Image analysis for classification of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography

Xin Qi,[1] Yinsheng Pan,[1] Michael V. Sivak, Jr.,[2] Joseph E. Willis,[3] Gerard Isenberg,[2] and Andrew M. Rollins[1,2*]

[1]*Departments of Biomedical Engineering, Case Western Reserve University, Cleveland, OH 44106, USA*
[2]*Departments of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA*
[3]*Departments of Pathology, Case Western Reserve University, Cleveland, OH 44106, USA*
**Rollins@case.edu*

Abstract: Barrett's esophagus (BE) and associated adenocarcinoma have emerged as a major health care problem. Endoscopic optical coherence tomography is a microscopic sub-surface imaging technology that has been shown to differentiate tissue layers of the gastrointestinal wall and identify dysplasia in the mucosa, and is proposed as a surveillance tool to aid in management of BE. In this work a computer-aided diagnosis (CAD) system has been demonstrated for classification of dysplasia in Barrett's esophagus using EOCT. The system is composed of four modules: region of interest segmentation, dysplasia-related image feature extraction, feature selection, and site classification and validation. Multiple feature extraction and classification methods were evaluated and the process of developing the CAD system is described in detail. Use of multiple EOCT images to classify a single site was also investigated. A total of 96 EOCT image-biopsy pairs (63 non-dysplastic, 26 low-grade and 7 high-grade dysplastic biopsy sites) from a previously described clinical study were analyzed using the CAD system, yielding an accuracy of 84% for classification of non-dysplastic vs. dysplastic BE tissue. The results motivate continued development of CAD to potentially enable EOCT surveillance of large surface areas of Barrett's mucosa to identify dysplasia.

OCIS codes: (110.4500) Optical coherence tomography; (170.2150) Endoscopic imaging; (100.2960) Image analysis.

References and links

1. K. R. DeVault, "Epidemiology and significance of Barrett's esophagus," Dig. Dis. **18**(4), 195–202 (2000).
2. R. E. Sampliner; Practice Parameters Committee of the American College of Gastroenterology, "Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus," Am. J. Gastroenterol. **97**(8), 1888–1895 (2002).
3. R. Tutuian, and D. O. Castell, "Barrett's esophagus prevalence and epidemiology," Gastrointest. Endosc. Clin. N. Am. **13**(2), 227–232 (2003).
4. M. Solaymani-Dodaran, R. F. Logan, J. West, and T. Card, "Mortality associated with Barrett's esophagus and gastroesophageal reflux disease diagnoses-a population-based cohort study," Am. J. Gastroenterol. **100**(12), 2616–2621 (2005).
5. M. Vieth, B. Schubert, K. Lang-Schwarz, and M. Stolte, "Frequency of Barrett's neoplasia after initial negative endoscopy with biopsy: a long-term histopathological follow-up study," Endoscopy **38**(12), 1201–1205 (2006).
6. H. Sayana, S. Wani, and P. Sharma, "Esophageal adenocarcinoma and Barrett's esophagus," Minerva Gastroenterol. Dietol. **53**(2), 157–169 (2007).
7. C. C. Maley, and A. K. Rustgi, "Barrett's esophagus and its progression to adenocarcinoma," J. Natl. Compr. Canc. Netw. **4**(4), 367–374 (2006).

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

8. J. L. Hornick, and R. D. Odze, "Neoplastic precursor lesions in Barrett's esophagus," Gastroenterol. Clin. North Am. **36**(4), 775–796, v (2007).
9. P. Sharma, K. McQuaid, J. Dent, M. B. Fennerty, R. Sampliner, S. Spechler, A. Cameron, D. Corley, G. Falk, J. Goldblum, J. Hunter, J. Jankowski, L. Lundell, B. Reid, N. J. Shaheen, A. Sonnenberg, K. Wang, W. Weinstein; AGA Chicago Workshop, "A critical review of the diagnosis and management of Barrett's esophagus: the AGA Chicago Workshop," Gastroenterology **127**(1), 310–330 (2004).
10. B. E. Bouma, G. J. Tearney, C. C. Compton, and N. S. Nishioka, "High-resolution imaging of the human esophagus and stomach in vivo using optical coherence tomography," Gastrointest. Endosc. **51**(4), 467–474 (2000).
11. M. V. J. Sivak, Jr., K. Kobayashi, J. A. Izatt, A. M. Rollins, R. Ung-Runyawee, A. Chak, R. C. Wong, G. A. Isenberg, and J. Willis, "High-resolution endoscopic imaging of the GI tract using optical coherence tomography," Gastrointest. Endosc. **51**(4), 474–479 (2000).
12. X. D. Li, S. A. Boppart, J. Van Dam, H. Mashimo, M. Mutinga, W. Drexler, M. Klein, C. Pitris, M. L. Krinsky, M. E. Brezinski, and J. G. Fujimoto, "Optical Coherence Tomography: Advanced Technology for the Endoscopic Imaging of Barrett's Esophagus," Endoscopy **32**(12), 921–930 (2000).
13. S. Jäckle, N. Gladkova, F. Feldchtein, A. Terentieva, B. Brand, G. Gelikonov, V. Gelikonov, A. Sergeev, A. Fritscher-Ravens, J. Freund, U. Seitz, S. Schröder, and N. Soehendra, "In vivo endoscopic optical coherence tomography of esophagitis, Barrett's esophagus, and adenocarcinoma of the esophagus," Endoscopy **32**(10), 750–755 (2000).
14. G. A. Isenberg, and M. V. Sivak, Jr., "Gastrointestinal optical coherence tomography," Tech. Gastrointest. Endosc. **5**(2), 94–101 (2003).
15. V. Westphal, A. M. Rollins, J. Willis, M. V. Sivak, Jr., and J. A. Izatt, "Correlation of endoscopic optical coherence tomography with histology in the lower-GI tract," Gastrointest. Endosc. **61**(4), 537–546 (2005).
16. P. R. Pfau, M. V. Sivak, Jr., A. Chak, M. Kinnard, R. C. Wong, G. A. Isenberg, J. A. Izatt, A. Rollins, and V. Westphal, "Criteria for the diagnosis of dysplasia by endoscopic optical coherence tomography," Gastrointest. Endosc. **58**(2), 196–202 (2003).
17. G. A. Isenberg, M. V. Sivak, Jr., A. Chak, R. C. K. Wong, J. E. Willis, B. Wolf, D. Y. Rowland, A. Das, and A. M. Rollins, "Accuracy of endoscopic optical coherence tomography in the detection of dysplasia in Barrett's esophagus: a prospective, double-blinded study," Gastrointest. Endosc. **62**(6), 825–831 (2005).
18. J. A. Evans, J. M. Poneros, B. E. Bouma, J. Bressner, E. F. Halpern, M. Shishkov, G. Y. Lauwers, M. Mino-Kenudson, N. S. Nishioka, and G. J. Tearney, "Optical coherence tomography to identify intramucosal carcinoma and high-grade dysplasia in Barrett's esophagus," Clin. Gastroenterol. Hepatol. **4**(1), 38–43 (2006).
19. X. Qi, M. V. Sivak, Jr., G. Isenberg, J. E. Willis, and A. M. Rollins, "Computer-aided diagnosis of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography," J. Biomed. Opt. **11**(4), 044010 (2006).
20. C. Pitris, C. Jesser, S. A. Boppart, D. Stamper, M. E. Brezinski, and J. G. Fujimoto, "Feasibility of optical coherence tomography for high-resolution imaging of human gastrointestinal tract malignancies," J. Gastroenterol. **35**(2), 87–92 (2000).
21. J. M. Poneros, S. Brand, B. E. Bouma, G. J. Tearney, C. C. Compton, and N. S. Nishioka, "Diagnosis of specialized intestinal metaplasia by optical coherence tomography," Gastroenterology **120**(1), 7–12 (2001).
22. J. M. Poneros, "Diagnosis of Barrett's esophagus using optical coherence tomography," Gastrointest. Endosc. Clin. N. Am. **14**(3), 573–588, x (2004).
23. L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology **215**(2), 554–562 (2000).
24. B. Sahiner, N. Petrick, H. P. Chan, L. M. Hadjiiski, C. Paramagul, M. A. Helvie, and M. N. Gurcan, "Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization," IEEE Trans. Med. Imaging **20**(12), 1275–1284 (2001).
25. R. L. Ellis, A. A. Meade, M. A. Mathiason, K. M. Willison, and W. Logan-Young, "Evaluation of computer-aided detection systems in the detection of small invasive breast carcinoma," Radiology **245**(1), 88–94 (2007).
26. S. B. Göktürk, C. Tomasi, B. Acar, C. F. Beaulieu, D. S. Paik, R. B. Jeffrey, Jr., J. Yee, and S. Napel, "A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography," IEEE Trans. Med. Imaging **20**(12), 1251–1260 (2001).
27. R. M. Summers, "Challenges for computer-aided diagnosis for CT colonography," Abdom. Imaging **27**(3), 268–274 (2002).
28. K. Suzuki, F. Li, S. Sone, and K. Doi, "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," IEEE Trans. Med. Imaging **24**(9), 1138–1150 (2005).
29. D. G. Vince, K. J. Dixon, R. M. Cothren, and J. F. Cornhill, "Comparison of texture analysis methods for the characterization of coronary plaques in intravascular ultrasound images," Comput. Med. Imaging Graph. **24**(4), 221–229 (2000).
30. K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Computerized diagnosis of breast lesions on ultrasound," Med. Phys. **29**(2), 157–164 (2002).
31. C. M. Chen, Y. H. Chou, K. C. Han, G. S. Hung, C. M. Tiu, H. J. Chiou, and S. Y. Chiou, "Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks," Radiology **226**(2), 504–514 (2003).

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

32. S. Joo, Y. S. Yang, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features," IEEE Trans. Med. Imaging **23**(10), 1292–1300 (2004).
33. K. Horsch, M. L. Giger, C. J. Vyborny, and L. A. Venta, "Performance of computer-aided diagnosis in the interpretation of lesions on breast sonography," Acad. Radiol. **11**(3), 272–280 (2004).
34. B. Sahiner, H. P. Chan, M. A. Roubidoux, L. M. Hadjiiski, M. A. Helvie, C. Paramagul, J. Bailey, A. V. Nees, and C. Blane, "Malignant and benign breast masses on 3D US volumetric images: effect of computer-aided diagnosis on radiologist accuracy," Radiology **242**(3), 716–724 (2007).
35. K. G. A. Gilhuijs, M. L. Giger, and U. Bick, "Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging," Med. Phys. **25**(9), 1647–1654 (1998).
36. W. Chen, M. L. Giger, L. Lan, and U. Bick, "Computerized interpretation of breast MRI: investigation of enhancement-variance dynamics," Med. Phys. **31**(5), 1076–1082 (2004).
37. P. M. White, E. M. Teasdale, J. M. Wardlaw, and V. Easton, "Intracranial aneurysms: CT angiography and MR angiography for detection prospective blinded comparison in a large patient cohort," Radiology **219**(3), 739–749 (2001).
38. F. M. Hall, "Improved sensitivity of mammography with computer-aided detection," AJR Am. J. Roentgenol. **182**(6), 1598–1599 (2004).
39. S. H. Taplin, C. M. Rutter, and C. D. Lehman, "Testing the effect of computer-assisted detection on interpretive performance in screening mammography," AJR Am. J. Roentgenol. **187**(6), 1475–1482 (2006).
40. S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," Radiology **232**(2), 578–584 (2004).
41. D. Gur, J. H. Sumkin, H. E. Rockette, M. Ganott, C. Hakim, L. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**(3), 185–190 (2004).
42. M. L. Giger, N. Karssemeijer, and S. G. Armato 3rd, "Computer-aided diagnosis in medical imaging," IEEE Trans. Med. Imaging **20**(12), 1205–1208 (2001).
43. I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," IEEE Trans. Med. Imaging **25**(4), 385–405 (2006).
44. R. L. Van Uitert, and R. M. Summers, "Automatic correction of level set based subvoxel precise centerlines for virtual colonoscopy using the colon outer wall," IEEE Trans. Med. Imaging **26**(8), 1069–1078 (2007).
45. S. Timp, C. Varela, and N. Karssemeijer, "Temporal change analysis for characterization of mass lesions in mammography," IEEE Trans. Med. Imaging **26**(7), 945–953 (2007).
46. M. A. Kupinski, and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," IEEE Trans. Med. Imaging **17**(4), 510–517 (1998).
47. N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," Med. Phys. **23**(10), 1685–1696 (1996).
48. Z. Huo, M. L. Giger, D. E. Wolverton, W. Zhong, S. Cumming, and O. I. Olopade, "Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection," Med. Phys. **27**(1), 4–12 (2000).
49. B. van Ginneken, S. Katsuragawa, B. M. ter Haar Romeny, K. Doi, and M. A. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," IEEE Trans. Med. Imaging **21**(2), 139–149 (2002).
50. X. Qi, M. V. Sivak, Jr., D. L. Wilson, and A. M. R. Rollins, "Computer Aided diagnosis (CAD) of dysplasia in Barrett's esophagus (BE) using endoscopic optical coherence tomography (EOCT)," Gastroenterology **126**, A351 (2004).
51. G. W. Falk, T. W. Rice, J. R. Goldblum, and J. E. Richter, "Jumbo biopsy forceps protocol still misses unsuspected cancer in Barrett's esophagus with high-grade dysplasia," Gastrointest. Endosc. **49**(2), 170–176 (1999).
52. A. M. Rollins, S. Yazdanfar, M. Kulkarni, R. Ung-Arunyawee, and J. A. Izatt, "In vivo video rate optical coherence tomography," Opt. Express **3**(6), 219–229 (1998).
53. A. M. Rollins, R. Ung-Arunyawee, A. Chak, R. C. K. Wong, K. Kobayashi, M. V. Sivak, Jr., and J. A. Izatt, "Real-time in vivo imaging of human gastrointestinal ultrastructure by use of endoscopic optical coherence tomography with a novel efficient interferometer design," Opt. Lett. **24**(19), 1358–1360 (1999).
54. A. Das, M. V. Sivak Jr., A. Chak, R.C.K. Wong, V. Westphal, A.M. Rollins, J. Izatt, G.A. Isenberg, and J. Willis, "Role of high-resolution endoscopic imaging using optical coherence tomography (OCT) in patients with Barrett's esophagus (BE)," Gastrointest. Endosc. **51**, AB93, Part92 (2000).
55. Y. Chen, A. D. Aguirre, P. L. Hsiung, S. Desai, P. R. Herz, M. Pedrosa, Q. Huang, M. Figueiredo, S. W. Huang, A. Koski, J. M. Schmitt, J. G. Fujimoto, and H. Mashimo, "Ultrahigh resolution optical coherence tomography of Barrett's esophagus: preliminary descriptive clinical study correlating images with histology," Endoscopy **39**(7), 599–605 (2007).
56. J. M. Schmitt, A. Knüttel, and R. F. Bonner, "Measurement of optical properties of biological tissues by low-coherence reflectometry," Appl. Opt. **32**(30), 6032–6042 (1993).

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

57. J. M. Schmitt, A. Knüttel, M. Yadlowsky, and M. A. Eckhaus, "Optical-coherence tomography of a dense tissue: statistics of attenuation and backscattering," Phys. Med. Biol. **39**(10), 1705–1720 (1994).
58. B. B. Chaudhuri, P. Kundu, and N. Sarkar, "Detection and gradation of oriented texture," Pattern Recognit. Lett. **14**(2), 147–153 (1993).
59. W. Karlon, J. W. Covell, A. D. Mcculloch, J. J. Hunter, and J. H. Omens, "Automated measurement of myofiber disarray in transgenic mice with ventricular expression of ras," Anat. Rec. **252**(4), 612–625 (1998).
60. N. I. Fisher, Statistical analysis of circular data, (Cambridge: Cambridge University Press, 1993).
61. I. Pitas, Digital image processing algorithms and applications, (John Wiley & Sons, 2000).
62. M. Tuceryan, and A. K. Jain, "Texture analysis," the Handbook of Pattern Recognition and Computer Vision, 207–248 (1998).
63. D. Harwood, T. Ojala, M. Pietikainen, S. Kelman, and L. Davis, "Texture classification by center-symmetric auto-correlation using Kullback discrimination of distributions," Pattern Recognit. Lett. **16**(1), 1–10 (1995).
64. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst. Man Cybern. **3**(6), 610–621 (1973).
65. R. M. Haralick, "Statistical and structural approaches to texture," Proc. SPIE **67**, 786–804 (1979).
66. M. H. Horng, Y. N. Sun, and X. Z. Lin, "Texture feature coding method for classification of liver sonography," Comput. Med. Imaging Graph. **26**(1), 33–42 (2002).
67. K. Laws, "Rapid texture identification," Proc. SPIE **238**, 376–380 (1980).
68. J. Reunanen, I. Guyon, and A. Elisseeff, "Overfitting in making comparisons between variable selection methods," J. Mach. Learn. Res. **3**(7-8), 1371–1382 (2003).
69. L. Yu, and H. Liu, "Efficient feature selection via analysis of relevant and redundancy," J. Mach. Learn. Res. **5**, 1205–1224 (2004).
70. I. T. Jolliffe, Principal component analysis, (Springer-Verlag New York Inc, 1986).
71. R. A. Fisher, "The use of multiple measurements in taxonomic problems," Ann. Eugen. **7**, 179–188 (1936).
72. T. M. Cover, and P. E. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory **13**(1), 21–27 (1967).
73. M. Caudill, and C. Butler, Understanding neural networks: computer explorations, (Cambridge, MA: The MIT Press **1 &2,** 1992).
74. T. Kohonen, Self-organization and associative memory, (Berlin:Springer-Verlag, 1987).
75. L. Breiman, J. H. Friedman, R. A. Olsen, and C. J. Stone, Classification and regression tree, (Chapman & Hall, New York, 1984).
76. B. D. Ripley, "Pattern Recognition," (1996).
77. J. Park, and D. W. Edington, "A sequential neural network model for diabetes prediction," Artif. Intell. Med. **23**(3), 277–293 (2001).
78. B. Efron, "Bootstrap methods: another look at the jackknife," Ann. Stat. **7**(1), 1–26 (1979).
79. B. Efron and R.J. Tibshirani, "An introduction to the bootstrap," (1993).
80. T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning: data mining, inference and prediction, (Springer-Verlag New York Inc, 2001).
81. L. Breiman, "Bagging predictors," Mach. Learn. **24**(2), 123–140 (1996).
82. L. Breiman, "Arcing classifiers," Ann. Stat. **26**, 801–849 (1998).
83. S. H. Yun, G. J. Tearney, B. J. Vakoc, M. Shishkov, W. Y. Oh, A. E. Desjardins, M. J. Suter, R. C. Chan, J. A. Evans, I. K. Jang, N. S. Nishioka, J. F. de Boer, and B. E. Bouma, "Comprehensive volumetric optical microscopy in vivo," Nat. Med. **12**(12), 1429–1433 (2007).
84. B. J. Vakoc, M. Shishko, S. H. Yun, W. Y. Oh, M. J. Suter, A. E. Desjardins, J. A. Evans, N. S. Nishioka, G. J. Tearney, and B. E. Bouma, "Comprehensive esophageal microscopy by using optical frequency-domain imaging (with video)," Gastrointest. Endosc. **65**(6), 898–905 (2007).
85. G. Zuccaro, N. Gladkova, J. Vargo, F. Feldchtein, E. Zagaynova, D. Conwell, G. Falk, J. Goldblum, J. Dumot, J. Ponsky, G. Gelikonov, B. Davros, E. Donchenko, and J. Richter, "Optical coherence tomography of the esophagus and proximal stomach in health and disease," Am. J. Gastroenterol. **96**(9), 2633–2639 (2001).
86. D. C. Adler, C. Zhou, T. H. Tsai, H. C. Lee, L. Becker, J. M. Schmitt, Q. Huang, J. G. Fujimoto, and H. Mashimo, "Three-dimensional optical coherence tomography of Barrett's esophagus and buried glands beneath neosquamous epithelium following radiofrequency ablation," Endoscopy **41**(9), 773–776 (2009).
87. A. M. Prasad, L. R. Iverson, and A. Liaw, "Newer classification and regression tree techniques: bagging and random forests for ecological prediction," Ecosystems (N. Y.) **9**(2), 181–199 (2006).

1. Introduction

Barrett's esophagus (BE) is a pre-malignant precursor lesion for esophageal adenocarcinoma [1–9]. Endoscopic optical coherence tomography (EOCT) has demonstrated interpretable high-resolution images of mucosa and submucosa in the gastrointestinal (GI) tract [10–15] and is capable of identifying dysplasia in gastrointestinal mucosal tissue [16–19]. BE is the most extensively studied potential GI application of EOCT [12,13,17,18,20–22].

Computer-aided diagnosis (CAD) is a relatively new field that aims to provide a diagnosing physician with quantitative and objective measures and interpretation of image

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

features. In the last two decades CAD has been applied using several imaging modalities (e.g. X-ray mammography and computed tomography (CT) [23–28] ultrasound [29–34] and magnetic resonance imaging (MRI) [35–37]). These computerized analysis schemes are being developed to aid in detection of lesions and distinguishing between malignant and benign lesions in order to improve both sensitivity and specificity of detection. Many studies have shown that CAD has the potential to increase the sensitivity [23,25,38,39] and the specificity [40,41] of diagnostic imaging. The merit of computer-aided analysis of image features lies in the objectivity and reproducibility of the measures of specific features. There is a large body of knowledge developed over the past 10-15 years with new computer techniques and new image processing methods being applied frequently to new imaging modalities [25,26,28,34,42–45]. For example, mammography CAD researchers have found that different types of lesions require different approaches to detection [46]. A wide range of feature extraction, feature selection, and classification methodologies have been explored [46–49].In general, CAD systems attempt to quantify image features associated with pathology and analyze these data to provide the physician with an objective assessment of disease state. The conventional paradigm envisions that the CAD output will be used by the physician as a "second opinion" with the final diagnosis to be made by the physician [42].

Although the specialized intestinal metaplasia associated with Barrett's esophagus is readily distinguishable from normal esophageal mucosa using EOCT [12,13,20–22], discerning dysplasia within Barrett's esophagus is more challenging [16–19]. CAD techniques may aid the effort to identify dysplasia in BE by providing objective and quantitative interpretation of clinical EOCT images reducing inter- and intra-observer variability and by detecting subtle changes in image features associated with the transformation. Furthermore, the future role of EOCT in BE surveillance is likely to involve comprehensive imaging of the involved segment of the esophagus, resulting in the acquisition of hundreds or possibly thousands of images per patient. Analysis of these data by the unaided human reader will not be practical. Therefore CAD is an essential component of the potential clinical application of EOCT to BE management.

For our clinical trials of EOCT imaging in BE, we have employed CAD methods for detecting dysplasia and classifying EOCT images of Barrett's esophagus [19,50]. Here, we expand upon these methods and describe them in detail and evaluate their suitability for detection of dysplasia in BE using EOCT, in the context of a generalizable tissue classification scheme. The analysis is extended to include multiple texture feature analysis methods as well as image features specifically designed to quantify observed EOCT image characteristics associated with dysplasia. A two-step image feature selection process is introduced to reduce risk of over-training the CAD system. Several multivariate classification methods are investigated, as well as the effects of bootstrapping and aggregating the training data. We furthermore extend the analysis to include the use of multiple image frames per examination site and investigate three-category classification.

The paper is organized as follows. Section 2 introduces the data analyzed and methods used for CAD consisting of the steps of segmentation, feature extraction, feature selection, classification and validation. The results of feature extraction, selection and classification and validation are reported in Section 3. Finally the results are discussed in Section 4.

2. Materials and methods

Under a protocol approved by the Institutional Review Board of University Hospitals Case Medical Center, we enrolled patients undergoing surveillance for Barrett's esophagus in a study designed to assess EOCT imaging. The protocol specified that surveillance should be conducted according to the "Seattle protocol" [51], with biopsies being obtained in four quadrants at two centimeter intervals along the entire length of esophagus involved by Barrett's changes.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

The EOCT system used to obtain these data has been described previously [52,53]. A digital stream of EOCT images was obtained at each biopsy site prior to removal of the actual biopsy. For this purpose, a 2.4 mm diameter EOCT probe was designed for use with a 2-channel endoscope fitted with an endoscopic mucosal resection (EMR) cap. The cap, a transparent, plastic cylinder beveled at the distal end, fits tightly on the end of the endoscope. When the tip of the endoscope was deflected toward the wall of the esophagus, a small circular portion of the esophageal mucosa was fixed by the cap, thereby negating the effects of esophageal motion. With the EOCT probe inserted through one of the two accessory channels in the endoscope, a portion of the esophageal mucosa within the area encircled by the cap was imaged. The EOCT probe and study protocol were designed such that the probe did not make physical contact with the tissue, thus avoiding compressing and altering the appearance of the mucosa [15]. As a digital stream of images was being obtained, a biopsy forceps inserted through the second endoscope channel was used to obtain a specimen. With the cap properly aligned on the tip of the endoscope, the biopsy forceps entered the EOCT field of view, providing assurance that the tissue imaged by EOCT was the same tissue removed by biopsy. Using this system, EOCT images were precisely correlated to the histopathologic diagnosis at each biopsy site [17,19].

For the present analysis, we selected image-biopsy pairs from screening procedures in 33 patients. An EOCT video stream (approximately 20 frames) was recorded at each biopsy site. The images selected at each biopsy site were the last several images recorded immediately before the biopsy forceps entered the EOCT field of view. The frame acquisition rate of the EOCT system was 4/second. We assumed that an endoscopist could reasonably hold the EOCT probe in place for 2-3 seconds. Therefore, the maximum number of frames chosen for analysis per biopsy site in this study is 10.

The EOCT image stream for each biopsy site was reviewed by EOCT experts (XQ, MVS). Also, each biopsy was evaluated by an experienced GI pathologist (JEW). The criteria for image-biopsy pair selection for inclusion in this study were as follows: 1) the biopsy forceps entered the EOCT field of view, ensuring image-biopsy correlation, 2) the EOCT probe did not contact the tissue, 3) the biopsy was graded as non-dysplastic (ND) BE or as low-grade dysplasia (LD) or high-grade dysplasia (HD). Biopsies graded as indefinite dysplasia (IND) or as cancer were excluded. The study was intended to determine the utility of EOCT and CAD for detecting dysplasia in BE. Therefore, cancer was excluded from this analysis because cancerous tissue can be readily identified endoscopically, and IND was excluded for lack of a determined gold standard diagnosis. We expect that an eventual clinical EOCT-based CAD system would not provide a classification of IND, opting instead for the more conservative classification of dysplasia, since the diagnosis will be confirmed by histological analysis of a biopsy. Of a total of 314 biopsy sites imaged from 33 patients, 96 biopsy sites obtained from 12 patients were included in this analysis. A total of 690 EOCT images were included (421 from 63 ND biopsy sites, 202 from 26 LD biopsy sites, and 67 from 7 HD biopsy sites).

The observed characteristics of dysplasia in BE include decreased scattering of light and focal loss of mucosal structures [16–18,54,55]. We and others [55] have also observed striped patterns that are common within non-dysplastic BE EOCT images when the probe is not making contact with the tissue surface, but uncommon within dysplastic BE EOCT images. These are the image features that we selected for quantitative analysis. Figure 1 shows representative EOCT images of Barrett's esophagus without dysplasia (A), with low grade dysplasia (B), and with high grade dysplasia (C) which illustrate these image features. Within each EOCT image the region of interest (ROI) representing the site of the biopsy was segmented for analysis using a segmentation method reported previously [19], removing the EMR cap and tissue outside of the cap. A binary mask was created from the ROI to select the region to be analyzed by the feature extraction algorithms. Figure 2 shows an EOCT image of Barrett's esophagus, the segmented ROI and its binary mask.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

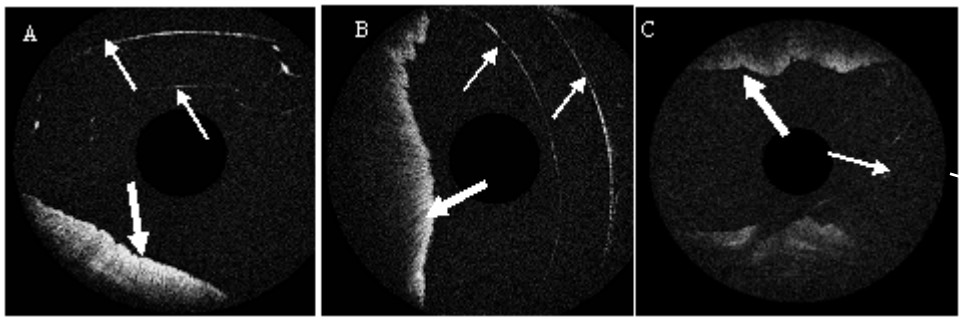


Fig. 1. Representative EOCT images of Barrett's esophagus without dysplasia (A), with low grade dysplasia (B), and with high grade dysplasia (C). Artifacts from the EMR cap are indicated by narrow arrows. Broad arrows indicate the tissue under analysis. Non-dysplastic tissue in Fig. 1(A) shows high intensity, more stripes and more local structure within the ROI. Low-grade dysplastic tissue in Fig. 1(B) shows lower intensity, less striping, and less local structure compared with ROI in Fig. 1(A). High-grade dysplastic tissue in Fig. 1(C) shows still lower intensity, no apparent stripes and still less local structure compared with Fig. 1(A) and 1(B).

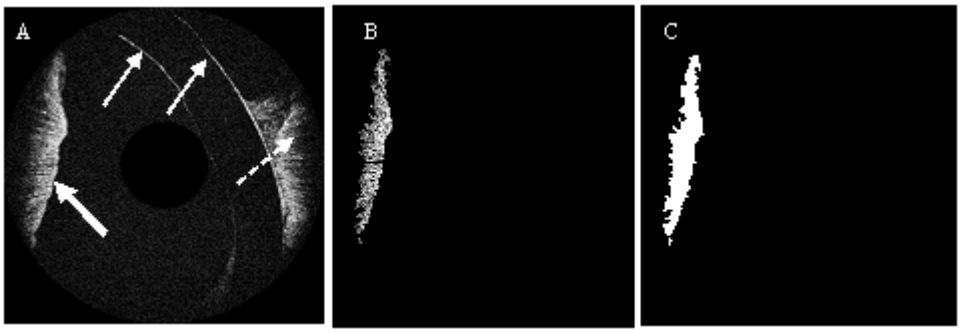


Fig. 2. EOCT image of Barrett's esophagus (A), segmented ROI (B) and its binary mask (C). Figure 2(A). shows an unprocessed EOCT image of Barrett's esophagus including noise and cap artifacts (narrow arrows), tissue outside the cap (dashed arrow) and tissue under analysis (broad arrow); Fig. 2(B). shows the segmented ROI with the background noise and artifacts removed; Fig. 2(C). shows the segmented ROI represented as a binary mask.

As described in detail hereafter, a total of 39 image features were quantified from the segmented ROI within each EOCT image. From those 39 image features, 18 were selected as candidates to be used for classification based on their 2-category classification power (using ROC analysis). To reduce risk of over-fitting and to remove highly correlated features, the first 5 principal components of the 18 selected features, accounting for 94% of the variance, were used as the final features for image and site classification. All image analysis and statistical analysis presented here was carried out using signal processing and image processing toolboxes in MATLAB R2006b (The Math Works Inc,. Natick, MA, USA), and S-PLUS 7.0 (The Insightful Corp., Seattle, WA, USA).

2.1 Feature extraction

2.1.1 Image intensity: single-scattering model

One observed characteristic of dysplasia in BE is decreased scattered light intensity, presumably resulting from the altered optical scattering properties of the dysplastic tissue. Dysplasia is characterized by cellular changes, including the size, shape and density of nuclei within the epithelial layer. These changes affect the light scattering by the tissue. The profile of each axial line (A-scan) in the EOCT image is related to the backscattered power as a function of depth and the optical properties of the tissue. In order to quantify the scattering

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

intensity by a tissue site, we estimated the optical properties of the tissue from the EOCT images. The single scattering model used here assumes that light attenuation occurs along the path of the wave both before and after it is backscattered, that higher-order scattering processes (multiple scattering) can be neglected, and that the beam was small and had a low beam-divergence angle. Under these assumptions, the backscattered power P(z) returning from a depth z can be approximated by the method suggested by Schmitt et al as Eq. (1) [56,57].

$$P(z) = P_i A(z) \frac{L_c}{2} \mu_b \exp(-2\mu_t z) \quad (1)$$

Here P_i is the incident power measured at depth z = 0; A(z) is the beam divergence function; L_c is the coherence length of the light source; μ_b is the backscattering coefficient of the tissue; and μ_t is attenuation coefficient of the tissue. The backscatter and attenuation coefficients are reasonable image features to grossly represent tissue scattering, and therefore image backscatter intensity. The tissue backscattering and attenuation coefficients can be estimated by fitting the logarithmically compressed EOCT A-scan data to a line as a function of the depth z. Then the intercept of the line can be equated to $\ln(\frac{L_c}{2}\mu_b)$ and the slope can be equated to $-2\mu_t$, yielding quantities representing the A-scan that are direct functions of μ_b and μ_t.

In this study, each EOCT image ROI was filtered with a rolling average of 9 adjacent A-scans to reduce noise. Then, for each pixel, the intercept and slope in the direction of the A-scan were estimated within an approximately 200 micrometer window around the pixel. Under the single-scattering assumption outlined above, the EOCT data should always have a positive intercept and a negative slope. To remove the influence of bad estimates, A-scan segments with a positive slope or negative intercept were excluded from the intensity analysis, as were any segments deeper than excluded regions., The mean, standard deviation and dominant peak of the histogram of slope and intercept over the entire ROI were extracted as possible classification features representing the intensity characteristic of the image.

2.1.2 Stripe detection

Another observed characteristic of EOCT images of dysplasia in BE is a lack of the stripe-like patterns that are observed in non-dysplastic BE EOCT images. These are seldom found in dysplastic BE EOCT images. Although the physical origin of the stripes has not been proven, they are probably associated with the surface texture of the tissue. They have been observed by us and others [55] when the EOCT catheter probe is not in contact with the tissue surface. Because they are often observed in non-dysplastic tissue and seldom observed in dysplastic tissue, they may be a good feature for tissue classification. Figure 3(A) shows the obvious stripe pattern within a non-dysplastic BE EOCT image and Fig. 3(B) shows no obvious stripe pattern within a dysplastic BE EOCT image. Because the EOCT probe scans in a radial manner, the stripe-like patterns within non-dysplastic BE EOCT images appear in a radial orientation. For tissue classification, the radial stripe density and relative orientation were quantified by a local intensity gradient-based method.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

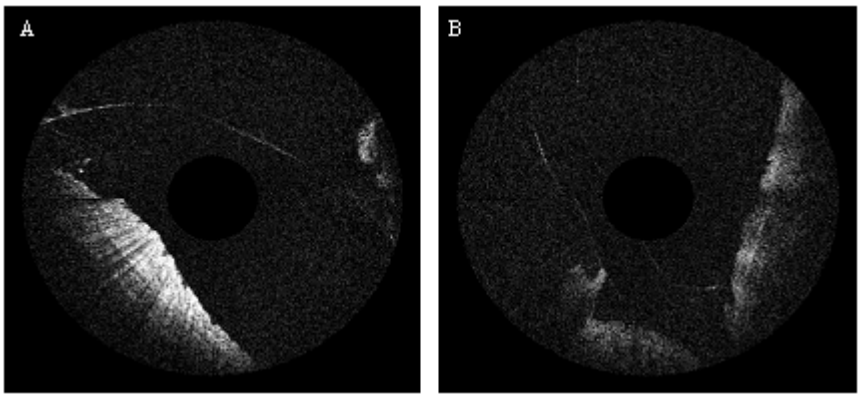


Fig. 3. Stripe-like patterns in BE. Figure 3(A) shows the obvious stripe pattern within a non-dysplastic BE EOCT image; Fig. 3(B) shows no obvious stripe pattern within a high-grade dysplastic BE EOCT image

Local orientation and density of stripes was quantified by calculating local intensity gradients in small sub-regions within the ROI of BE EOCT images [58,59]. The gradient magnitude and the angle of the stripes were calculated using a 5x5 Prewitt gradient kernel. The edges of the ROI were removed by morphological erosion to avoid edge effects. The dominant local stripe orientation was determined within 5x5 pixel sub-regions using maximum likelihood estimation [58,59]. Because the gradient magnitude and the angle within sub-regions can be described as circular data [60], the maximum likelihood of the stripe orientation within a sub-region was estimated using a von Mises distribution, which is analogous to a normal distribution for circular data [60]. The relative orientation was then determined as the difference between the local dominant angle and direction of the sub-region to center of probe. For tissue classification, three parameters were quantified: the density of stripes, defined as the number of radially-orientated stripes divided by the major axis length of the ROI, and the mean and the standard deviation of the relative orientation.

2.1.3 Texture analysis

Another observed characteristic of dysplastic mucosa in BE is loss of structure associated with normal histological organization, presumably resulting from the altered tissue architecture of the dysplastic tissue [10,17,54]. Therefore such a structure loss can be quantified as texture features, such as smoothness, coarseness and homogeneity etc. in EOCT images, which could potentially serve for tissue classification. There are three principal approaches to defining qualities of image texture: statistical methods, structural methods and model based methods [61,62]. With a statistical approach, the relationships between each pixel and its neighboring pixels are quantified by their spatial distribution of gray values. A set of local descriptors are extracted, such as energy, entropy, coarseness, contrast etc. With a structural approach, an image is considered as being composed of a set of texture units or primitives. The methods of analysis usually depend upon the geometric properties of these texture units. Model based texture analysis methods are based on the construction of an image model that can be used not only to describe texture, but also to synthesize it. The model parameters capture the essential perceived qualities of texture, such as Markov random fields, fractals et al. The fact that the perception of texture has many different dimensions is an important reason why there is no single method of texture representation which is adequate for a variety of textures. Statistical and structural approaches are often combined to extract texture features [62].

In the present study, the size, shape, and orientation of the ROI were different for each EOCT image. Model based approaches, which are sensitive to size and shape of the ROI, would not be readily adapted for texture analysis in this case. Statistical and structural texture features tend to vary locally and so these methods are not sensitive to variation in ROI size.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

Based on these observations, we chose to make use of statistical and structure approaches for texture analysis in this study of BE EOCT images. Because these EOCT image data were obtained using a radial scanning probe, rotation invariant features were investigated for texture analysis for this study. Center-symmetric auto-correlation (CSAC) [63] and co-occurrence matrices (COOC) [64,65] are statistical methods. The texture feature coding method (TFCM) [66] is a combined statistical and structural approach. We chose to investigate these three methods for this study because they capture rotation- and intensity-invariant texture features and are not sensitive to ROI size.

The CSAC method can be regarded as a generalization of Laws' kernel method [67]. It measures covariance of any local center-symmetric patterns. Two local center-symmetric auto-correlations, linear and rank-order (SAC and SRAC), together with a related covariance measure (SCOV) and variance ratio (SVR), within-pair variance (WVAR) and between-pair variance (BVAR) were calculated. All of these are rotation-invariant measures [63].

Co-occurrence matrices (COOC) (also called spatial gray-level dependence matrices) were first proposed by Haralick *et al* [64,65] and are based on the estimation of the intensity second-order joint conditional probability density functions (*pdf*) for various distances and for four specified directions (0°, 45°, 90°, and 135°) between two pixels. Texture features calculated using the COOC quantify the distribution of gray-level values within an image. For this study, four texture features were calculated from the co-occurrence matrices within the segmented ROIs, contrast, correlation, energy and homogeneity. Contrast is a measure of the gray-level variation between pairs of image elements. Correlation is sensitive to uniform and repeated structures. Energy is sensitive to image regions that have only a small number of intensity distribution patterns; it is an indicator of uniformity or smoothness. Homogeneity is sensitive to images with lower contrast values. Within the segmented ROI of each EOCT image, these four features were calculated in four different directions within a 3 pixel distance.

The texture feature coding method (TFCM) [66] is a coding scheme in which each pixel is represented by a texture feature number (TFN). The TFN of each pixel is generated based on a 3x3 texture unit as well as the gray-level variations of its eight surrounding pixels. The TFNs are used to generate a TFN histogram from which texture feature descriptors are quantified. In this work, we calculated coarseness, homogeneity, mean convergence and variance. Coarseness measures drastic intensity change in the 8-connective neighborhood. Homogeneity measures the total number of pixels whose intensity have no significant change in the 8-connective neighborhood. Mean convergence indicates how closely the texture approximates the mean intensity within a texture unit. Variance measures deviation of TFNs from the mean. Code entropy, which measures the information content of coded TFNs, was also calculated, in four Orientations; 0°, 45°, 90° and 135°.

2.2 Feature selection

From the feature extraction methods described above, a total of 39 image features were quantified within the segmented ROI of each EOCT image. Classification based on multiple image features has the advantage of increasing accuracy by increasing the amount of information used. However, making use of too many image features derived from a limited training data set increases the risk of over-fitting. This will decrease the robustness of the system when classifying data outside of the training set [68,69]. Therefore it is necessary to select a limited number of image features to balance accurate and robust classification. Here, we used a two-stage feature selection process. First, image features were not equally correlated with the histopathological diagnosis, so image features were initially screened individually to remove those that were not strong classifiers. Second, image features that are strong classifiers may be correlated with each other and therefore redundant, so redundant features were removed. For this process we used the single EOCT images immediately previous to the appearance of the biopsy forceps for each biopsy case.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

For the first feature selection stage, individual image features were screened for their ability to separate dysplastic and non-dysplastic populations. The receiver operating characteristic (ROC) curves of each feature were calculated to find the features with higher classification accuracy. Features with areas under ROC curve smaller than 0.7 were rejected. This threshold is subjective, but because it is only the first stage of feature selection, small changes of the threshold would not be expected to significantly affect the final classification result. The list of calculated image features and corresponding areas under the ROC curve is shown in Table 1.

For the second stage of feature selection, to remove redundancy due to highly correlated features, principal component analysis (PCA) was applied to the set of image features selected from stage one [70]. Each principal component is orthogonal and represents a linear combination of the original variables. The first few principal components typically account for most of the variance in the original data. Therefore, the first five principal components were selected as the parameters to be input to the multivariate classification algorithms. From the loading of the first five principal components, image features accounting for the most variance are shown in Fig. 4.

2.3 Classification and validation

Multivariate analysis was applied to potential image classification parameters to understand the relationships between the variables, and to predict generalizability to new data. The image parameters analyzed were the first five principal components of selected features from the single images previous to forceps appearance at each biopsy site, as described above. We carried out and compared the performance of four types of multivariate analysis for EOCT image classification of dysplasia in BE: linear and quadratic discriminant analysis (LDA & QLA), K-nearest neighbor (*k*-NN) classification, two types of neural network (NN) (single-hidden-layer NN (SLNN) and learning vector quantization (LVQ) network), and classification tree. Here, a binary classification was used into two categories, non-dysplastic (ND) and dysplastic (D). The dysplastic category includes both low-grade and high-grade dysplastic sites. Because of the moderate sample size (12 patients, 96 biopsy sites), to validate the classification based on the selected parameters, we applied leave-one-patient-out cross-validation to calculate the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy with their corresponding 95% confidence intervals (CI).

Fisher [71] introduced a linear discriminant analysis (LDA) seeking a linear combination of the variables (here image features) with a maximal ratio of the separation of the class means to the within-class variance, in which extreme values are obtained from eigen values and eigenvectors. Bayesian classification with pooled covariance matrix and equal prior probabilities for each group was used as the rule for classification. Quadratic discriminant analysis (QDA) is closely related to LDA, where the boundaries of the decision regions are quadratic surfaces in feature space instead of linear plane for LDA. The advantage of LDA and QDA is that they are easily calculated. But LDA and QDA require normality assumptions for training data sets.

The *k*-nearest neighbor algorithm (*k*-NN) [72] is a method for classifying objects based on the closest training samples in the feature space. The training samples are mapped into multidimensional feature space, and partitioned into regions by class labels. Testing data is assigned to the class *c* if it is the most frequent class label among the *k* nearest training samples, usually using Euclidian distance. The best choice for *k* depends upon the data; generally, larger values of *k* reduce the effect of noise on the classification, but make boundaries between classes less distinct. A good *k* can be selected by cross-validation, running the nearest neighbor classifier on the learning set only. Because of the simplicity and because no normality assumption is required, *k*-NN classification has been widely used in pattern recognition.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

A neural network [73] is a computer-intensive, algorithmic procedure for transforming inputs into desired outputs using highly connected networks of relatively simple processing units (neurons or nodes). Many different NN architectures can be employed for pattern recognition. SLNN implements the well-known statistical techniques of linear regression and generalized linear models. Here we used the independent logistic sigmoidal activation function applied to each of the outputs independently. The classification of image features using LVQ [74] is based on a comparison with a number of so-called prototype vectors. LVQ is composed of a competitive layer and a linear layer. Vector quantization is one example of competitive learning. The competitive layer automatically learns to classify input vectors in a supervised manner. In vector quantization, the network is given by prototypes and the changes of the weights of the network related to the *k*-nearest neighbor algorithm (*k*-NN). The linear layer transforms the competitive layer's classes into target classification defined by target results.

The classification tree is another non-parametric classification method. The root is the top node of the tree, and features are passed down the tree with decisions being made at each node until a terminal node or leaf is reached. Each non-terminal node contains a question on which a split is based. Each leaf contains the label of a classification [75]. A classification tree partitions the feature space into a set of sub-regions corresponding to the leaves, since each image will be classified by the label of the leaf it reaches. Thus decision trees can be seen as a hierarchical way to describe a partition of feature space. Using tree methods for classification results in a series of logical if-then conditions (tree nodes), it is conceptually simple yet powerful, and provides a structured description of the feature space [76]. Here we applied entropy [75] as a measure of impurity of nodes to construct the whole classification tree.

The advantage of LDA and QDA is that they are easily calculated. But LDA and QDA require normality assumptions for training data sets. The performance of *k*-NN varies according to the choice of *k*. The best choice for *k* depends upon the data; generally, larger values of *k* reduce the effect of noise on the classification, but make boundaries between classes less distinct. Neural networks based on artificial intelligence principles have the powerful capacity to define a non-linear separation of classes. Almost any finite-dimensional vector function on a compact set can be approximated. However, training NN are computationally expensive and can suffer from "interference" in that new data can cause it to forget some of what it has learned on old data.

Preliminary comparison of the results of these classification methods (shown in Tables 2, 3, and the first row of Table 4) reveal that the classification tree achieved classification accuracy superior to k-NN and neural network methods, and equivalent as to LDA and QDA methods. Moreover, unlike LDA and QDA, the classification tree does not assume normality nor that the underlying relationships between the predictor variables and the dependent variable are linear or quadratic. Therefore, we employed the classification tree method for subsequent investigations presented here.

One major problem with classification trees (also NN) is their high variance. Often a small change in the training data can result in a very different series of splits, making interpretation somewhat precarious. The major reason for this instability is the hierarchical nature of the process. The optimization of a classifier using a learning sample only leads to over-fitting and bias in both model selection and error rate estimation. Independent test samples are available for some problems [77], but this is usually not the case in medical statistics due to small learning samples.

Bootstrapping is a data re-sampling method that iteratively trains and evaluates a classifier in order to improve its performance. Bootstrapping is applied to estimate an approximate population distribution from the empirical distribution of the observed data, and can be implemented in the case where a set of observations can be assumed to be from an independent and identically distributed population. It constructs a number of re-samples of the

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

observed data set (of equal size to the observed data set), each of which is obtained by random sampling with replacement from the original data set [78].

Bootstrapping is becoming the most popular method of testing mediation because it does not require the normality assumption to be met, and it can be effectively utilized with smaller data sizes (such as N<20) [79]. Combining bootstrapping with classification (called bootstrapping aggregating or bagging), reduces the variance in classification [80]. The misclassification error estimated by the bootstrap with observations from the learning sample is unbiased. Bagging classification trees results in substantial reduction of misclassification error in many applications [81,82]. We classified the EOCT images based on the selected image parameters using bagging of classification trees, with 10, 30, 50, 70 and 100 times bagging. Using bootstrapped leave-one-patient-out training sets, one classification tree was constructed for each subsample. Then each test image was classified by majority voting of those bagging classification trees.

2.4 Biopsy site classification using multiple frames per biopsy site

For the EOCT image/biopsy pairs included in this study, the maximum number of frames chosen for analysis per biopsy site was 10. Not all biopsy sites had 10 frames available. In this unbalanced data set, among the 96 biopsy sites (690 EOCT images) included, the number of usable frames ranged from one to more than ten per biopsy site. Cinematic viewing of multiple EOCT images of a tissue site is extremely advantageous to a human reader classifying the site, as compared to examination of a single image. Therefore, we have investigated whether use of multiple images of a biopsy site by our CAD algorithm would improve classification accuracy. This initial investigation uses the simplest possible method of making use of multiple images.

For this analysis, biopsy sites were classified in three ways. First, sites were classified into two categories as non-dysplastic (ND) or dysplastic (D). Second, sites were classified into two categories as non-high-grade dysplastic (NHD) or high-grade dysplastic (HD). Third, sites were classified into three categories as non-dysplastic (ND), low-grade dysplastic (LD), or high-grade dysplastic (HD).

For 2-category classification, we defined V (values 1-10) as the number of frames to be considered in the classification algorithm. For each value of V, we included in the analysis only those biopsy sites for which at least V frames were available. We defined R as the number of frames (out of V) that had to be classified as positive for the biopsy site to be classified as positive. For each V, we analyzed data for values of R from one to V. For each V and R, we calculated the resultant sensitivity and specificity for classifying the biopsy site, using the pathologist's diagnosis of the biopsy as the reference standard. These data were compiled in sensitivity, specificity, and accuracy tables according to changing V and R.

For 3-category classification (ND vs. LD vs. HD), classification was calculated using classification tree with bagging using the most conservative criteria for detecting dysplasia. Within the site, if at least one HD image was found, this site was classified as HD. If no HD image and at least one LD image was found, this site was classified as LD. Otherwise the site was classified as ND.

3. Results

3.1 Feature selection

We extracted a total of 40 image features within the segmented ROIs. Eighteen image features had areas under the ROC curve that were >0.7. These features were: mean intercept from intensity model; number of stripes per unit length from stripe detection; mean SVR, VAR and BVAR from CSAC; homogeneity, CE 0°, CE 45°, CE 90° and CE 135° from TFCM; contrast 0°, correlation 45°, correlation 90°, correlation 135°, energy 0°, energy 45°, energy 90° and energy 135° from COOC. Table 1 shows the extracted image features and their areas under

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

the ROC curve. Highlighted features are those for which the areas under ROC curve were larger than 0.7.

Table 1. Extracted image features and corresponding areas under the ROC curve. Highlighted features are those for which the areas under ROC curve were larger than 0.70

Intensity Model		Stripe Measures		Texture Methods	
				CSAC	
Feature	ROC	Feature	ROC	Feature	ROC
Mean Slope	0.53	Number of Stripes per unit length	0.77	SCOV	0.57
Sd Slope	0.51			SAC	0.68
Mean Intercept	0.72	Mean Angle Difference	0.69	SVR	0.79
Sd Intercept	0.67			VAR	0.80
Dominant Slope	0.57	Sd Angle Difference	0.69	BVAR	0.80
Dominant Intercept	0.68			WVAR	0.63
Texture Methods		Texture Methods		Texture Methods	
TFCM		COOC		COOC	
Coarseness	0.69	Contrast 0^0	0.78	Energy 0^0	0.78
Homogeneity	0.74	Contrast 45°	0.68	Energy 45^0	0.77
MC	0.69	Contrast 90°	0.65	Energy 90^0	0.78
Variance	0.65	Contrast 135°	0.69	Energy 135^0	0.77
$CE0^0$	0.85	Correlation 0°	0.69	Homogeneity 0°	0.55
$CE45^0$	0.82	Correlation 45^0	0.80	Homogeneity 45°	0.62
$CE90^0$	0.82	Correlation 90^0	0.78	Homogeneity 90°	0.61
$CE135^0$	0.83	Correlation 135^0	0.79	Homogeneity 135°	0.62

In this analysis, the first principal component explained 0.69 of the variance, and the first five principal components together explained 0.94 of the variance. These were the parameters used for classification. The loading of the first five principal components is shown in Fig. 4. From this plot, we can observe that the image features contributing most to the variance of the data are: mean VAR from CSAC, CE 0° from TFCM, contrast 0° from COOC, mean intercept from intensity model and number of stripes per unit length from stripe detection. Figure 5 shows the ROC curves of the first five principal components evaluated individually as classification parameters, and each corresponding area under its ROC curve.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

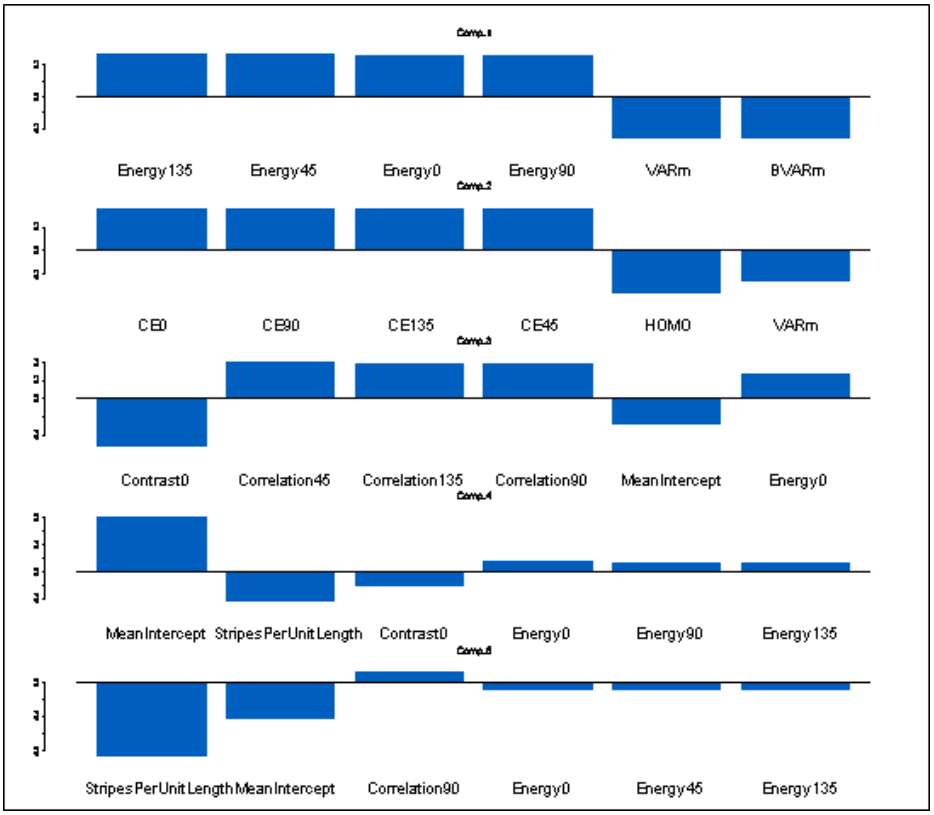
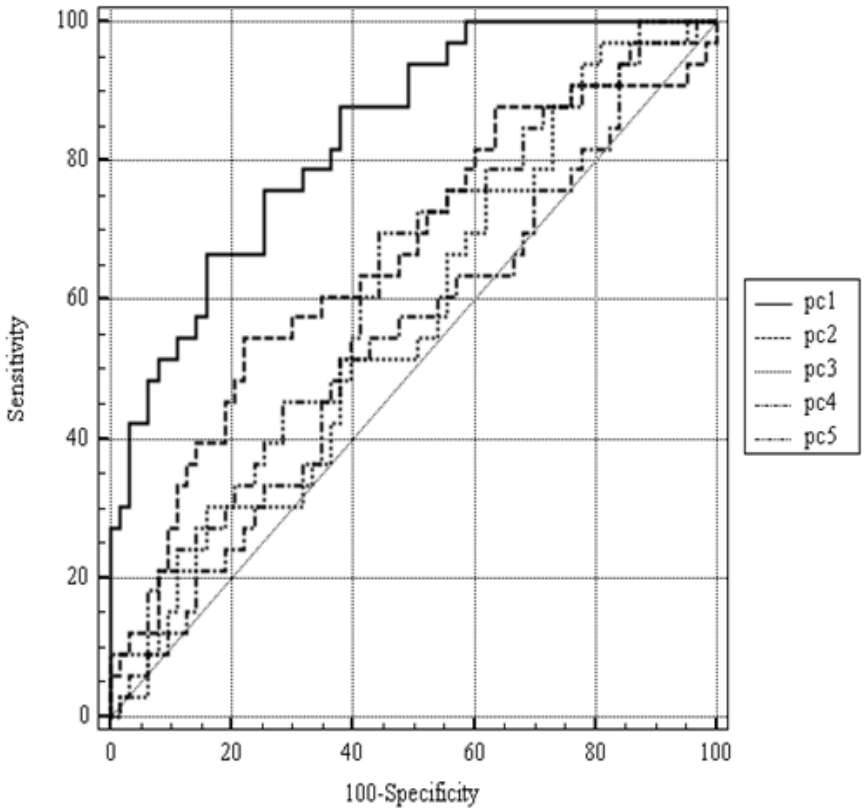


Fig. 4. Loadings plot of the first five principal components from PCA. It can be seen that the image features that contribute most to the variance present in the data are: Energy 1350 from COOC, CE 00 from TFCM, contrast 00 from COOC, mean intercept from intensity model and number of stripes per unit length from stripe detection.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010



Area under ROC curve	pc1	pc1	pc3	pc4	pc5
	0.84	0.66	0.57	0.54	0.62

Fig. 5. ROC curves of the first 5 principal components of 18 selected features and each corresponding area under its ROC curve.

3.2 Classification and validation

Table 2 reports classification results using leave-one-patient-out cross-validation by LDA, QDA, SLNN and LVQ. Table 3 reports the *k*-NN classification results for values of *k* ranging from one to six using leave-one-patient-out cross-validation. It can be seen that k-NN classification achieved the best classification of dysplasia on our EOCT data set when k = 3. Table 4 shows the classification result using the classification tree method without bagging, and with 10, 30, 50, 70 and 100 times bagging using one image per biopsy site. From these results it can be seen that misclassification errors were much higher without bagging than with bagging. At 50 times bagging, the sensitivity and specificity significantly improved. Beyond 50 times bagging, the classification accuracy did not improve further.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

Table 2. Classification of dysplasia in BE using EOCT using leave-one-patient-out cross-validation by linear and quadratic discriminant analysis (LDA & QLA), (single-hidden-layer NN (SLNN) and learning vector quantization (LVQ) network. ND: non-dysplastic; D: dysplastic; PPV: positive predictive value; NPV: negative predictive value

ND vs. D	LDA (95% CI)	QDA (95% CI)	SLNN (95% CI)	LVQ (95% CI)
Sensitivity (95% CI)	0.63 (0.47-0.78) (21/33)	0.61 (0.44-0.75) (20/33)	0.55 (0.38-0.70) (18/33)	0.97 (0.83-0.99) (32/33)
Specificity (95% CI)	0.76 (0.64-0.85) (48/63)	0.78 (0.66-0.86) (49/63)	0.78 (0.66-0.86) (49/63)	0.40 (0.29-0.52) (25/63)
PPV (95% CI)	0.58 (0.42-0.73) (21/36)	0.59 (0.42-0.74) (20/34)	0.56 (0.39-0.72) (18/32)	0.46 (0.35-0.57) (32/70)
NPV (95% CI)	0.80 (0.68-0.88) (48/60)	0.79 (0.67-0.87) (49/62)	0.77 (0.65-0.85) (49/64)	0.96 (0.80-0.99) (25/26)
Accuracy (95% CI)	0.72 (0.62-0.80) (69/96)	0.72 (0.62-0.80) (69/96)	0.70 (0.60-0.78) (67/96)	0.59 (0.49-0.69) (57/96)

Table 3. k-nearest neighbor classification results for various choices of k using leave-one-patient-out cross-validation. ND: non-dysplastic; D: dysplastic; PPV: positive predictive value; NPV: negative predictive value

ND vs. D	$k = 1$ (95% CI)	$k = 2$ (95% CI)	$k = 3$ (95% CI)	$k = 4$ (95% CI)	$k = 5$ (95% CI)	$k = 6$ (95% CI)
Sensitivity	0.30 (0.17-0.47) (10/33)	0.30 (0.17-0.47) (10/33)	0.36 (0.22-0.53) (12/33)	0.18 (0.08-0.35) (6/33)	0.21 (0.10-0.35) (7/33)	0.15 (0.06-0.31) (5/33)
Specificity	0.68 (0.56-0.78) (43/63)	0.68 (0.56-0.78) (43/63)	0.70 (0.58-0.80) (44/63)	0.70 (0.58-0.80) (44/63)	0.75 (0.63-0.84) (47/63)	0.65 (0.53-0.76) (41/63)
PPV	0.33 (0.19-0.51) (10/30)	0.33 (0.19-0.51) (10/30)	0.39 (0.24-0.56) (12/31)	0.24 (0.11-0.44) (6/25)	0.30 (0.15-0.51) (7/23)	0.19 (0.07-0.37) (5/27)
NPV	0.65 (0.53-0.76) (43/66)	0.65 (0.53-0.76) (43/66)	0.68 (0.56-0.78) (44/65)	0.62 (0.50-0.72) (44/71)	0.64 (0.53-0.74) (47/73)	0.59 (0.48-0.70) (41/69)
Accuracy	0.55 (0.45-0.65) (53/96)	0.55 (0.45-0.65) (53/96)	0.58 (0.48-0.68) (56/96)	0.52 (0.42-0.62) (50/96)	0.56 (0.46-0.66) (54/96)	0.48 (0.38-0.58) (46/96)

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

Table 4. Classification tree results using one image per biopsy site without bagging, and with 10, 30, 50, 70 and 100 times bagging using leave-one-patient-out cross-validation. ND: non-dysplastic; D: dysplastic; PPV: positive predictive value; NPV: negative predictive value

ND vs. D	Without bagging	10 bagging	30 bagging	50 bagging	70 bagging	100 bagging
Sensitivity (95% CI)	0.64 (0.47-0.78) (21/33)	0.55 (0.38-0.70) (18/33)	0.55 (0.38-0.70) (18/33)	0.67 (0.50-0.80) (22/33)	0.67 (0.50-0.80) (22/33)	0.67 (0.50-0.80) (22/33)
Specificity (95% CI)	0.76 (0.64-0.85) (48/63)	0.81 (0.69-0.89) (51/63)	0.94 (0.84-0.98) (59/63)	0.90 (0.80-0.96) (57/63)	0.90 (0.80-0.96) (57/63)	0.90 (0.80-0.96) (57/63)
PPV (95% CI)	0.58 (0.42-0.73) (21/36)	0.60 (0.42-0.75) (18/30)	0.82 (0.61-0.93) (18/22)	0.79 (0.60-0.90) (22/28)	0.79 (0.60-0.90) (22/28)	0.79 (0.60-0.90) (22/28)
NPV (95% CI)	0.80 (0.68-0.88) (48/60)	0.74 (0.66-0.86) (51/66)	0.80 (0.69-0.87) (59/74)	0.84 (0.73-0.91) (57/68)	0.84 (0.73-0.91) (57/68)	0.84 (0.73-0.91) (57/68)
Accuracy (95% CI)	0.72 (0.62-0.80) (69/96)	0.72 (0.62-0.80) (69/96)	0.80 (0.71-0.87) (77/96)	0.82 (0.73-0.89) (79/96)	0.82 (0.73-0.89) (79/96)	0.82 (0.73-0.89) (79/96)

3.3 Site classification using multiple frames per biopsy site

Results of using multiple frames per biopsy site for binary classification of non-dysplastic sites (ND, negative) vs. dysplastic sites (D, positive) are summarized in Table 5. The table presents the sensitivity and specificity of classification using number of frames (V) ranging from 1 to 10, and numbers of positive images required to classify the site as positive (R) ranging from one to V. The table also shows how many of the biopsy sites were included in the analysis for each value of V (i.e. number of sites for which enough OCT images were available). For V = 1 and R = 1, sensitivity was 0.76 (95% confidence interval (CI): 0.59-0.87) and specificity was 0.79 (95% CI: 0.68-0.88) and accuracy was 0.77 (95% CI: 0.68-0.84). Using ten frames per biopsy site and requiring at least four frames to be positive to classify the site as positive (V = 10, R = 4) yielded a sensitivity of 0.78 (95% CI: 0.54-0.92) and a specificity of 0.90 (95% CI: 0.67-0.98) and accuracy was 0.84 (95% CI: 0.68-0.93). This was the best classification result for this experiment. In this data set, 39% of the sites had ten frames available.

Results using multiple frames per biopsy site for binary classification of non-high-grade dysplastic (NHD) vs. high-grade dysplastic (HD) sites are summarized in Table 6. Using V = 1 and R = 1 yielded a sensitivity of 0.14 (95% confidence interval (CI): 0.06-0.31) and a specificity of 1.00 (95% CI: 0.93-1.00). The best classification result was obtained using V = 9, R = 1 or 2, which yielded a sensitivity of 0.67 (95% CI: 0.46-0.85) and a specificity of 1.00 (95% CI: 0.83-1.00). 44% of the sites had at least nine frames available. From the results shown, using multiple frames per biopsy site classification achieved higher sensitivity and specificity than using only one single frame per biopsy site. Table 7 contains positive predictive values (PPV) and negative predictive values (NPV) for the V = 1 and R = 1 cases and the best cases from the ND vs. D results and the NHD vs. HD results, using multiple frames per biopsy site. Three-category classification (ND sites vs. LG sites vs. HD sites) results are shown in Table 8. Table 8 shows the raw classification results. Using the conservative criteria for multi-image classification, no high-grade dysplastic biopsy sites were misclassified as low-grade or non-dysplastic. Furthermore, no low-grade dysplastic biopsy sites were misclassified as non-dysplastic. However, several non-dysplastic sites and low-grade sites were misclassified as high-grade dysplastic. Table 9 shows the calculation time for each module using MATLAB R2006b on an Intel CPU with 3.25 GB of RAM computer.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

Table 5. Sensitivity and specificity of binary classification of non-dysplastic sites vs. dysplastic sites in BE for varying numbers of images per site used for classification (V) and numbers of frames out of V that must be classified as positive for the biopsy site to be classified as positive (R) using leave-one-patient-out cross validation

Sensitivity Specificity	R = 1	R = 2	R = 3	R = 4	R = 5	R = 6	R = 7	R = 8	R = 9	R = 10	Number of usable sites (of 96)
	0.76										
V = 1	0.79										96
	0.82	0.49									
V = 2	0.65	0.94									95
	0.88	0.64	0.46								
V = 3	0.59	0.81	0.98								92
	0.87	0.71	0.52	0.32							
V = 4	0.58	0.71	0.94	0.98							79
	0.90	0.72	0.55	0.45	0.28						
V = 5	0.53	0.71	0.89	0.96	1.00						74
	0.89	0.78	0.74	0.48	0.48	0.26					
V = 6	0.54	0.68	0.85	0.93	0.98	1.00					68
	0.88	0.76	0.72	0.56	0.44	0.36	0.24				
V = 7	0.51	0.69	0.89	0.91	0.94	0.97	1.00				60
	0.91	0.76	0.76	0.71	0.52	0.38	0.33	0.24			
V = 8	0.46	0.58	0.85	0.85	0.92	0.92	0.96	1.00			47
	0.90	0.74	0.74	0.73	0.60	0.42	0.320.	0.26	0.16		
V = 9	0.44	0.57	0.83	0.84	0.83	0.91	91	1.00	1.00		42
	0.89	0.83	0.78	**0.78**	0.67	0.56	0.39	0.33	0.28	0.17	
V = 10	0.42	0.60	0.79	**0.90**	0.90	0.95	0.95	0.95	1.00	1.00	37

Table 6. Sensitivity and specificity of binary classification of non-high-grade dysplastic sites vs. high-grade dysplastic sites in BE for varying numbers of images per site used for classification (V) and numbers of frames out of V that must be classified as positive for the biopsy site to be classified as positive (R) using leave-one-patient-out cross validation

Sensitivity Specificity	R = 1	R = 2	R = 3	R = 4	R = 5	R = 6	R = 7	R = 8	R = 9	R = 10	Number of usable sites (of 96)
	0.14										
V = 1	1.00										96
	0.43	0.14									
V = 2	0.99	1.00									95
	0.43	0.14	0.14								
V = 3	0.97	1.00	1.00								92
	0.43	0.29	0.14	0.14							
V = 4	0.94	0.97	1.00	1.00							79
	0.43	0.43	0.29	0.14	0.14						
V = 5	0.94	0.99	0.99	1.00	1.00						74
	0.57	0.43	0.29	0.14	0.14	0.14					
V = 6	0.93	0.98	0.98	1.00	1.00	1.00					68
	0.57	0.43	0.29	0.29	0.14	0.14	0.14				
V = 7	0.94	0.96	0.98	1.00	1.00	1.00	1.00				60
	0.67	0.67	0.33	0.33	0.17	0.17	0.17	0.17			
V = 8	0.95	0.98	0.98	1.00	1.00	1.00	1.00	1.00			47
	0.67	0.67	0.33	0.33	0.33	0.17	0.17	0.17	0.00		
V = 9	**1.00**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		42
	0.67	0.67	0.50	0.33	0.33	0.17	0.17	0.17	0.00	0.00	
V = 10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	37

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

Table 7. Positive predictive value (PPV) and negative predictive value (NPV) for ND vs. D classification and NHD vs. HD classification using multiple frames per biopsy site. Shown are the V = 1, R = 1 case and the best case

	PPV (95% CI)	NPV (95% CI)
ND vs. D		
V = 1, R = 1	0.66 (0.50-0.79)	0.86 (0.75-0.93)
V = 10, R = 4 (best case)	0.88 (0.63-0.98)	0.81 (0.59-0.93)
NHD vs. HD		
V = 1, R = 1	1.00 (0.17-1.00)	0.94 (0.87-0.97)
V = 9, R = 1 (best case)	1.00 (0.45-1.00)	0.95 (0.82-0.99)

Table 8. Three-category site membership (by most conservative criteria) using leave-one-patient-out cross-validation. LD: low-grade dysplasia; ND: non-dysplasia

Pathology (Reference standard)				
CAD (Predicted)		ND	LD	HD
	ND	41	**0**	**0**
	LD	0	8	**0**
	HD	22	18	7

Table 9. Calculation time of each module using MATLAB R2006b on an Intel CPU with 3.25 GB of RAM

Segmentation	**Feature Extraction** (unit: second)				
	Intensity Model	Stripe Detection	Texture CSAC	Texture TFCM	Texture COOC
6.09	24.23	32.88	2.19	2.74	0.68
Classification					
k-NN	LDA	QDA	LVQ	SLNN	Tree
1.52	0.26	0.27	1540.20	40.01	4.15

4. Discussion and conclusion

The EOCT system used here and the experimental protocol were designed to acquire EOCT images precisely correlated with biopsies. An EOCT for clinical surveillance will comprehensively image the entire affected esophageal wall, not only individual locations [83,84] The image analysis presented here should be translatable to such comprehensive imaging. Based on these results, it is apparent that CAD has the potential to aid in detecting the presence or absence of dysplasia in surveillance of large surface areas of Barrett's mucosa using EOCT.

Automatic segmentation of the region of interest (ROI) for analysis will be important for surveillance of dysplasia in Barrett's esophagus using EOCT because of the large number of images acquired. The segmentation of ROI for comprehensive imaging will be simplified compared with the segmentation method used in this work [19] because the images will not include artifacts such as the EMR cap or tissue outside the cap.

Our observation of stripe-like patterns that are usually present in non-dysplastic BE EOCT images, but usually not present within dysplastic BE EOCT images, is supported by a study conducted by Chen et al [55], where OCT images were recorded of pinch biopsies taken during BE surveillance procedures. However, these stripe-like patterns are not observed in some BE images recorded using other EOCT systems [10,85] which made use of probes in physical contact with the tissue. Furthermore, similar stripe-like features are observable in other pitted gastrointestinal mucosae such as stomach and colon (e.g. Figure 4 in [10] and Figs. 1, 2 in [11], and Fig. 6 in [86].) Because the feature is observed when the probe in not in contact with the tissue, but not observed when the probe is in contact, and because the stripes are typically oriented in the direction of the illuminating probe light, it is likely that the stripes are associated with the surface topology related to the pit structure of the mucosa. The effects of probe-tissue contact on OCT detection of dysplasia within BE is currently under investigation, but it is apparent that contact obscures this feature by smoothing surface topology.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

Of the 39 image features quantified in this study, 18 (46%) were selected for classification because they were found to be significant classifiers (area under ROC > 0.7) in the initial screen. The highest individual area under the curve is 0.85. Several features not selected have areas under the curve close to 0.7. From the principal component analysis, the first five principal components account for 94% of the variance within the selected features. This indicates that there is a high level of correlation amongst the calculated features . From the loading plots shown in Fig. 4, It can be seen that the first three principal components (PC 1-3) are composed mostly of texture features. PC 1 is dominated by CSAC features VAR and BVAR and the COOC energy features. PC 2 is dominated by TFCM features besides CSAC feature VAR. PC 3 consists of the COOC contrast and correlation features. PC 4 and PC 5 are composed primarily of the mean intercept and the number of stripes, respectively, with little contribution from texture features. Therefore it is apparent that while there is a high level of correlation among texture features, most of this correlation is between the various features calculated using the same method. The features calculated from the attenuation model and the stripes feature are largely independent. Whether the specific image features that appeared as the strongest classifiers in t his study also emerge as the strongest classifiers in future studies, using different EOCT systems and imaging protocols, is unclear. For example, improved resolution or signal to noise ratio can be expected to improve classification accuracy. However, the feature extraction methods described here all yielded strong classifiers, and the feature selection methods applied here are generalizable.

In this study, classifying images using bagging of classification trees [87] reduced variance and misclassification error due to outliers. Without bagging, misclassification errors were much higher than with bagging; for 50 times bagging of classification trees, sensitivity and specificity significantly improved results without bagging. Additionally, in this study, each feature was equally weighted. In the future, once good training data is built, different weights can be applied to each feature when using tree methods, which may improve diagnostic accuracy.

The method of using multiple images evaluated in this study is the simplest and most conservative possible approach. The results shown in Table 5 and Table 6, show that making use of multiple images from a site has the potential improve the sensitivity and specificity of classification. Compared with the results of using one image per biopsy site with bagging (Table 4), the sensitivity of using multiple frames per biopsy site for binary classification of non-dysplastic sites vs. dysplastic sites is improved significantly (0.78 vs. 0.67), however both approaches achieved similar accuracy (0.84 vs. 0.82). This is because the data are unbalanced, with more negative sites than positive sites so that the improved sensitivity does not strongly affect the overall accuracy. The improved sensitivity is reflected in the PPV, however, which is improved from 0.79 to 0.88 by use of multiple images at a classification site. Making use of multiple frames makes sense intuitively; a human reader much more easily interprets a cine-stream of EOCT images than a single image. Also, use of multiple frames may support investigation of 3D image features when 3D data are available. More sophisticated methods should make better use of the information present in multiple images at a site or of 3D image sets.

Results of CAD classification in this study must be compared cautiously with previously-reported classification by endoscopists [17] or with our previously reported CAD [19]. That said, in [17], the EOCT system with human readers resulted in an accuracy of 78% for detection of dysplasia in patients with Barrett's esophagus compared with 82% accuracy presented here. In this study, image selection criteria for CAD were different than those of [17]. Of a total of 405 image-biopsy pairs collected, 96 met inclusion criteria for the present study. But for each image and biopsy pair used in the present study, there was a high degree of certitude that the biopsy forceps entered the EOCT field of view, thereby ensuring nearly perfect image-biopsy correlation. On the other hand, the endoscopists grading the images in the previous study had access to the real-time image stream consisting of many images at each

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

biopsy site, while the CAD algorithm in this study analyzed each image independently before grouping them according to biopsy site. However, the results presented here still indicate that CAD has the potential to be at least as accurate as humans for identifying dysplasia in EOCT images of BE.

The set of data analyzed in this report is similar to the set analyzed in our previous CAD report [19] because they originate from the same clinical study, but they are not the same data set because the current set was selected blindly using criteria based only on image quality so it would not be biased by previous classification results. For comparison we analyzed the current data set using the method reported in [19] and found that the classification accuracy is worse for the current data set than for the set analyzed in [19]. In other words, the current data set is noisier than that used in [19]. Therefore, while the methods reported here offer improvement over those reported in [19], this improvement is not reflected clearly in a direct comparison of the best reported values of sensitivity and specificity. However, the comparative analysis of methods reported here do make clear a number of useful observations such as: Multivariate classification is likely more robust and accurate than single parameter classification. The technique of bagging of training data improves classifier performance. Use of multiple EOCT images at a single examination location improves classification accuracy over use of a single image. Three-category classification was achieved with very low false negative rate using a classification tree and multiple images per site.

In conclusion, we present a generalizable method for developing CAD classifiers for EOCT, and the results presented here demonstrate that CAD has the potential to accurately detect the presence or absence of dysplasia for surveillance of Barrett's esophagus using EOCT. CAD quantifies classification criteria, eliminating inter-observer variability and potentially allowing further stratification.

#132754 - $15.00 USD Received 2 Aug 2010; revised 7 Sep 2010; accepted 7 Sep 2010; published 9 Sep 2010

Appendix

Nomenclature
BE: Barrett's esophagus
GI: gastrointestinal
ND: non-dysplastic
D: dysplasia or dysplastic
LD: low-grade dysplasia or dysplastic
HD: high-grade dysplasia or dysplastic
NHD: non-high-grade dysplasia or dysplastic
ROI: region of interest
CAD: computer-aided diagnosis
CSAC: center symmetric auto-correlation
TFCM: texture feature coding method
COOC: co-occurrence matrix
SCOV: gray scale texture covariance
VAR: local variance
BVAR: between-pair variance
WVAR: within-pair variance
SVR: variance ratio
SAC: normalized gray scale texture covariance
TFN: texture feature number
MC: mean convergence
CE: code entropy
CI: confidence interval
ROC: receiver operating characteristic
PCA: principal component analysis
LDA: linear discriminant analysis
QDA: quadratic discriminant analysis
k-NN: *k*-nearest neighbor
NN: neural network
SLNN: single-hidden layer NN
LVQ: learning vector quantization
Bagging: bootstrap aggregating
PPV: positive predictive value
NPV: negative predictive value

Acknowledgments

The authors acknowledge the contributions of Brian Wolf, and Matthew Ford and the financial support of the National Institutes of Health (CA94304 and CA114276). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.