Comparison of Classification Methods of Barret’s and Dysplasia in the Esophagus from In Vivo Optical Coherence Tomography Images

Christos Photiou¹, George Plastiras¹, Guillermo Tearney², Costas Pitiris¹

¹ KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, 75 Kallipoleos St, Nicosia, Cyprus, 1678
² Wellman Center for Photomedicine, Massachusetts General Hospital and Harvard Medical School, 55 Fruit Street, Boston, MA 02114, USA.

ABSTRACT
Endoscopic Optical Coherence Tomography (EOCT) systems can perform in vivo, real-time, high-resolution imaging of the human esophagus and, thus, play an important role in the earlier diagnosis and better prognosis of esophageal diseases such as Barrett’s, dysplasia and adenocarcinoma. However, the high image throughput and massive data volumes make manual evaluation of the generated information extremely difficult. Unfortunately, the algorithms, developed thus far, have not been able to provide effective computer-aided diagnosis. In this study, we compare different machine learning methods for tissue segmentation and classification of esophageal tissue in in vivo OCT images. An automated algorithm was developed, capable of discriminating normal tissue from Barrett’s Esophagus (BE) and dysplasia. The classification was based on various features of the epithelium, extracted from EOCT images, such as intensity-based statistics, the group velocity dispersion (GVD), estimated from the image speckle, and the scatterer size, calculated using the bandwidth of the correlation of the derivative (COD) method. The comparison and evaluation of various machine learning techniques has shown that a neural network based approach provided the best performance, classifying Barret’s esophagus and dysplasia, for individual A-Scans, with an accuracy of 89%.

1. INTRODUCTION
Barrett’s esophagus (BE) is a condition where the normal squamous epithelium of the esophagus transforms to columnar-like, due to gastroesophageal reflux. The frequency of adenocarcinoma in patients with BE is up to 35 times greater than the general population. Furthermore, over the last few decades, the incidence of BE is rapidly increasing in western countries. BE progresses through different stages of dysplasia before developing into esophageal adenocarcinoma thus providing a window for an early detection of the disease, which significantly improves patient prognosis. Patients with confirmed BE undergo periodic endoscopic surveillance with systematic biopsies. However, this procedure suffers from important limitations. Sampling error can lead to misdiagnosis since only a small proportion of the metaplastic BE epithelium is covered by the biopsies (4-6% of the area) [1,2]. Furthermore, the size and morphology of the samples obtained can cause diagnostic uncertainty since it limits inter-observer agreement between pathologists and causes delays in the histopathologic processing [3]. Endoscopic Optical Coherence Tomography (EOCT) systems can acquire cross-sectional images of the microscopic structure of the esophageal layers. By analyzing the microstructure of the epithelium in OCT images, researchers aim to characterize the state of the esophageal tissue, discriminating between normal esophagus and BE with or without dysplasia [4,5]. In the last 20 years, computer-aided diagnosis was exploited in many contexts to provide tissue characterization and classification between malignant and benign regions. Computational methods for the analysis of esophageal OCT images have also been demonstrated [6,7]. In this study, we classify different regions of esophageal OCT images, utilizing a fully automated algorithm for image segmentation and feature extraction. Several machine learning classifiers were evaluated for accuracy and capability to differentiate normal from abnormal tissues and BE vs. dysplasia from in vivo acquired data.

2. METHODOLOGY
A. Image and Data Processing
The data in this study was collected with a swept source EOCT system with a center wavelength of 1300 nm, an axial resolution of 10 μm and an A-Scan rate of 40 kHz. Each catheter rotation produced 2,048 A-Scan, displayed in real-time, and multiple cross-sectional esophageal images were collected as the catheter was manually pulled up from the gastroesophageal junction. In vivo data was collected from healthy volunteers and patients with esophageal disease enrolled...
in a study at Massachusetts General Hospital, approved by the Partner’s Internal Review Board (IRB). Normal, BE or dysplastic tissue regions were annotated on the OCT images by an expert. The part of the OCT image corresponding to the epithelium was automatically segmented and was used to extract image features for the training and testing of the machine learning classifier models.

B. Feature Extraction

The classification of the esophageal tissue was performed using various features extracted from the epithelial portion of the OCT images. These features included:

1. Intensity Statistics
   The statistics of the intensity of the OCT images were estimated for the portion of the OCT image corresponding to the epithelium. The statistics were calculated for each A-Scan but, also, separately for the upper and lower parts of the epithelium to capture the differences between the basal layers and the luminal surface of the epithelium. For each half, the calculated statistics included the mean, variance, standard deviation, skewness, kurtosis, median, and the total, minimum and maximum intensities. In addition, the sliding standard deviation of every statistic, an estimate of the variation between the values of adjacent A-Scans, was also included resulting in a total of 36 intensity-based statistics for every A-Scan.

2. Group Velocity Dispersion (GVD)
   Studies have shown that tissue dispersion could be used as a biomarker of early disease changes to enhance the diagnostic potential of OCT. For this study, we estimated tissue dispersion using the image’s speckle, a technique that does not require the presence of distinct reflectors in the imaged tissue and is, therefore, applicable to in vivo imaging. This method compares the image speckle from different portions of the segmented area, beginning from the top surface of the epithelium and gradually progressing to the bottom (~ 0.5 mm depth). The GVD is calculated, for each A-Scan, from the speckle width degradation that is proportional to the point spread function (PSF) degradation. [8]

3. Scatterer Size
   The average scatter size for each A-Scan of the epithelial layer was estimated using the bandwidth of the correlation of the derivative (COD) method. The COD is a new spectroscopic metric that extracts information about the modulation of depth-resolved spectra, as predicted by Mie theory, and can be used to calculate the average scatterer size. For each region of the epithelium, the spectrum was extracted using autoregressive spectral estimation. To estimate the scatterer size, the first derivative of the spectrum was calculated followed by its autocorrelation. The lag location of the first minimum of the autocorrelation was used to estimate the scatter size using a function derived from curve fitting of the expected COD calculated from Mie Theory. [9]

C. Feature Selection and Classification

In order to select the features that result in the best classification results, feature selection and optimization was performed. This process included utilization of a paired t-test and Multivariate Analysis of Variance (MANOVA). For the classification, five classifiers were, initially, evaluated: Discriminant Analysis (DA), Naïve-Bayes (NB), Decision Trees (DT), k-nearest neighbor (KNN) and Ensemble of Decision Trees (EDT). The performance of each classifier model was verified using leave-one-out cross-validation and was, subsequently, applied to entire images and volumes. Subsequently, using Python and Keras, different neural networks (NN) were constructed, varying the number of hidden layers and neurons, and each one evaluated on the same dataset. The optimal NN, consisted of 1 input, 1 output and 7 hidden layers with a total of 3,643 trainable parameters. A Rectified Linear Unit (ReLU) was used as an activation function with a learning rate of 0.01 as an optimizer. Each neighborhood of the epithelium was classified first as normal vs. abnormal and, subsequently, the abnormal areas were classified as BE vs. dysplasia. [10-13]
3. EXPERIMENTAL RESULTS

Figure 1 shows an example of an in vivo OCT image of the esophagus. In real-time, the image is displayed in the standard Cartesian coordinates (Fig. 1A) which more accurately represent the geometry of the esophagus. However, for processing purposes, the image was retained in polar coordinates (Fig. 1B) so that individual A-Scans could be processed separately. The regions of the images in this dataset were annotated as normal, BE and dysplasia (Fig. 1B, yellow boxes). The epithelium was automatically segmented using automatic thresh-holding and appropriate morphological operations, thus, identifying the top and bottom boarders of the epithelium (Fig. 1B, red and green lines). The segmented regions (Fig1C) were used in the machine learning process.

![Figure 1](image_url)

**Figure 1.** (A) In vivo OCT image of the human esophagus in Cartesian coordinates. (B) Same image in polar coordinates with the red and green lines indicating the top and bottom borders of the automatically segmented epithelial region. The yellow boxes indicate annotated dysplastic (a), BE (b) and Normal (c) regions. (C) Zoomed regions corresponding (from top to bottom) to the dysplastic, BE and normal annotated areas respectively.

Table 1 summarizes the performance of all models for normal vs abnormal classification (left) and Barrett’s vs. dysplasia (right). The performance of each classification scheme was verified using a leave-one image-out-cross validation approach for the same features selected based on t-testing and MANOVA (p-value <0.05). Mean, variance and median, of both the upper and lower parts of the segmented epithelial regions, along with the maximum of the lower area and the GVD appeared to be the most significant features. Normal vs. abnormal discrimination was more challenging to perform with the accuracy ranging from 63 to 73%, sensitivity from 25 to 80% and specificity from 57 to 69%. The neural network approach resulted in the best accuracy (73%), 80% sensitivity and 64% specificity. For Barrett’s vs. dysplasia, the classifiers resulted in accuracy values ranging from 60 to 89%, sensitivity from 45 to 80% and specificity from 71 to 92%. Again, the neural network classifier had the best accuracy (89%), with sensitivity and specificity of 71% and 92% respectively.

|       | N vs A |       | B vs D |       |
|-------|--------|-------|--------|-------|
|       | Sensitivity [\%] | Specificity [\%] | Accuracy [\%] | AUC | G-mean | Sensitivity [\%] | Specificity [\%] | Accuracy [\%] | AUC | G-mean |
| LDA   | 29     | 62    | 67     | 0.66 | 0.42   | 80    | 91    | 74     | 0.57 | 0.85   |
| k-NN  | 31     | 57    | 64     | 0.63 | 0.41   | 67    | 71    | 82     | 0.57 | 0.74   |
| NB    | 33     | 69    | 68     | 0.68 | 0.48   | 45    | 71    | 60     | 0.63 | 0.56   |
| EDT   | 25     | 63    | 70     | 0.69 | 0.40   | 69    | 87    | 75     | 0.60 | 0.77   |
| DT    | 36     | 58    | 63     | 0.61 | 0.46   | 57    | 77    | 69     | 0.59 | 0.66   |
| NN    | 80     | 64    | 73     | 0.76 | 0.72   | 71    | 92    | 89     | 0.66 | 0.81   |

Table 1. Classification results for Normal vs Abnormal (left) and Barret’s vs Dysplasia (right) using leave-one image out-cross-validation.
4. CONCLUSIONS

Given the results presented above, the automated algorithm proposed could be developed to perform in vivo OCT esophageal tissue segmentation and classification. Using machine learning with a neural network, the algorithm can distinguish Barrett’s esophagus from dysplasia with an accuracy of 89%, sensitivity of 71% and specificity of 92%. These results are very promising and improve on the current state of the art. However, further evaluation is needed, with a larger number of images from more patients, to optimize the classifier models and create a system that can be used for effective computer aided diagnosis from OCT images.

Acknowledgments

This study is funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE) and from the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

5. REFERENCES

1. J Evans, et al, “Optical Coherence Tomography to Identify Intramucosal Carcinoma and High-Grade Dysplasia in Barrett’s Esophagus”, Clinical Gastroenterology and Hepatology, Volume 4, Issue 1, 38 – 43 (2006)
2. G.Isenberg, et al, “Accuracy of endoscopic optical coherence tomography in the detection of dysplasia in Barrett’s esophagus: a prospective, double-bIinded study”, Gast. Endos. Volume 62, 825-831. (2005)
3. Downs-Kelly, E., et al, “Poor interobserver agreement in the distinction of high-grade dysplasia and adenocarcinoma in pretreatment Barrett’s esophagus biopsies”, 103(9), 2333-40. (2008)
4. Gora, M.J., et al, “Tethered capsule endomicroscopy enables less-invasive imaging of gastrointestinal tract microstructure”, Nat Med. 19(2), 238-40. (2013)
5. Yun, S.H., et al, “Comprehensive volumetric optical microscopy in vivo”, Nature medicine, 12(12), p.1429. (2006)
6. X. Qi, et al, “Image analysis for classification of dysplasia in Barrett’s esophagus using endoscopic optical coherence tomography”, Biom. Opt Expr. Sep 9 (13):825-847. (2010)
7. J. Ughi, et al, “Automated segmentation and characterization of esophageal wall in vivo by tethered capsule optical coherence tomography endomicroscopy”, Biom. Opt Expr. 1 Feb. Vol.7, No. 2. (2016)
8. C.Photiou, et al, “Using speckle to measure tissue dispersion in optical coherence tomography”, Biom. Opt Expr. May 1; 8(5): 2528–2535. (2017)
9. Kassinopoulos, M., et al, “Correlation of the derivative as a robust estimator of scatterer size in optical coherence tomography (OCT)”, Biomed Opt Expr. 8(3), 1598-606.(2017)
10. Mika et al., “Neural net. for signal proc. IX”, IEEE, pp. 41–48. (1999)
11. J. Quinlan, , “Induction of Decision Trees”, Machine learning, vol. 1, no. 1, pp. 81–106. (1986)
12. D. Opitz and R. Maclin, “Popular ensemble methods: An empirical study”, Jour. of AI res vol. 11, pp. 169–198. (1999)
13. J. Erickson, et al, “Machine Learning for Medical Imaging”.Radiographics. 37(2): 505–515. (2017)