril M, Hasan M, Crook JE, Gomez V, Raimondo M, Woodward T, Wolfsen HC, Wallace MB. Comparison of probe-based confocal laser endomicroscopy with virtual chromoendoscopy for classification of colon polyps, *Gastroenterology* 2010; **138**(3):834-842 [PMID:19909747, DOI:10.1053/j.gastro.2009.10.053]

9. **Wallace MB**, Fockens P. Probe-based confocal laser endomicroscopy. *Gastroenterology* 2009, **136**(5):1509-1513 [PMID:19328799, DOI:10.1053/j.gastro.2009.03.034]

10. **Paris Workshop.** The Paris endoscopic classification of superficial neoplastic lesions: Esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc* 2003; **58**(Suppl):S3-43 [PMID:14652541, PII:S0016-5107(03)02159-X]

11. **Kiesslich R**, Burg J, Vieth M, Gnaendiger J, Enders M, Delaney P, Polglase A, McLaren W, Janell D, Thomas S, Nafe B, Galle PR, Neurath MF. Confocal laser endoscopy for diagnosing intraepithelial neoplasias and colorectal cancer in vivo. *Gastroenterology* 2004; **127**:706-713 [PMID:15362025, DOI:10.1053/j.gastro.2004.06.050]

12. **Schlemper RJ**, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, Dixon MF, Fenoglio-Preiser CM, Fléjou JF, Geboes K, Hattori T, Hirota T, Itabashi M, Iwafuchi M, Iwashita A, Kim YI, Kirchner T, Klimpfinger M, Koike

M, Lauwers GY, Lewin KJ, Oberhuber G, Offner F, Price AB, Rubio CA, Shimizu M, Shimoda T, Sipponen P, Solcia E, Stolte M, Watanabe H, Yamabe H. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000; **47**:251-255 [PMID:10896917, DOI:10.1136/gut.47.2.251]

13. **Rubio CA**, Nesi G, Messerini L, Zampi GC, Mandai K, Itabashi M, Takubo K. The Vienna classification applied to colorectal adenomas. *J Gastroenterol Hepatol* 2006; **21**:1697-1703 [PMID:16984592, DOI:10.1111/j.1440-1746.2006.04258.x]

14. **Zhang J**, Lazebnik S, Schmid C. Local features and kernels for classification of texture and object categories: a comprehensive study. *Int. J. Comput. Vis.* 2007; **73**: 213-238 [DOI:10.1007/s11263-006-9794-4]

15. **André B**, Vercauteren T, Buchner AM , Wallace MB, Ayache N. A smart atlas for endomicroscopy using automated video retrieval. *Med Im Analysis* 2011; **15**(4):460-476. IEEE [PMID:21414833, DOI:10.1016/j.media.2011.02.003]

16. **Vercauteren T**, Perchant A, Malandain G, Pennec X, Ayache N. Robust mosaicing with correction of motion distortions and tissue deformation for in vivo fibered microscopy. *Med Image Anal* 2006, **10**(5):673-692 [PMID:16887375, DOI:10.1016/j.media.2006.06.006]

17. **Becker V**, Vercauteren T, von Weyern CH, Prinz C, Schmid RM, Meining A. High resolution miniprobe-based confocal microscopy in combination with video-mosaicing. *Gastrointest Endosc* 2007; **66**(5):1001-1007 [PMID:17767932, DOI:10.1016/j.gie.2007.04.015]

18. **Dundar M**, Fung G, Bogoni L, Macari M, Megibow A, Rao RB. A methodology for training and validating a CAD system and potential

pitfalls. *Int J Comput Assisted Radiol Surg* 2004; 1010-1014 [DOI:10.1016/j.ics.2003.10.002]

19. **Sheskin, DJ**. Handbook of parametric and nonparametric statistical procedures. Chapman & Hall/CRC, 5th Revised edition, 2011

20. **Jones B**, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *Brit Med J* 1996; **313**:36-39 [BMJ 313 : 36]

21. **Gomez V**, Buchner AM, Dekker E, van den Broek FJ, Meining A, Shahid MW, Ghabril MS, Fockens P, Heckman MG, Wallace MB. Interobserver agreement and accuracy among international experts of probe-based confocal laser microscopy (pCLE) in predicting colorectal neoplasia. *Endoscopy* 2010; **42**(4):286-291 [DOI:10.1016/S0016-5085(09)62365-9]

22. **Buchner AM**, Gomez V, Gill KR, Ghabril M, Scimeca D, Shahid MW, Achem SR, Picco MF, Riegert-Johnson D, Raimondo M, Wolfsen HC, Woodward TA, Hasan MK, Wallace MB. The learning curve for in vivo probe based Confocal Laser Endomicroscopy (pCLE) for prediction of colorectal neoplasia. *Gastrointest Endosc* 2009; **69**(5):AB364-AB365 [DOI:10.1016/j.gie.2009.03.1086]

23. **André B**, Vercauteren T, Buchner AM, Shahid MW, Wallace MB, Ayache N. An image retrieval approach to setup difficulty levels in training systems for endomicroscopy diagnosis. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 2010; **6362**:480–487 [PMID:20879350, DOI:10.1007/978-3-642-15745-5_59]

24. **Brydges RN**. A critical reappraisal of self-learning in health professions education: Directed self-guided learning using simulation modalities. *University of Toronto,* 2009*;* http://hdl.handle.net/1807/19178

25. **Khalid O**, Radaideh S, Cummings OW, O'Brien MJ, Goldblum JR, Rex DK. Reinterpretation of histology of proximal colon polyps called hyperplastic in 2001. *World J Gastroenterol* 2009; **15**(30):3767-70 [PMID:19673017, DOI:10.3748/wjg.15.3767]