

Comparison of a Classification Software based on Image Retrieval with the Off-Line Diagnosis of Expert Endoscopists for probe-based Confocal Laser Endomicroscopy (pCLE) of Colorectal Polyps

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

BACKGROUND AND AIMS pCLE (Cellvizio, Mauna Kea Technologies) enables *in vivo* microscopic imaging of the epithelium in real-time during ongoing endoscopy. An image retrieval software prototype for automatic classification of pCLE images, recently developed to assist the endoscopists in the *in vivo* pCLE diagnosis of colorectal polyps, has the great potential of decreasing inter-observer agreement while increasing diagnostic performance of endoscopists. This study aims at comparing the performances of the classification software with the performance of pCLE diagnosis established off-line by expert endoscopists. **METHODS** Intravenous fluorescein pCLE imaging of colorectal lesions was performed on patients undergoing surveillance colonoscopies, followed by polypectomies. Histopathology was used as gold standard for the differentiation between neoplastic and non-neoplastic lesions. The pCLE sequences, recorded for each polyp, were analyzed off-line by 2 expert endoscopists, blinded to the endoscopic characteristics and histopathology. These pCLE videos, along with their histopathology diagnosis, were used to train the classification software which is a content-based image retrieval technique followed by *k*-nearest neighbor classification. All evaluations were performed using leave-one-patient-out (LOPO) cross-validation to avoid bias. **RESULTS** 135 colorectal lesions, including 6 serrated adenoma cases, were imaged in 71 patients. Based on histopathology, 93 of these 135 lesions were neoplastic and 42 were non-neoplastic. No statistical significance was found for the difference between the performance of software classification (accuracy 89.6%, sensitivity 92.5%, specificity 83.3%, using LOPO) and the performance of off-line diagnosis of pCLE established by the expert endoscopists (accuracy 89.6%, sensitivity 91.4%, specificity 85.7%). The 95% confidence intervals for equivalence testing (-0.073 to 0.073 for accuracy, -0.068 to 0.089 for sensitivity, -0.18 to 0.13 for specificity) are sufficiently small to suggest statistical equivalence. The -0.18 lower bound for the specificity should be sufficient if the classification software is only taken as a second-reader tool to support pCLE diagnosis. **CONCLUSIONS** The image retrieval software for automatic classification of pCLE sequences of colorectal polyps achieves a high performance which is statistically comparable to that of off-line diagnosis of pCLE sequences established by expert endoscopists. A fortiori, the classification software should be useful, not only to train non-expert endoscopists, but also to assist any endoscopist in *in vivo* pCLE diagnosis. **DISCUSSION** The proposed software is not a “black box” but an informative tool based on the query by example model that produces, as intermediate results, visually similar annotated pCLE videos directly interpretable by the endoscopist.

Table 1. Performance comparison between automatic software classification and off-line expert diagnosis of pCLE sequences, for the differentiation between neoplastic and non-neoplastic colorectal polyps. TP, TN, n, n1 and n2 successively indicate “true positives”, “true negatives”, “number of pCLE sequences”, “number of neoplastic pCLE sequences” and “number of non-neoplastic pCLE sequences”.

	(1) Automatic Software Classification (LOPO)	(2) Off-line Expert Diagnosis of pCLE
Accuracy		
%	89.6	89.6
(TP+TN)/n	121 / 135	121 / 135
Sensitivity		
%	92.5	91.4
TP/n1	86 / 93	85 / 93
Specificity		
%	83.3	85.7
TN/n2	35 / 42	36 / 42
Statistical significance between (1) and (2)		
McNemar's test, alpha=0.05		
Accuracy: (<i>P</i> , power)		(not significant, 2.5%)
Sensitivity: (<i>P</i> , power)		(not significant, 6.5%)
Specificity: (<i>P</i> , 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

Figure 1. Typical result of automatic retrieval of pCLE video sequences, represented by mosaic images and annotated with their histopathology diagnosis.

