21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Automated Classification of Local Patches in Colon Histopathology *

Habil Kalkan[1,3], Marius Nap[2], Robert P.W. Duin[1], Marco Loog[1]
[1]*Pattern Recognition Laboratory, Delft University of Technology, The Netherlands*
[2]*Atrium Medical Center, Heerlen, The Netherlands*
[3]*Department of Computer Engineering, Suleyman Demirel University, Isparta, Turkey*
habilkalkan@sdu.edu.tr, M.Nap@atriummc.nl, r.duin@ieee.org, M.loog@tudelft.nl

Abstract

An automated histology analysis is proposed for classification of local image patches of colon histopathology images into four principle classes: normal, cancer, adenomatous and inflamed classes. Shape features based on stroma, lumen and imperfectly segmented nuclei are combined with texture features for classification. The classification is analyzed under the three scenarios: normal vs. abnormal, cancer vs. non-cancer and four-class classification on a labeled dataset consisting of 2000 patches per class which were collected from 55 different slices. The proposed method achieves 79.28% mean accuracy between normal and abnormal; 87.67% accuracy between cancer and non-cancer and 75.15% between the four classes with equal class priories.

1. Introduction

Worldwide, colorectal cancer (CRC) is reported as the third leading cause of cancer-related death in Western world with more than 103,000 new cases of CRC predicted in United States for 2012 [1]. Similar to other cancer types, early detection of CRC is vitally important for successful treatment and the most accurate way in diagnosis is accepted to be the histopathological analysis of the biopsy samples taken during colonoscopy. The biopsy specimen of the colon is stained and observed by pathologists under a microscope. However, visual observation, is tedious and subject to inter-observer variation [10].

The presence, level and the type of malignancy drastically changes the structures (stroma, lumina, gland, nuclei, etc) of tissue. This changing in structure are captured by texture and shape features in computational histopathology [7]. Sertel et.al [11] categorized the nerve histopathology images into stroma rich and stroma poor classes for nervous cancer prognosis using the Haralick and Local Binary Pattern (LBP) features. Doyle et. al [4] used texture features derived from Gabor filter banks and second-order statistics for detecting the prostatic adenocarcinoma. Esgiar et. al [5] integrated the fractal features to texture features and enhanced the classification results on a dataset with 44 normal and 58 benign colon tissue. Lim et. al [8] used Hough Transform based texture features on colon dataset with 25 images from normal, adenomatous and cancer classes.

In addition to texture features, shape features based on glands, nuclei and also the arrangement of the nuclei in the tissue can have diagnostic significance for some kind of malignancy in histopathology [4, 2]. Doyle et. al [4] evaluated the distribution of the nuclei based on graph features for breast histopathology with 30 cancerous and 18 benign images. It has, however, been indicated that the histological objects in the tissue may not need to be perfectly detected for good tissue classification when a comprehensive set of features is available [3].

The majority of the studies in computational histopathology are performed on either classification of benign and cancer images or grading the cancer if present [5, 3, 4]. However, besides cancer, precancerous or low-cancer risk diseases are also important to detect for successful medical treatment. Lim et. al [8] perform classification of adenomatous, cancer and normal classes from colon histopathology using 25 objects per class. Fiscor et. al [6] separated the inflammation in colon sections using 69 objects in four classes.

In this paper, we propose a study on automated classification of colon histopathology images into normal, inflamed, adenomatous and cancer classes using both textural and shape features. We performed the study

*This work is partially supported by the TUBITAK-BIDEB 2219 program

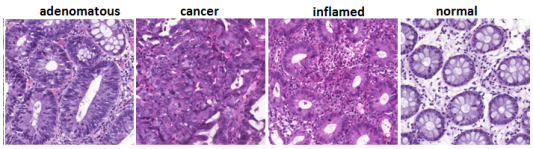

Figure 1. Patch images from four different classes.

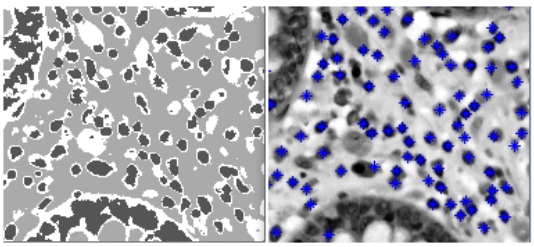

Figure 2. Segmented tissue image (left) and the detected nuclei (centers marked with blue).

on, respectively, large dataset with 2,000 samples from each class. Further details on the data employed together with the used feature extraction, selection, and classification algorithms are presented in Section 2. Experimental results and conclusions are given in Sections 3 and 4, respectively.

2. Dataset

A total of 55 H&E stained colon slices which are collected from 36 different patients at Atrium Medical Centrum are used in this study. The slices which are about 70.000×120.000 dimension are automatically divided into patches at 1024×1024 dimension. Each individual patch is labeled by an expert pathologist and assigned to one of the four classes: *normal*, *cancer*, *adenomatous* and *inflamed* (Fig. 1). In total, we use 2,000 patches per *normal*, *cancer*, *adenomatous* and *inflamed* classes, respectively, to yield a total of 8,000 patches.

3. Feature Extraction and Classification

3.1. Image preprocessing and nuclei detection

Tissue images in RGB space are segmented into background, stroma, cellular and lumina regions using k-means clustering. The mean of the clusters are every time initialized by the following set of RGB triplets: $\{(0,0,0),(0.3,0.2,0.4),(0.9,0.9,0.9),(0.7,0.5,0.7)\}$ (Fig.2). The background segment is discarded in the analysis and used for masking the patch images. The regions of the remaining three segments are used for feature extraction that is detailed below. In parallel, the RGB images are transformed into gray-scale and then normalized by stretching for nuclei detection.

Nuclei are detected by using Laplacian of Gaussian (LoG) blob detection. The size of the LoG filter needs to match the size of the nuclei and while nuclei in the colon tissue may have slight variations in scale, we use Lindeberg's blob detection algorithm with automatic scale selection over a range of scales [9]. With this approach, isolated nuclei at various sizes are detected. Beyond some limited visual inspection, we have not evaluated the accuracy of our nuclei detector. Following [3], we expect however that the detection does not have to be overly precise. In a last step, the connected nuclei which frequently occur around the glands are eliminated.

3.2. Structural features

We evaluated structural features based on stroma, cellular and lumina segments in tissue. In addition, the number of isolated nuclei in the tissue and also the arrangement of these nuclei have significant importance for malignancy detection. To capture these nuclei-based features, we divide the patches into 16 sub-patches and consider only the sub-patches containing tissue. Then the following features are extracted to get the local information in tissue: the number of nuclei per tissue area in sub-patches, the ratio of the each structures (stroma, cellular and lumen) to the tissue area and the pairwise ratio of the structures in sub-patches. The density of these structures are obtained from the segmentation result which was explained in previous section. Finally, the variance and the mean of these extracted features are used to get global descriptors for each patch image.

3.3. Texture features

The existence of malignancy and also the level of malignancy may drastically change the texture of the tissue and any change in texture may also cause a change in color distribution after H&E staining. To capture this, we evaluate 32 bins color channel histograms of R, G, B components of raw image after removing the background pixels. That makes a total of 96 color features for each patch.

In addition to color features, the raw RGB images are transformed into HSV domain and Gabor and second order statistical features are extracted on H,S,V and normalized gray value component which is also

used in nuclei detection. For second order features, we first evaluate the gray-level co-occurrence matrix, $M(\theta) \in \mathbb{R}^{N \times N}$ in four different directions where $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. Then we evaluate the following Haralick features: homogeneity, contrast, energy and correlation for each directions to yield 16 features. For Gabor features, we construct a set of Gabor filters $G(f, \theta)$ for orientation parameters $\theta \in \{0, \frac{\pi}{8}, \frac{2\pi}{8}, ..., \frac{7\pi}{8}\}$ and frequency parameters $f \in \{k, 2k, ..., 8k\}$ where k is selected as 1/16. With the inclusion of the filter at with parameters $(0, 0)$, we obtain 65 Gabor filters that are applied to the tissue images. In addition, the average, variance and minimum-to-maximum ratio of these filtered images are evaluated for the Gabor features which makes a total of 195 features for each image.

3.4. Feature selection and classification

Using the techniques described in this section, we extracted a total of 1108 texture and 19 shape features for each patch. Our feature set contains a, potentially large, number of correlated features, which may negatively affect the classification accuracy. Therefore, we selected the best 120 features by forward feature selection algorithm considering the sum of Mahalanobis distances [13] as feature subset selection criteria. We repeat the feature selection procedure for each classification problem, separately.

4. Experimental Results

The patch images may contain tissue at various density (Fig.1) and patches having respectively small tissue ratio (tissue/patch area) may not provide sufficient information for analysis. Therefore, we ignore the patches having tissue ratio less than 0.25 for the analysis. Classification performance under the following two two-class (*normal* versus *un-normal* and *cancer* versus *non-cancer*) and four-class scenarios is studied using 10-fold cross validation. The classification accuracies are reported for these three problems and for the two-class problems we also report the AUC (area under the ROC) values. Results are obtained using a *logistic regression* classifier.

4.1. Two class classification

Normal, *cancer* and the pre-cancerous classes like *inflammation* or *adenomatous* are the mainly encountered classes in colon histopathology. The *inflamed* abnormality is accepted as low cancer risk than the *adenomatous* classes [12]. These pre-cancerous cases are exposed to different medical treatment than the cancer. In reality, a pathologist labels the whole slice which includes a large number of patches. The patches in a slice may be at different classes. However, a slice having only one cancer patch leads to be labeled as cancer. In other words, a first objective in histopathology is to detect the cancer which requires urgent treatment. If no cancer is observed in a slice, then the attention is given to high cancer-risk classes, like *adenomatous*.

Another objective, however, might be the automated detection of *normal* slices as the first step. This could be used to significantly decrease the labor cost, as such slices could be discarded without the need for the pathologist to study them. Therefore, we generated two different two-class problems: *normal* versus *abnormal* to emphasize the normal patch detection and *cancer* versus *non-cancer* to emphasize the cancer patch detection. In the first problem, we construct the class *abnormal* by collecting samples from *cancer*, *inflamed* and *adenomatous* labeled patches against to *normal* class. However, for the second problem, we construct the *non-cancer* class by the collecting samples from *normal*, *adenomatous* and *inflamed* labeled patches against the *cancer* class.

Using a *logistic regression* classifier with equal class priories, accuracies of 65.35% and 93.21% are obtained for *normal* and *abnormal* classes in the first classification problem, respectively. However, the accuracies of 79.64% and 95.70% are obtained for *cancer* and *non-cancer* classes, respectively, in the second classification problem. We also measured the area under the ROC curves and found values of 0.90 and 0.95 for the *normal* vs. *abnormal* and *cancer* vs. *non-cancer* classification, respectively.

4.2. Four class classification

For the four class classification problem, we randomly divide the dataset into training and test set and present the confusion matrix over the classified test samples. In total, 890, 887, 966 and 874 patches are used for testing from the *adenomatous*, *cancer*, *inflamed* and *normal* classes, respectively.

It is observed that 48 of the 887 *cancer*-labeled patches are misclassified as *normal* and 25 of the 874 *normal*-labeled patches are misclassified as *cancer*. A higher number of mistakes is observed between *inflamed*, *normal* and *adenomatous*.

5. Discussion and Conclusion

A computational histology analysis is proposed for the classification of local patches in colon slices. Among the 1127 extracted features, we selected the best

	Estimated Labels			
	Adeno.	Cancer	Inflamed	Normal
Adeno.	667	84	59	189
Cancer	44	834	73	48
Inflamed	67	71	703	159
Normal	85	25	89	799
Total	863	1014	924	1195

Table 1. Confusion matrix with the number of the patches assigned to each class.

120 features for the analysis. It is observed that these features are selected from the shape, Haralick and Gabor features (obtained at two different scales). An analysis is performed based on two class and four class classification.

In *cancer* vs *non-cancer* classification, the false negative rate (*cancer* patches that are detected *non-cancer*) is expected to lower. However, false positives (non-cancer patches that are detected cancer) is tolerable at patch level. In contrast, at *normal* vs. *abnormal* classification, false positives (*abnormal* patches detected as *normal*) is intolerable. In both of the two class classification problem, we achieve promising AUC results which are: 0.90 and 0.95.

In the four-class classification, slightly worse results are obtained compared to the two-class problems, which can largely be explained from the simple fact that we consider more classes. Another study on multi-class colon tissue classification has been performed by [8] using 25 patches from *cancer*, *adenomatous* and *normal* classes. They performed tests with 10 patches from each class and showed that only the two of the adenomatous patches were mis-classified as normal. Similarly, Fiscor et. al [6] separated the inflammation in colon sections using 69 objects in four classes. However, the dataset they used, are not adequate for generalization in computational histopathology. Because, tissue histology is highly heterogenous and it needs more training data for generalization. Compared to [8, 6], we achieve more accurate results even in the case of four classes. The lowest errors in four-class classification are observed between the *cancer* and *normal* classes. Whereas, higher errors are observed between the *adenomatous*,*inflamed* and *normal* classes. The mean classification accuracy in four class classification is 75.15% which can be regarded poor in pattern recognition problem. However, patches in histopathology may include malignancy together with the normal structure. That explains the high error rate between the *normal*, *adenomatous* and *inflamed* classes.

In practice, a pathologist labels whole slice. Therefore, in the future, the proposed patch-based classification problem will be extended to slice-based classification to better aid to pathologist for histopathological image analysis.

References

[1] *Cancer Facts and Figures*. American Cancer Society, Atlanta, Ga, 2012.

[2] D. E. Axelrod, N. A. Miller, H. L. Lickley, J. Qian, W. A. Christens-Barry, Y. Yuan, Y. Fu, and J. A. Chapman. Effect of quantitative nuclear image features on recurrence of ductal carcinoma in situ (dcis) of the breast. *Cancer Informatics*, (4):99–109, 2008.

[3] L. E. Boucheron, B. S. Manjunath, and N. R. Harvey. Use of imperfectly segmented nuclei in the classification of histopathology images of breast cancer. *IEEE ICASSP.*, pages 666–669, 2010.

[4] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. *IEEE Int. Sym. on Bio. Imaging: From Nano to Macro*, pages 496–499, 2008.

[5] A. N. Esgiar, R. N. G. Naguib, B. S. Sharif, M. K. Bennett, and A. Murray. Fractal analysis in the detection of colonic cancer images. *IEEE Trans.on Infor. Tech. in Biomedicine*, 1(6), 2002.

[6] L. Ficsor, V. S. Varga, A. Tagscherer, Z. Tulassay, and B. Molnar. Automated classification of inflammation in colon histological sections based on digital microscopy and advanced image analysis. *Cytometry Part A*, 73A:230–237, 2008.

[7] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabbushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *IEEE Rev in Bio. Eng.*, 2, 2009.

[8] L. A. G. Lim, R. N. G. Naguib, E. P. Dadios, and J. M. C. Avila. Analysis of colonic histopathological images using pixel intensities and hough transform. *Philippine Science Letters*, 1(3), 2010.

[9] T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int. Journal of Computer Vision*, 3(11):283–318, 1993.

[10] P. G. Putten, L. Hol, H. Dekken, K. J. H, E. Ballegooijen M, Kuipers, and L. M. E. Inter-observer variation in the histological diagnosis of polyps in colorectal cancer screening. *Histopathology*, 6(58):974–981, 2011.

[11] O. Sertel, J. Kong, H. Shimada, U. Catalyurek, J. Saltz, and M. N. Gurcan. Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. *Pattern Recognition*, 42:1093–1103, 2009.

[12] T. Tanaka. Effect of quantitative nuclear image features on recurrence of ductal carcinoma in situ (dcis) of the breast. *Journal of Carcinogenesis*, 5(8), 2009.

[13] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, London, UK, 2008.