Automatic Classification of Endoscopic Images for Premalignant Conditions of the Esophagus

Davide Boschetto[a,b], Gloria Gambaretto[b], and Enrico Grisan[b]

[a]IMT School for Advanced Studies Lucca, Lucca, Italy
[b]Department of Information Engineering, Universitá degli Studi di Padova, Padova, Italy

ABSTRACT

Barretts esophagus (BE) is a precancerous complication of gastroesophageal reflux disease in which normal stratified squamous epithelium lining the esophagus is replaced by intestinal metaplastic columnar epithelium. Repeated endoscopies and multiple biopsies are often necessary to establish the presence of intestinal metaplasia. Narrow Band Imaging (NBI) is an imaging technique commonly used with endoscopies that enhances the contrast of vascular pattern on the mucosa. We present a computer-based method for the automatic normal/metaplastic classification of endoscopic NBI images. Superpixel segmentation is used to identify and cluster pixels belonging to uniform regions. From each uniform clustered region of pixels, eight features maximizing differences among normal and metaplastic epithelium are extracted for the classification step. For each superpixel, the three mean intensity of each color channel are firstly selected as features. Three added features are the mean intensity for each superpixel after separately applying to the red-channel image three different morphological filters (top-hat filtering, entropy filtering and range filtering). The last two features require the computation of the Grey-Level Co-Occurrence Matrix (GLCM), and are reflective of the contrast and the homogeneity of each superpixel. The classification step is performed using an ensemble of 50 classification tree, with a 10-fold cross-validation scheme by training the classifier at each step on a random 70% of the images and testing on the remaining 30% of the dataset. Sensitivity and Specificity are respectively of 79.2% and 87.3%, with an overall accuracy of 83.9%.

1. INTRODUCTION

Gastroesophageal tumor is one of the most common cause of death among industrialized countries, with a high mortality rate due to the difficulties of early discovery and treatment. The main precancerous lesion is known as Barretts Esophagus,[1] a condition in which any extent of metaplastic columnar epithelium that predisposes to cancer development replaces the stratified squamous epithelium that normally lines the distal esophagus above the gastroesophageal junction (GEJ). Controversies and debates about the pathogenesis, the management and the definition of this common disorder have been numerous in the last decades, and a clear consensus is still missing.[2–7] BE develops as a consequence of gastroesophageal reflux disease, which damages the distal esophagus epithelium and stimulates healing through columnar metaplasia rather than through the regeneration of more squamous cells.[8] Endoscopically, the columnar epithelium and the squamous epithelium are distinguishable: the former has a pink color with a coarse texture, while the latter exhibits pale colour and glossy appearance. An example of this difference can be seen in Fig. 1. Repeated endoscopies and multiple biopsies are often necessary to establish the presence of intestinal metaplasia, since the visual confirmation of the presence of columnar epithelium above the GEJ is not enough to diagnose BE in a patient. Generally, in addition to that, an histological confirmation of columnar metaplasia in esophageal biopsy specimens is necessary, although there is a debate on which kind of metaplasia should lead to the final BE diagnosis (namely, if goblet cells are required or not to diagnose BE with certainty).[9]

Various imaging techniques can be used to screen the status of the gastrointestinal tract, but the most used are conventional white light endoscopy (WLE) and narrow band imaging (NBI). WLE uses a conventional RGB filter, a light source and a sensor able to record high resolution images. Narrow Band Imaging, instead, uses an additional filter that splits white light into two specific lights with narrowed bandwidths (blue and green,

Further author information: (Send correspondence to Davide Boschetto)
Davide Boschetto: E-mail: davide.boschetto@imtlucca.it

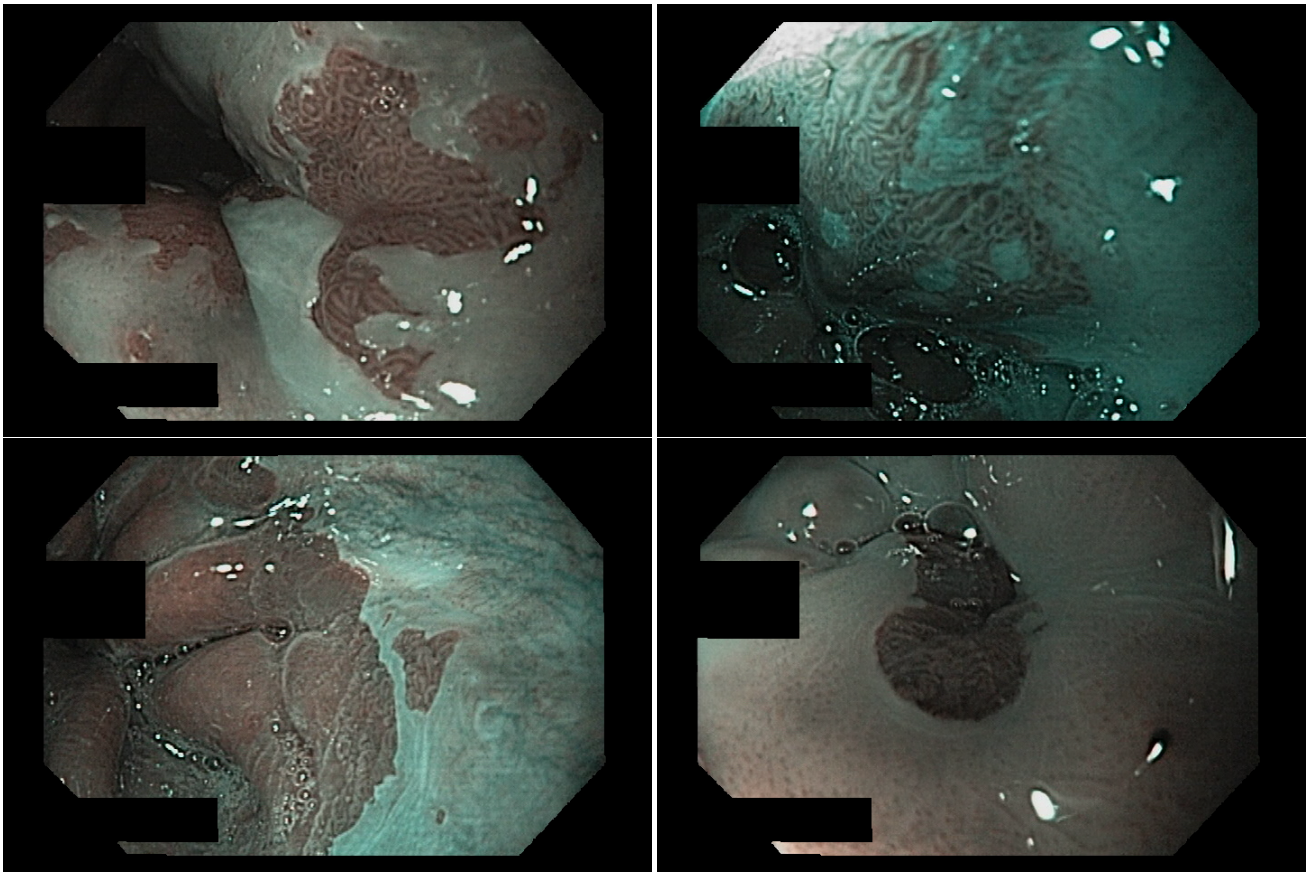

Figure 1. Four images from the dataset, showing pale squamous epithelium and pink metaplasia, masked to remove personal information.

at 400-430nm and 530-550nm respectively), canceling the contribution of the red light. In this way, since blue and green light possess more superficial penetration than the red light, pit patterns and vasculature texture are enhanced in the resulting images. Automatic classification of endoscopic images can be used by experts as a support system, narrowing down the critical regions in the images, where abnormalities are most likely to be discovered, giving the experts a chance to perform targeted biopsies, instead of performing potentially damaging and not targeted biopsies using the four-quadrant protocol. This work presents a novel local feature extraction method, by first performing a rough clustering-based segmentation using superpixels for the first time in endoscopic literature, and classifying each superpixel according to the calculated features in each of them using random forests, a well-established classification method. The main aim of the method is the classification of each superpixel as normal or metaplastic, to be used as first step in a CADSS with the purpose of helping experts grading the severity of the patients.

2. MATERIALS

In this study, 116 NBI images were obtained from clinical checkups conducted at Istituto Oncologico Veneto (IOV) in Padova, Italy, in which each patient underwent a surveillance endoscopy (Olympus CV-180). Each image's resolution is 720×480 pixels. In order to provide a ground truth, all images have been manually analyzed, providing an outline of the eventual lesion in each of the images. Typical images from the dataset and its manually defined ground truth are shown in Fig. 1 and Fig. 2, respectively.

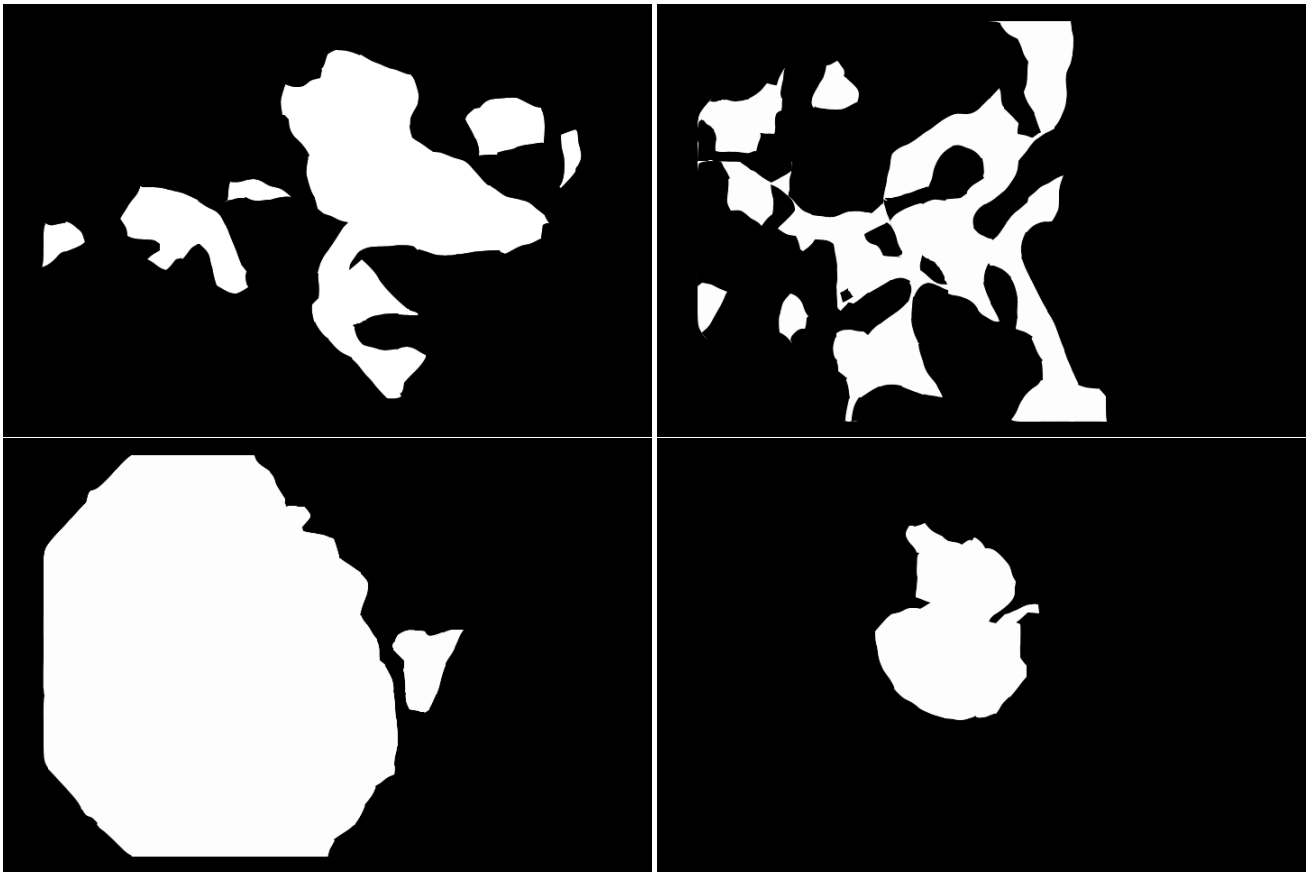

Figure 2. Manually defined ground truth for each of the four images of the dataset showed in Fig. 1. The personal information mask has not been applied, for visualization purposes.

3. METHODS

This method aims at the construction of a segmentation mask identifying a candidate region of the image with the highest possibility of being premalignant, according to its texture and color. This is performed by processing the image with a computer vision technique called superpixel segmentation, in particular using the SLIC implementation.[10] The purpose of this process is to create clusters of spatially connected pixels exhibiting similar color and texture. Each of the superpixels is then analyzed, and 8 features are extracted from each of them, to be fed to an ensemble of 50 decision trees, trained using 10-fold cross validation, with a randomly selected 70% of the dataset as training set and the remaining 30% as testing set at each step of the cross-validation procedure.

3.1 Superpixel segmentation

As pre-processing steps, each RGB image of the dataset was first edited to remove labels and personal information about each patient and exam by applying a binary mask, as shown in Fig. 1. Then, each image was normalized, transforming each color channel to zero mean and unitary variance. Segmentation via superpixel is then performed by grouping pixels into perceptually meaningful atomic regions, used to replace the rigid structure of the pixel grid. Many computer vision algorithms use superpixels as their building blocks,[11,12] given their straightforwardness and the ease of their implementation. A commonly used superpixel implementation is the Simple Linear Iterative Clustering (SLIC):[10] this implementation, based on k-means clustering, is fast to compute, memory efficient, simple to use, and outputs superpixels that adhere well to image boundaries. SLIC implementation clusters pixels of the image in the combined five-dimensional color and image plane space to efficiently generate compact and nearly uniform superpixels, imposing a degree of spatial regularization to

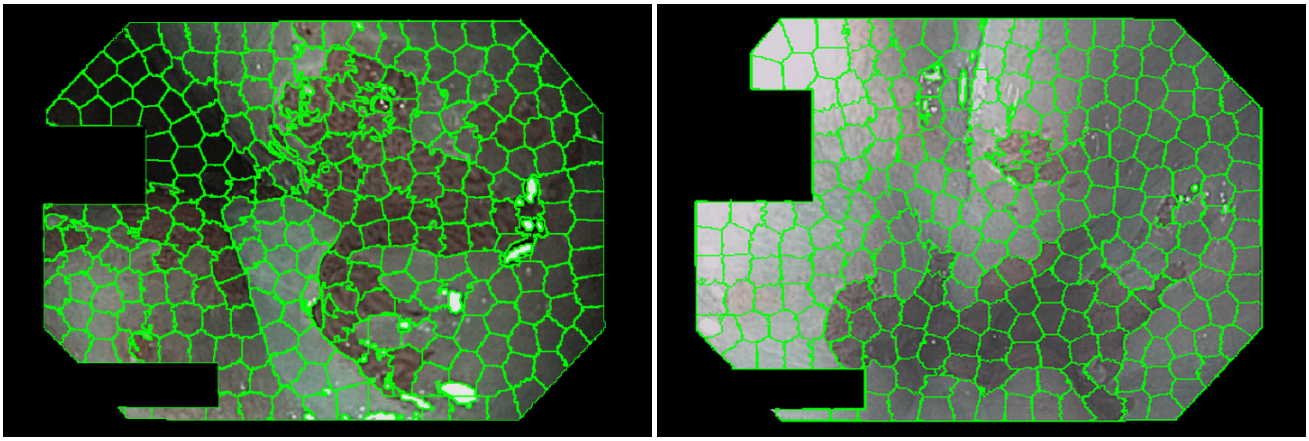

Figure 3. Two images from the dataset, with superpixel segmentation superimposed. As this figure shows, superpixels cluster together pixels exhibiting similar texture and are spatially close one another.

extracted regions. Briefly, the method starts with an initialization step, where all cluster centers are sampled on a regular grid. Centers are then moved to the lowest grayscale gradient position in a 3×3 neighborhood, to avoid centering a superpixel on an edge or on a noisy pixel. Then, a search region is defined for each center, and each pixel belonging to each region is assigned to the closest center. After this, an update step adjusts the cluster centers to be the mean five-dimensional vector of all pixels belonging to the cluster. A post-processing steps assigns disjoint pixels to nearby superpixels to enforce connectivity. Two different images from the dataset, with superpixels superimposed, are shown in Fig. 3. This step has been implemented with MATLAB R2015b, using an implementation of SLIC superpixels by *vlfeat*.[13] This technique only requires two parameters to set: the desired size of each superpixel N and a regularization parameter λ, that tweaks the smoothness of their contours.

3.2 Feature Extraction

To be able to distinguish among normal and metaplastic superpixels, some features needs to be defined on the basis of what normal and metaplastic appearance should mean. A total of 8 groups of features has been identified from each superpixel as such:

- Mean intensities I_r, I_g and I_b: color is clearly among the peculiar differences among the distinction from normal to metaplastic epithelium. Hence, the three different average intensity of the image in RGB color space were selected as features;
- Mean intensity after Top-Hat filtering, I_{th}: Morphological top-hat filtering is used to find the brightest spots in images with not uniform background, correcting the effects of an uneven illumination. Top-hat filtering results in images with smoother and darker image intensity, but with an enhancement of biological patterns of interest as vascularization and tissue rugosity. Since this filtering can only be performed to monochrome images, we extracted this feature from the red channel, the most discriminative of the three among normal and metaplastic tissue;
- Mean intensity after Entropy Filtering, I_{en}: Entropy filtering calculates, for each pixel, the entropy value of a neighborhood around that pixel. Entropy is a measure of randomness, indicating how much information (i.e., contrast, relative to greyscale intensity variations) is encoded in an image.
- Mean intensity after Range Filtering, I_{ra}: Range filtering enhances regions exhibiting sudden intensity changes among the image. Given this, this filtering emphasizes patterns and borders of the object depicted in the images. Range and entropy filtering are useful, since the two extracted features are statistical measures of the texture of an image, providing information about the local variability of their pixels.

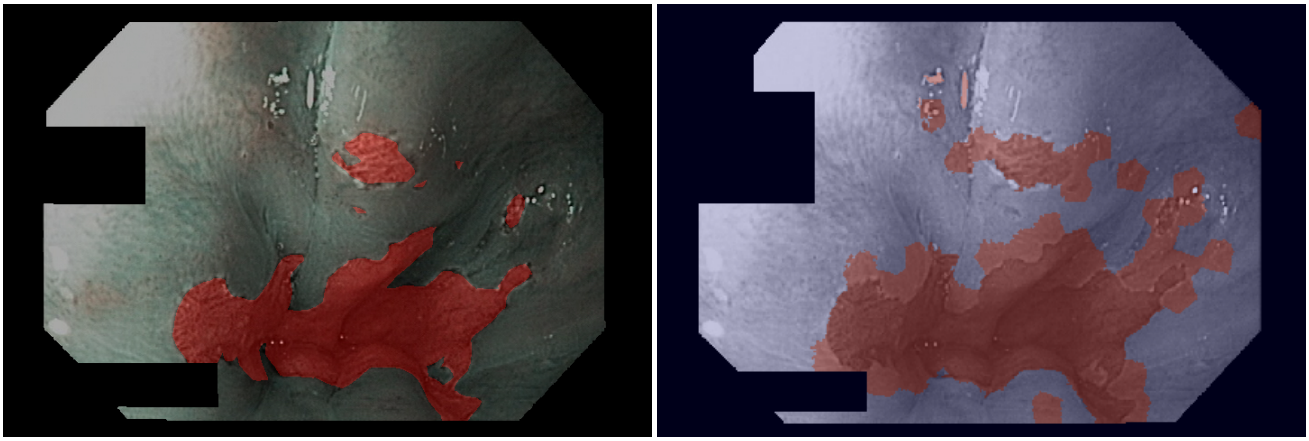

Figure 4. Comparison among metaplastic region according to ground truth (top) and our algorithm (bottom).

- Contrast C_S and Homogeneity H_S from the Gray Level Co-Occurrence Matrix (GLCM): GLCM is a statistical method of examining texture considering the spatial relationship of pixels. It calculates how often pairs of pixels with specified values and spatial locations occur in an image, building a 8×8 occurrence matrix. Extracting statistical measures from this matrix provide information about the specific texture. From this analysis, contrast (local variations in the GLCM) and homogeneity (how close the distribution of the elements in the GLCM is to its diagonal values) measures have been included in the feature set.

3.3 Classification with random forests

For each superpixel, the probability of it being part of a villus fold is computed as the score of a binary random forest[14] classifier using 50 classification trees. To obtain a robust estimation of the performances of our classification method, a cross-validation approach has been chosen. For each step of the 10 folds, the random forest has been trained on a random sample consisting of 70% of the images in the dataset, and tested on the remaining 30%.

4. RESULTS

Using 116 RGB images, our method achieves 83.9% accuracy, with 79.2% sensitivity and 87.3% specificity. In Fig.4 a comparison among manually obtained ground truth and automatic results is shown for one of the images in the dataset. As shown, the detected metaplastic region strongly correlates with the labeled metaplastic mucosa. False positives are included in the metaplastic mucosa due to light reflectance and/or shadowing artifacts, which are more similar in texture to metaplastic tissue than to squamous epithelium.

5. CONCLUSIONS

In this work we have presented a fast, efficient, stable and reliable method for the automatic detection of metaplastic regions in NBI endoscopic images, using SLIC superpixel segmentation and random forest classification. This work represents the first step of a bigger pipeline in which each region will be classified to the most probable grade of the disease, in a CADSS fashion. Moreover, this tool will be tested on different dataset of images, from different fields of medical imaging having the same needs as CLE. With more quantitative tools, experts will have a chance to perform targeted biopsies instead of using the four-quadrant standard protocol, improving the accuracy of the exam and improving the examination's accuracies.

ACKNOWLEDGEMENTS

We want to thank Istituto Oncologico Veneto (http://www.ioveneto.it) for their kindness and for providing us the dataset used in this work.

REFERENCES

[1] Barrett, N., "Chronic peptic ulcer of the oesophagus and 'oesophagitis," *B.R. J. Surg.* , 175–182 (1950).

[2] Dunbar, K. and Spechler, S., "Controversies in barrett's esophagus," *Mayo Clinic Proc.* **89**, 973–984 (2014).

[3] Falk, G., "Barrett's esophagus," *Gastroenterology* , 1569–1591 (2002).

[4] Mannath, J. and Ragunath, K., "Reflux and barrett's disease," *Endoscopy* , 34–37 (2010).

[5] Sharma, P., "Barrett's esophagus," *N. Engl. J. Med.* , 2548–2556 (2009).

[6] Ishimura, N. et al., "Barrett's esophagus: endoscopic diagnosis," *Ann. N.Y. Acad. Sci.* , 53–75 (2011).

[7] Baldaque-Silva, F. et al., "Endoscopic assessment and grading of barrett's esophagus using magnification endoscopy and narrow band imaging: impact of structured learning and experience on the accuracy of the amsterdam classification system.," *Scand J Gastroenterol.* **48**, 160–167 (Feb 2013).

[8] Spechler, S., "Barrett's esophagus and risk of esophageal cancer: a clinical review," *JAMA* **310**, 627–636 (2013).

[9] Spechler, SJ. et al., "American gastroenterological association technical review on the management of barrett's esophagus," *Gastroenterology* **310**, 627–636 (2013).

[10] Achanta, R. et al., "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans Pattern Anal Mach Intell* **34**, 2274–2282 (Nov 2012).

[11] Fulkerson, B., Vedaldi, A., and Soatto, S., "Class segmentation and object localization with superpixel neighborhoods," 670–677 (2009). Computer Vision, 2009 IEEE 12th International Conference on.

[12] Li, Y. et al, "Lazy snapping," **3**(23), 303–308 (2004). ACM Transactions on Graphics (SIGGRAPH).

[13] Vedaldi, A. and Fulkerson, B., "VLFeat: An open and portable library of computer vision algorithms." http://www.vlfeat.org/ (2008).

[14] Breiman, L., "Random forests," *Machine Learning* **45**, 5–32 (2001).