TEXTURE BASED CLASSIFICATION OF HYPERSPECTRAL COLON BIOPSY SAMPLES USING CLBP

Khalid Masood, Nasir Rajpoot

Department of Computer Science, University of Warwick, UK

ABSTRACT

Computer aided diagnosis (CAD) is aimed at supporting the pathologists in their diagnosis. In this paper, we present an algorithm for texture-based classification of colon tissue patterns. In this method, a single band is selected from its hyperspectral cube and spatial analysis is performed using circular local binary pattern (CLBP) features. A novel method for feature selection is presented resulting in the best feature set without actually running the classifier. Classification results using Gaussian kernel SVM, with an accuracy of 90%, demonstrate that texture analysis based on CLBP features is able to distinguish the benign and malignant patterns.

***Index Terms*—** Histopathology, Computer-Aided Diagnosis (CAD), Colon Biopsy Classification, Hyperspectral Imaging

1. INTRODUCTION

Colon cancer is one of the leading causes of death from cancer in the Western and industrialised world [1]. Early diagnosis of colon cancer can lead to an effective treatment. Routine screening, however, means an increased workload for the pathologists, and hence the need for development of robust automated techniques to assist the manual diagnosis [2]. Automated techniques are also necessary because there is a significant difference in the diagnosis by pathologists – due to subjectivity, fatigue or tiredness – which can potentially result in different treatment plans.

In a computer-assisted colon biopsy diagnosis setting, the task of the computer algorithm is to label a colon biopsy slide as normal or malignant. The approach in [3] uses hyperspectral images of colon biopsy slides, whereby the classification algorithm is based on spectral analysis to discriminate between normal and cancerous biopsies of the colon tissue. Tissue segmentation is supervised, in the sense that different points in a biopsy image are labelled manually. The algorithm automatically assigns a class to each labelled location by using Laplacian eigenmaps. Each biopsy slide has 40 to 60 gland nuclei after the segmentation. If a pre-determined fraction of nuclei is malignant, then the slide is classified as cancerous. However, with a diagnostic efficiency of 82% there is room for improvement before such a system can be deployed in a clinical screening.

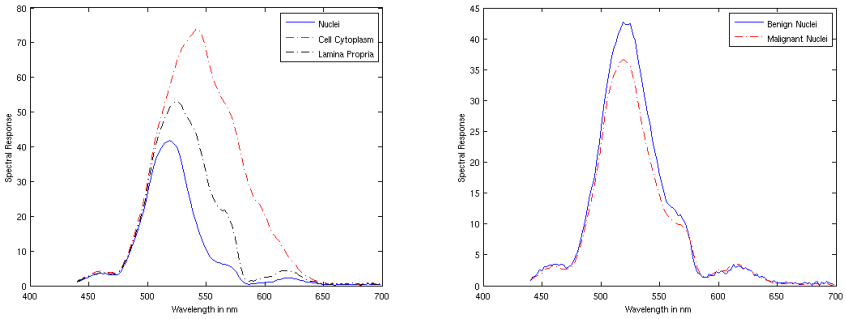
In this paper, we propose a colon biopsy classification algorithm based on spatial analysis of hyperspectral image data taken from colon biopsy samples. First, a single spectral band is selected from all the available spectral bands. Then, spatial analysis is performed on this spectral band. Using circular local binary pattern (CLBP) algorithm, patterns in the biopsy samples are represented by a feature vector. A feature selection algorithm is employed on this feature vector to save the computational cost and to discard any redundancy in the data. Classification is achieved using subspace projection methods such as principal component analysis (PCA), linear discriminant analysis (LDA) as well as support vector machines (SVM). A side advantage of performing the spatial analysis on a single band is to investigate whether comparable performance can be achieved with carefully selected spectral band to those machines which use full 3D hyperspectral data [3, 4, 5]. If so, additional acquisition, storage, and computational cost on a hyperspectral or multispectral data can be safely avoided.

The remainder of this paper is organized as follows. Section 2 describes the hyperspectral data. Textural analysis and feature selection algorithm are described in Section 3 and 4, respectively. In Section 5, experimental results are presented. Finally, conclusions are drawn from this study in Section 6.

2. THE DATASET

Our experimental dataset consists of carefully selected spectral band from the hyperspectral cubes of colon biopsy samples on a tissue micro-array. Biopsy samples are taken from 32 different patients and are stained with Hematoxylin & Eosin ($H\&E$). Hyperspectral imaging setup consists of tuned light source based on a digital mirror device which transmits any combination of light frequencies [6]. Tuned light source generates 128 wavelengths within the visual range of 440-700 *nm* with a resolution of 2 *nm*. Nikon Biophot microscope with a CCD camera is used for 40X magnification. Hyperspectral pictures of tissues obtained with the CCD camera are captured by a computer and a single data cube is collected for each biopsy. Each image cube has a spatial resolution of $491 \times 652 \times 128$ pixels.

Fig. 1(a) shows the spectral responses of nuclei, glandular cytoplasm and lamina propria of a benign tissue for all the



(a) Benign tissue components (b) Nuclei (benign and malignant tissue)

Fig. 1. Spectral response of hyperspectral images

available 128 spectral bands, while responses of benign and malignant nuclei are presented in Fig. 1(b). The spread of the benign and malignant nuclei is almost identical but benign nuclei have higher peak than malignant nuclei. It can be said that hyperspectral data is able to discriminate between different tissue parts but the spectral response of the two nuclei is not distinguishable. This motivates us to perform spatial analysis on this data for achieving reasonable classification results.

3. CIRCULAR LBP FEATURES

In our earlier work [7], 70th spectral band (588 *nm*) showed relatively high classification accuracy than other spectral bands. The same band is selected and CLBP features are computed on the band. The feature selection algorithm of Section 4 helps us select the best features.

The basic LBP operator is a gray scale invariant pattern measure complementing the texture in images [8]. In LBP method, texture is defined using local patterns on a pixel level. Each pixel is labelled with the code of the texture primitive that best matches the local neighbourhood. First, center pixel value is taken as a threshold and n neighbours of the pixel are selected. Each neighbouring pixel is given a weight based on its position and these weights are multiplied by the threshold values to generate basic LBP code. In circular LBP (which is used in our algorithm), symmetric neighbours in a circle are used for a particular radius. The reader is referred to [8] for more details of this method.

4. FEATURE SELECTION

Our feature selection algorithm is based on three measures related to the quality of clustering. These measures determine the separability and compactness of the clusters. First, Classification Scatter Index (CSI) C_o is a measure of compactness of clusters formed by a set of features [9],

$$C_o = \sum_{j=1}^{m} p_j \sigma_j \tag{1}$$

where p_j is the probability of j-th class, σ_j is the standard deviation of j-th class and m denotes the number of clusters, two in our case. We scale the values of C_o between 0 and 1, with 1 representing the most compact clustering, as follows,

$$C = e^{-aC_o^2} \tag{2}$$

where a is a positive constant.

Second, Rand index is a measure of similarity between two different clusters [10], and it ranges from zero when the two clusters are not similar to one when the clusters are exactly the same. The adjusted Rand index provides a standardised measure such that its expected value is zero when the partitions are selected at random and one, when the partitions match completely. Suppose there are two partitions of the same data with G_1 and G_2 having g_1 and g_2 clusters, respectively. The Rand index computes the proportion of n by 2 objects, n being the total number of points, that belong to both the partitions. It is defined as follows,

$$R = \frac{\binom{n}{2} + \sum_{i=1}^{g_1}\sum_{j=1}^{g_2} n_{ij}^2 - \frac{1}{2}\sum_{i=1}^{g_1}\left[\sum_{j=1}^{g_2} n_{ij}\right]^2 - \frac{1}{2}\sum_{j=1}^{g_2}\left[\sum_{i=1}^{g_1} n_{ij}\right]^2}{\binom{n}{2}} \tag{3}$$

where n_{ij} is the number of points in cluster i of G_1 that also belong to cluster j of G_2.

Third, silhouette index is a measure for each point in the clustering as to how similar that point is to points in its own cluster and to the points in other clusters [11]. Its value ranges between -1 to 1. It is defined as follows;

$$S(i) = \frac{min[D_b(i,k)] - D_w(i)}{max[D_w(i), min[D_b(i,k)]]} \tag{4}$$

where $D_w(i)$ is the average distance from the i-th point to the other points in its own cluster, while $D_b(i,k)$ is the average distance from the i-th point to points in another cluster k (it is the minimum average dissimilarity between the i-th element and any other cluster not containing the element). The silhouette coefficient S is the average of the coefficients S_i for all the points.

Taking the three indices described above, we compute a composite index I as the weighted average of C, R and S. The weights are calculated based on correlation coefficients of the indices with the clustering accuracy achieved by k-means algorithm. Composite index determines the basis for best feature selection and is defined as follows,

$$I = r_c C + r_r R + r_s S \tag{5}$$

where r_c, r_r and r_s are correlation coefficients for C, R and S respectively. CLBP(r,b) features are computed for 33 different combinations of radius r and number of neighbours b, with $r \in \{2,3\ldots,12\}$ and $b \in \{8,12,16\}$. Fig. 2 shows the surface obtained for I after sweeping through the ranges of r and b. We pick the best features sets that give the two highest values of composite indices, namely CLBP$(5,8)$ and CLBP$(5,12)$.

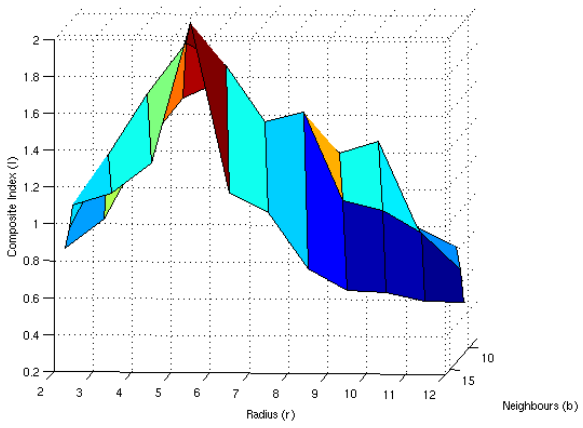


Fig. 2. Composite index I for various values of r and b

5. RESULTS AND DISCUSSION

Our dataset consists of 32 hyperspectral images of colon biopsy samples, taken from uniquely different patients, of which 24 samples are used for training and the remaining 8 samples comprise the testing set. Four-fold cross-validation is performed by repeating this process four times varying the composition of training and test datasets each time.

Experiments are conducted using three classifiers: PCA, LDA and SVM. Twenty principal components for PCA are selected empirically to yield the best classification accuracy. As there are only two classes in our problem, LDA uses only one principal component for projection of the data samples. This results in decreased performance of LDA as compared to PCA. To avoid this, we have used modular LDA approach[12]. The idea is to virtually increase the number of classes and introduce sub-classes in the training set. This is possible only if within-class scatter is large and there are enough training samples, which holds in our case. Hence, we divide our two main classes into four sub-classes. First two sub-classes are from benign samples and remaining two are for malignant samples. For SVM, Gaussian kernel is used and its kernel bandwidth is tuned. Once a kernel has been tuned at the time of training, the same kernel parameters are used in the testing stage.

Table 1 presents various classification performance measures for these classifiers using the two best CLBP feature sets. The performance measures can be expressed in terms of total numbers of true positives (TP), false positives (FP), true negatives (TN), false negatives (FN), positives (P) and negatives (N). Accuracy (Acc), Sensitivity (Sen) or Recall, and Specificity ($Spec$) are defined as follows,

$$Acc = \frac{TP+TN}{P+N}, Sen = \frac{TP}{TP+FN}, Spec = \frac{TN}{TN+FP}. \tag{6}$$

Positive Predictive Value (PPV) or Precision and Negative Predictive value (NPV) are defined as follows,

$$PPV = \frac{TP}{TP+FP}, NPV = \frac{TN}{TN+FN}. \tag{7}$$

The final measure, F-measure ($F-meas$), is the weighted harmonic mean between Recall and Precision. It is given by,

$$\text{F-measure} = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}}. \tag{8}$$

From Table 1, we observe that SVM yields best overall performance in case of CLBP$(5,12)$ whereas PCA leads the table in case of CLBP$(5,8)$ with SVM closely behind. It can also be observed that the highest sensitivity (87.5%), F-measure (90.3%) and NPV (88.2%) values are achieved by SVM with CLBP$(5,8)$. Specificity and PPV are highest (100%) for PCA using the same feature set. These results are in close agreement with our feature selection method, which returns CLBP$(5,8)$ as the best feature set. The power of non-linear classifier SVM is harnessed in case of CLBP$(5,12)$, our second best feature set, whereas PCA and SVM perform equally well in terms of the overall accuracy in case of CLBP$(5,8)$ due to PCA using a relatively large number of components and employing the best feature set.

Receiver operating characteristic (ROC) curves for both feature sets with SVM are presented in Fig. 3, which also gives area under convex hull (AUCH) of the ROC curves in both cases. The curves are obtained by varying margin of support vectors from the decision boundary. A reasonably high AUCH value of 0.9 is achieved by SVM with CLBP$(5,8)$.

CLBP$(5,12)$									
Classifier	n	k	Acc	Sen	Spec	PPV	NPV	F-meas	Ave
PCA	135	20	78.1	68.7	87.5	84.6	73.7	75.8	78.1
LDA	135	3	81.4	68.7	**93.7**	91.7	75	78.5	81.5
SVM	135	-	**87.5**	**81.3**	**93.7**	**92.8**	**83.3**	**86.7**	**87.6**
CLBP$(5,8)$									
PCA	59	20	**90.6**	81.2	**100**	**100**	84.2	89.6	**90.9**
LDA	59	3	84.4	81.2	87.5	86.6	82.3	86.6	84.3
SVM	59	-	**90.6**	**87.5**	93.7	93.3	**88.2**	**90.3**	90.6

Table 1. Performance measures (%) using the two best CLBP fetaure sets: n and k denote number of features and eigen components, respectively.

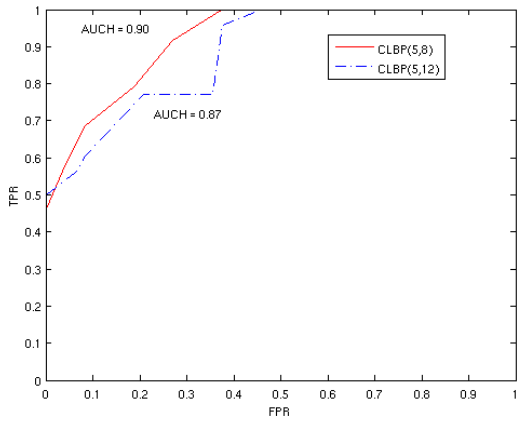


Fig. 3. ROC and AUCH for CLBP(5,8) and CLBP(5,12)

6. CONCLUSIONS

In this paper, we presented a novel algorithm for automatic colon biopsy classification. It is shown that spatial analysis of textural patterns corresponding to benign and malignant biopsies reveals the differences between them. A composite performance index, made up of a linear combination of three measures of clustering performance, is used to select the best feature set from a range of possible textural feature sets. The selected feature set yields the best performance in terms of sensitivity, specificity and area under the convex hull of the ROC curve.

7. REFERENCES

[1] R.S. Houlston, "Molecular pathology of colorectal cancer," *Journal of Clinical Pathology*, vol. 54, pp. 206–214, 2001.

[2] A. Ferrandez and J.A. Disario, "Colorectal cancer: Screening and surveillance for high-risk individuals," *Expert Review of Anticancer Therapy*, vol. 3(6), pp. 851–862, 2003.

[3] F. Woolfe, M. Maggioni, and R. Coifman et al, "Hyper-spectral microscopic discrimination between normal and malignant colon biopsies," *Technical Report, Applied Mathematics Programme, Yale University*, 2006.

[4] L.E. Boucheron, Z. Bi, N.R. Harvey, B.S. Manjunath, and D.L. Rimm, "Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery," *BMC Cell Biology*, 2007.

[5] K. M. Rajpoot and N. M. Rajpoot, "SVM optimization for hyperspectral colon tissue cell classification," *Proceedings MICCAI*, 2004.

[6] G. Davis, M. Maggioni, and R. Coifman et al, "Spectral/spatial analysis of colon carcinoma," *Journal of Modern Pathology*, vol. 16, pp. 320–332, 2003.

[7] K. Masood and N. Rajpoot, "Classification of colon biopsy samples by spatial analysis of a single spectral band from its hyperspectral cube," *Proceedings Medical Image Understanding and Analysis (MIUA)*, 2007.

[8] T. Maenpaa, T. Ojala, M. Pietikainen, and M. Sariano, "Robust texture classification by subsets of local binary patterns," *Proc. 15th International Conference on Pattern Recognition*, pp. 947–950, 2000.

[9] J. Kim S. Lee, K. Kim, J. Park, S. Park, and W. Moon, "Optimal custering of kinetic patterns on malignant breast lesions: Comparison between k-means clustering and three-time-points method in dynamic contrast-enhanced MRI," *Proc. of the 29th Annual International Conference of the IEEE EMBS*, 2007.

[10] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, pp. 193–218, 1985.

[11] S. Aranganayagi and K. Thangavel, "Clustering categorical data using silhouette coefficient as a relocating measure," *Conference on Computational Intelligence and Multimedia Applications IEEE: 13-17*, 2007.

[12] R. Huang, Q. Liu, and S. Ma, "Solving the small sample size problem of LDA," *Proc. of international conference on pattern recognition*, 2002.