Hyperspectral Colon Biopsy Classification into Normal and Malignant Categories

Khalid Masood, Nasir Rajpoot
Department of Computer Science, The University of Warwick

September 28, 2006

Abstract

Diagnosis and cure of colon cancer can be improved by efficiently classifying the colon tissue cells from biopsy slides into normal and malignant classes. This report presents the classification of hyperspectral colon tissue cells using morphology of gland nuclei of cells. The application of hyperspectral imaging techniques in medical image analysis is a new domain for researchers. The main advantage of using hyperspectral imaging is the increased spectral resolution and detailed subpixel information. The proposed classification algorithm is based on the subspace projection techniques. Support vector machine, with 3rd degree polynomial kernel, is employed in final set of experiments. Dimensionality reduction and tissue segmentation is achieved by Independent Component Analysis (ICA) and k-means clustering. Morphological features, which describe the shape, orientation and other geometrical attributes, are extracted in one set of experiments. Grey level co-occurrence matrices are also computed for the second set of experiments. For classification, linear discriminant analysis (LDA) with co-occurrence features gives comparable classification accuracy to SVM using a 3rd degree polynomial kernel. The algorithm is tested on a limited set of samples containing only ten biopsy slides and its applicability is demonstrated with the help of measures such as classification accuracy rate and the area under the convex hull of ROC curves.

Contents

List of Figures

1 Introduction

1.1 Colon Cancer

Colon cancer is a malignant disease of the large bowel. After lung and breast cancer, colorectal cancer (a combined term for colon and rectal cancer) is the most common cause of death for cancers in the Western world. The incidence of disease in England and Wales is about 30,000 cases/year, resulting in approximately 17,000 death/annum [19], and it has been estimated that at least half a million cases of colorectal cancer occur each year worldwide. It is caused by colonic polyps, an abnormal growth of tissue that projects in due course from the lining of the intestine or rectum, into colorectal cancer. These polyps are often benign and usually produce no symptoms. They may, however, cause painless rectal bleeding usually not apparent to the naked eye. The normal time for a polyp to reach 1 cm in diameter is five years or a little more. This 1 cm polyp will take around 5-10 years for the cancer to cause symptoms by which time it is frequently too late [17].

Diets low in fruits, vegetables, less protein from vegetable sources, high age and family history are associated with increased risk of polyps. Persons smoking more than 20 cigarettes a day are 250 percent more likely to have polyps as opposed to nonsmokers who otherwise have the same risks. There is an association of cancer risk with meat, fat or protein consumption which appear to break down in the gut into cancer causing compounds called carcinogens [13]. Smoking cessation is important to decrease the likelihood of developing colon cancer. Dietary supplementation with 1500 mg of calcium or more a day is associated with a lower incidence of colon cancer. Weight reduction may be helpful in reducing the risk for colorectal cancer. Daily exercise reduces the likelihood of developing colon cancer. Turmeric, the spice which gives curry its distinctive yellow color, may also prevent colon cancer [10].

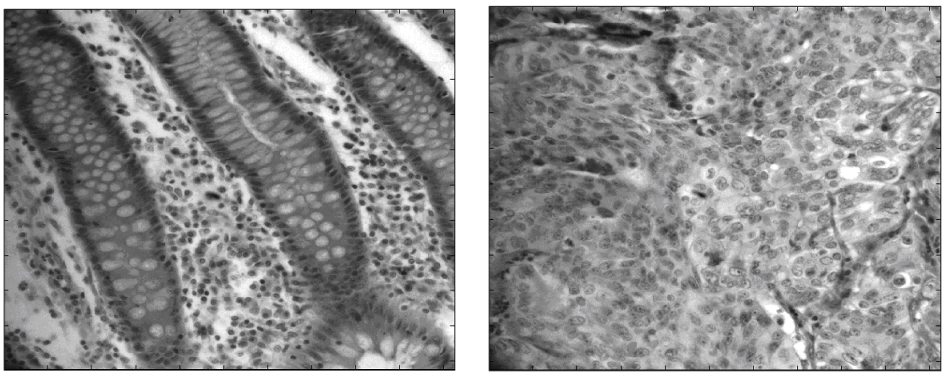
1.2 Hyperspectral Imaging

Hyperspectral imaging in laboratory experiments, is a non-contact sensing technique for obtaining both spectral and spatial information about a tissue sample. Hyperspectral imaging measures a spectrum for each pixel in an image. There are many types of spectroscopy which are being used to study the spectral signatures of individual cells and underlying tissue sections. In optical spectroscopy, which measures transmission through, or reflectance from, a sample by visible or near-infrared radiation at the same wavelength as the source, classification is done mostly by statistical measures [1].

Hyperspectral images are normally produced by emission of spectra from imaging spectrometers. Spectroscopy is the study of light that is emitted by or reflected from materials and its variation in energy with wavelength [15]. Spectrometers are used to make measurements of the light reflected from a test specimen. A prism in the centre of spectrometer splits this light into many different wavelength bands

and the energy in each band is measured by detectors which are different for each band. By using large number of detectors (even a few thousand), spectrometers can make spectral measurements of bands as narrow as 0.01 micrometers over a wide wavelength range, typically at least 0.4 to 2.4 micrometers (visible through middle infrared wavelength ranges). Most approaches to analyse hyperspectral images concentrate on the spectral information in individual image cells, rather than spatial variations within individual bands or groups of bands. The statistical classification (clustering) methods often used with multispectral images can also be applied to hyperspectral images but may need to be adapted to handle high dimensionality.

Recent developments in hyperspectral imaging have enhanced the usefulness of the light microscope [5]. A standard epifluorescence microscope can be optically coupled to an imaging spectrograph, with output recorded by a CCD camera. Individual images are captured representing Y-wavelength planes, with the stage successively moved in the X direction, allowing an image cube to be constructed from the compilation of generated scan files. Hyperspectral imaging microscopy permits the capture and identification of different spectral signatures present in an optical field during a single-pass evaluation, including molecules with overlapping but distinct emission spectra. High resolution characteristics of hyperspectal imaging is reflected in two sample images in Figure 1 of colon tissue cells.



(a) Normal cells (b) Malignant cells

Figure 1: Colon Tissue Imagery

1.3 Previous Approaches

It has been suggested that fatality due to colon cancer can be reduced if cancer is detected in its early stages. Regular screening is now being provided more often in the hospitals and colorectal cancers are resected in earlier stages. Due to increase in the workload associated with increased screening procedure, there is more demand for automated, quantitative analysis techniques capable of classifying the specimens efficiently. A number of studies have looked at various solutions for classification of biopsy images slides. However, despite showing promising re-

sults, the required accuracy for automated systems is still not achieved. Further research to make robust automated system is under way and in the next section some selected approaches for the solution of problems are presented. But, as mentioned above, most of these approaches need large refinements to make them practically applicable.

A computational model of light interaction with colon tissue and classification using the tissue reflectance spectra is analyzed in [22]. Three layers: mucosa, submucosa and smooth muscle, have different reflectance properties with the light incident on their surface. Different parameters characterising the mucosa layer include blood volume fraction, haemoglobin saturation, the size of the collagen fibres, their density and the layer thickness, have unique changes for the emitted spectra of normal and malignant colon tissues. Multispectral values can be obtained from the computed spectra by convolving each spectrum by a set of appropriate band pass filters. Twenty wavelengths were chosen to best describe the colon spectra. Fifty normal and seven cancerous tissues were used as an input to the system. There is an increased blood content, decreased collagen density and increase in the mucosal thickness for the abnormal tissue in comparison to the normal tissue of the colon. Thus it is claimed that above model correctly predicts the spectra of colon tissue and this is an agreement with known histological changes which occur with the development of cancer which alters the macroarchitecture of the colon tissue.

The approach proposed in [18] uses statistical scalar parameters for classification. Simple first order parameters and stochastic-geometric functions including volume and surface area per unit reference volume, the pair correlation function and the centred quadratic contact density function of epithelium were estimated to classify benign and malignant lesions of glandular tissues. Approaches of Linear discriminant analysis- the classical statistical reference method- and artificial neural networks, i.e. computational systems constructed in analogy to neurobiological models, are used for classification. Regardless of the input variables and of the classification procedure, the accuracy of classification is greater than 90 percent. The results show that malignant transformation was always accompanied by an increase of the surface area of epithelium per unit volume and a decrease of the absolute area between the pair correlation function and its reference lines. In malignant tumors, the proliferative activity of the epithell cells is anymore regulated by the normal metabolic or hormonal pathways, hence these cells can autonomously overgrow the tissue. The high classification accuracy is from measurements at low magnification, without taking any nuclear morphology or cytological details into account and cellular details at high magnification are not needed. Thus different statistical parameters correctly describe the change in tissue structure with the development of cancer.

Mass spectrometry, classification of samples using difference in their molecular weights, is used in [16] for the probabilistic classification of healthy vs. disease whole serum samples. The approach employs Principal Components Analysis (PCA) followed by Linear Discriminant Analysis (LDA) on mass spectrometry

data of complex protein mixtures. Analysis of mass spectra by manual inspection has been feasible for samples containing a small number of molecules. Classification is done on the principle that healthy spectrum lies closer to the healthy cluster than to the disease cluster (and vice-versa). Test spectra can than be classified by minimum distance to its nearest cluster. Linear discriminant analysis creates a hyperplane that maximises the between-class variance while minimising the within-class variance. Each spectrum is represented by a point in spectral-space, the set of all spectrum points in spectral-space reduced in dimensionality using Principal Components Analysis (PCA). The use of a probabilistic classification framework increases the predictive accuracy of the algorithm. The high accuracy shows that algorithm is able to determine the spectral differences existing between healthy and disease samples which are a function of molecular differences. Using the relative spectral differences, classification of the sample tissue is achieved with reasonable accuracy.

In [7], classification of tumors is done using expression levels of gene patterns in the tissue samples. The underlying principle in gene expression level is that sample from the healthy class share some common expression profile patterns unique to their class and same is true for the samples from the diseased class. In order to extract information from gene expression measurements, different methods have been employed to analyse this data including Support Vector Machines (SVMs), clustering methods, self organizing maps, and a weighted correlation method. Using oligonucleotide arrays, expression levels for 40 tumors and 20 normal colon tissues are recorded for 6500 human genes. Only 2000 genes with the highest minimal intensity across the tissues are selected for classification. Using a SVM with simple kernel and performing a leave one out method, error rates as low as 9 percent were noted when six tissues were mis-classified. Experiments are repeated with top 1000 genes and same result is obtained. SVMs, a supervised machine learning technique, have been shown to perform well in multiple areas of biological analysis including evaluating microarray expression data, detecting remote protein homologies, and recognising translation initiation sites. Experiments were repeated using p-norm perceptron algorithms and results obtained are comparable to those for the SVMs. The dataset contains a few samples and thus superiority of any method could not be proved. The SVM performs well using only a simple kernel and use of more complex kernel can achieve further good results.

In a similar approach to above, gene expression data is used to classify multiple tumour types [24]. The data consists of 190 human tumor samples containing 14 cancer types. The training set contained 144 samples and the test set contained 46 samples. For colorectal cancer, 8 samples are used for training and 5 samples for testing. Three different classifiers i.e. weighted voting, KNN and Support vector machines (SVMs) have been used. Among the thousands of genes, usually only a small portion of them are related to the distinction of tumour classes. Feature selection involves the choice of a subset of genes based on the strength of their correlation to separate tumor classes. Experiments were performed selecting the top 200 genes as features from 16063 genes, in terms of the absolute values of the

signal to noise ratio. The assumption is that tumor gene expression profiles might serve as molecular fingerprints that would allow for the accurate classification of tumors.

In [23], it is proposed that by modelling the orientational selectivity of neurons in stained histological specimens of colon tissue, classification can be done between images of normal, dysplastic (transitional) and cancerous samples. This classification is based on the assumption that normal colon tissue is characterised by a well-organised, strongly oriented structure, which breaks down due to disease, losing orientational coherence in dysplastic conditions and almost irregular and deformed structure in a cancerous state. Shape analysis is done using two metrics based on a bank of asymmetrical and symmetrical sample cells each covering 6 equally spaced orientations. Two parameters, total activation ratio (T) and the orthogonal activation ratio (T_1), based on the orientational preference response of stretched-Gabor filters to contrast edges and line-like features of a specific orientation, are calculated. These metrics yield high values for images of normal colon, while for dysplasia or cancer, the metrics values are comparatively low. Experiments are performed on a very low dataset and only 15 images, 5 for each case, have been used. There is also an overlap between measures for the dysplastic and cancerous samples. It is concluded that this approach can be used in parallel with texture based approaches using features such as the angular second moment, contrast function, correlation function, entropy and inverse difference moment to discriminate between normal and malignant specimens.

The approach in [20], which is the basis of our work, focuses on the hyperspectral analysis of colon tissue and cell classification is done using Support Vector Machines (SVMs). A Gaussian kernel with low bandwidth is used for classification. Morphology of cell shape is used as a basis for the discrimination of two classes. It is believed that there is a significant difference between the cell shape (structure) of normal and malignant tissues. The cell shape for normal cells is well oriented and have definite boundaries between cell constituents while cell shape of the malignant tissue is deformed and deshaped. Thus it is proposed that cell shape can be represented by a feature vector containing information about shape, size, orientation and geometry. This feature vector can produce significant discrimination between the classes and use of a simple linear classifier can do classification with reasonable accuracy. A feature vector is extracted which contains different morphological features obtained from four binarised images of the original biopsy slides. High classification accuracy is achieved by selecting an optimal set of parameters for the Gaussian kernel. The only limitation is that it is computationally intensive and the parameter selection procedure is exhaustive.

2 Dimensionality Reduction and Subspace Projection

There is a large redundant information in the subbands of hyperspectral imagery. Independent component analysis (ICA) is used to discard the redundancy and ex-

tract the variance among different wavelengths of spectra. K-means clustering is used to help the dimensionality reduction procedure and to segregate the biopsy slide into its cellular components. Subspace projection is achieved with principal component analysis (PCA) and linear discriminant analysis (LDA). A brief introduction to the mathematical derivation of all these four methods is presented in the following subsections.

2.1 Independent Component Analysis (ICA)

The objective of Independent Component Analysis (ICA) is to perform a dimension reduction approach to achieve decorrelation between independent components [25]. Let us denote by $X = (x_1, x_2, \ldots, x_m)^T$ a zero-mean m-dimensional variable, and $S = (s_1, s_2, \ldots, s_n)^T$, $n < m$, is its linear transform with a constant matrix W [17]:

$$S = WX$$

Given X as observations, ICA aims to estimating W and S. The goal of ICA is to find a new variable S such that transformed components s_i are not only uncorrelated with each other, but also statistically as independent of each other as possible. An ICA algorithm consists of two parts, an objective function which measures the independence between components, entropy of each independent source or their higher order cumulants, and the second part is the optimisation method used to optimise the objective function. Higher order cumulants like kurtosis, and approximations of negentropy provide one-unit objective function. A decorrelation method is needed to prevent the objective function from converging to the same optimum for different independent components. Whitening or data sphering project the data onto its subspace as well as normalizing its variance.

2.2 K-Means Clustering

Clustering is the process of partitioning or grouping a given set of patterns into disjoint clusters. This is done such that patterns in the same cluster are alike and patterns belonging to two different clusters are different. The k-means method has been shown to be effective in producing good clustering results for many practical applications [2]. The aim of the k-means algorithm is to divide m points in n dimensions into k clusters so that the within-cluster sum of squared distance from the cluster centroids is minimised. The algorithm requires as input a matrix of m points in n dimensions and a matrix of k initial cluster centres in n dimensions. The number of clusters k is assumed to be fixed in k-means clustering. Let the k prototypes $(w_1, \ldots, w_k)$ be initialised to one of the m input patterns $(i_1, \ldots, i_m)$. Therefore;

$$w_j = i_l, j \in \{1, \ldots, k\}, l \in \{1, \ldots, m\}$$

The appropriate choice of k is problem and domain dependent and generally a user must try several values of k. The quality of the clustering is determined by the

following error function:

$$E = \sum_{j=1}^{k} \sum_{i_l \in C_j} |i_l - w_j|^2$$

The direct implementation of k-means method is computationally very intensive.

2.3 Principal Component Analysis

PCA is a classical statistical method. This linear transform has been widely used in data analysis and compression [26]. The aim of PCA is to reduce the dimensionality from p variables to something much less while preserving the variance-covariance structure. More explicitly, we try to explain the variance-covariance structure through a few linear combinations of the original variables. Given N points in d dimensions PCA essentially projects the data points onto p, directions ($p < d$) which capture the maximum variance of the data. These directions correspond to the eigen vectors of the covariance matrix of the training data points. Intuitively PCA fits an ellipsoid in d dimensions and uses the projections of the data points on the first p major axis of the ellipsoid. A brief derivation of PCA is as follows.

Suppose there are N data points in d dimensions with the mean subtracted. The objective is to find c directions which capture the maximum variance of the data points. The data points can than be projected on these c directions also known as the basis. The projections are called as principal components. Let a_{nm} be the projection of the n^{th} data points on the m^{th} basis e_m. Using these projections and the basis functions a c^{th} order approximation of x_n is reconstructed as

$$x_n = \sum_{m=1}^{c} a_{nm} e_m$$

Defining a new $d \times c$ matrix E where each column is a basis function and a vector a_m where the i_{th} element is the projection on the i_{th} basis.

$$E = [e_1 | e_2 | \dots e_c]$$
$${a_n}^T = [e_{n1} | e_{n2} \dots e_{nc}]$$

Then x_n can be written as

$$x_n = E a_n$$

Now the problem can be stated in terms of finding E and a_n which minimise the reconstruction error over all the data points subject to the constraint that $E^t E = I$ i.e. essentially these are orthonormal directions. This problem can be formulated as follows:

$$min_{E,a_n} \frac{1}{2} \sum_{n=1}^{N} |x_n - E a_n|^2$$

PCA is the best linear representation and first c principal components have more variance than any other components. A drawback of PCA is that it considers only the global information of each training image and sometimes unwanted global information, which is not required for classification, is reflected in the top eigenvectors affecting the performance of the classifier.

2.4 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a supervised learning method which finds the linear combinations of features to discriminate between two or more classes of patterns. The LDA method is proposed to take into consideration global as well as local variation of training images. It uses both principal component analysis and linear discriminant analysis to produce a subspace projection matrix. The advantage is in using within-class information, minimising variation within each class and maximising between class variation, thus obtaining more discriminant information from the data [21].

A significant limitation of LDA occurs when the number of training samples is less than the dimensions of feature vectors of the sample. In this case, covariance matrix do not have full rank and within class scatter is not mininmised. This is called small sample size problem. Different solutions have been proposed to deal with it and most widely used method applies PCA firstly to obtain a full rank within-class scatter matrix. But the drawback is that it removes the null space to make the resulting within-class matrix full rank. There are other solutions which extract informations from this discriminant null space [11]. PCA, usually, performs better in small sample size problems.

In many practical problems, linear boundaries cannot separate the classes for discrimination. Support vector machines, using kernel techniques, extract non linear features by mapping the original observations into high dimensional non linear space with the help of kernel trick. In case of large training samples and linear separation boundary, LDA performs better than any other classifier.

LDA is based on the principle of controlled scatter of the training classes. Three scatter matrices, representing within-class (S_w), between-class (S_b) and total scatter (S_t), are computed for the input data set.

$$S_t = \sum_{n=1}^{M} (\Gamma_n - \Psi)(\Gamma_n - \Psi)^T$$

$$S_b = \sum_{i=1}^{C} N_i (\Psi_i - \Psi)(\Psi_i - \Psi)^T$$

$$S_w = \sum_{i=1}^{C} \sum_{\Gamma_k \in X_i} (\Gamma_k - \Psi_i)(\Gamma_k - \Psi_i)^T$$

where $\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n$ is the average feature vector of the training set and $\Psi_i = \frac{1}{|X_i|}\sum_{\Gamma_i \in X_i}\Gamma_i$ is the average feature vector of the ith individual class.

If S_w is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximises the ratio of the determinant of the within-class scatter matrix of the projected samples, i.e.

$$W_{opt} = argmax_w |\frac{W^T S_B W}{W^T S_w W}|$$

$$= [w_1, w_2, \cdots, w_m]$$

where $[w_i \ | i = 1, 2, \ldots \ m]$ is the set of generalised eigen vectors of S_B and S_w corresponding to the m largest generalised eigen values $\{\lambda_i \ | i = 1, 2, \ldots, m\}$, hence Equation becomes $S_B w_i = \lambda_i S_w w_i$ for $i = 1, 2, \cdots, m$

LDA algorithm takes advantage of the fact that, under some idealized conditions, the variation within class lies in a linear subspace of the image space. Hence, the classes are convex, and, therefore linearly separable. LDA is a class specific method, in the sense that it tries to 'shape' the scatter in order to make it more reliable for classification. It is clear that, although PCA achieves larger total scatter, LDA achieves greater between-class scatter, and consequently, yields improved classification [3].

2.5 Limitation of PCA and LDA

PCA is an unsupervised classifier, while LDA is used for supervised classification. Both are used for classification on the assumption of a linear boundary between discriminating classes. While PCA aims to extract a subspace in which the variance is maximised (or the reconstruction error is minimised), some unwanted variation may be retained in its top eigen vectors. Therefore, while the PCA projections are optimal in a correlation sense, these eigenvectors or bases may be suboptimal from the discrimination point of view [9]. LDA is a class-specific method, in the sense that it tries to shape the scatter in order to make it more reliable for classification. LDA is dependent on the size of training samples, when training sample size is low, PCA performs better than LDA. If there are nonlinear or second order dependencies i.e., pixelwise covariance among the pixels, and relationships among three or more pixels in an edge or curve, than linear classifiers like PCA and LDA fail to discriminate between the classes.

3 Methodolgy

The proposed classification algorithm consists of three modules as shown in Figure 2. Brief description of dimensionality reduction and feature extraction modules is given in the following sub-sections. Detailed description of the segmentation can be found in [20].

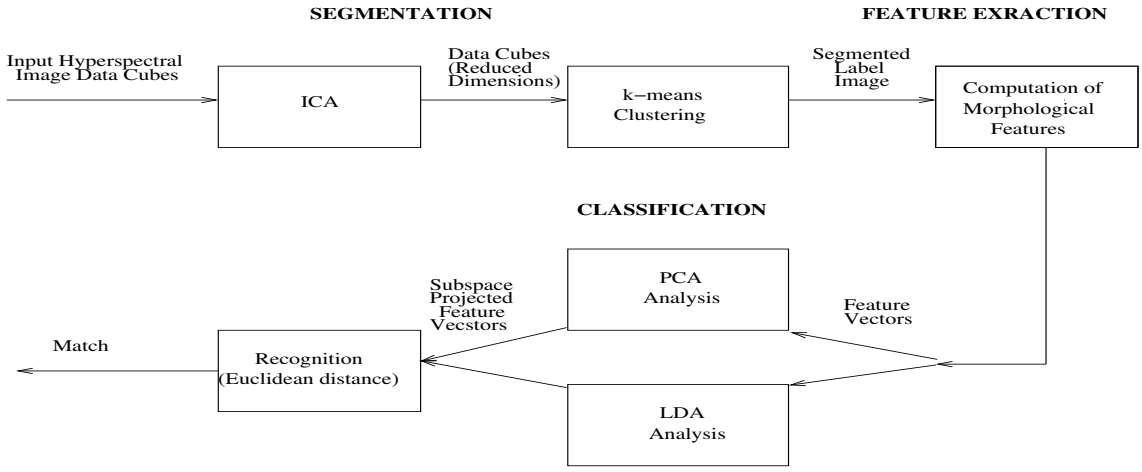


Figure 2: Classification Algorithm Block Diagram

3.1 Segmentation

High dimensional data in the form of cubes is obtained using hyperspectral imaging. For efficient processing this data has to be dimensionally reduced. Dimensionality reduction involves two steps, extraction of statistically independent components using Independent Component Analysis (ICA) and colour segmentation using k-means clustering. Flexible ICA (FlexICA) [12], a fixed point algorithm for ICA, adopting a generalised Gaussian density, is used for data sphering (whitening) and achieves considerable dimensionality reduction. Data is distributed towards heavy-tailedness by the high-emphasis filters. The data with reduced dimensionality is then fed to k-means clustering algorithm for segmentation.

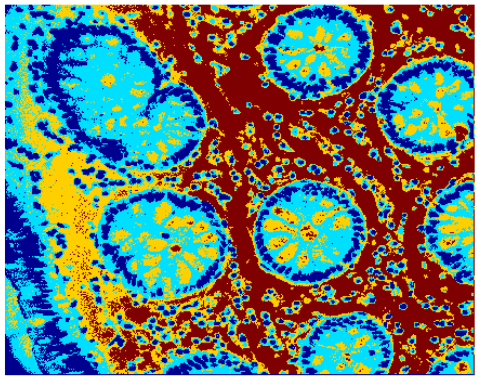
The initial data cube containing 28 subbands is segmented to four labelled parts in each image after the process of dimensionality reduction. Each slide of tissue cells is divided into four regions represented by four colours as shown in Figure 3. The four labelled parts are denoted by the following colours:

- nuclei : dark blue
- cytoplasm : light blue
- gland secretions : yellow
- stroma of the lamina propria : red

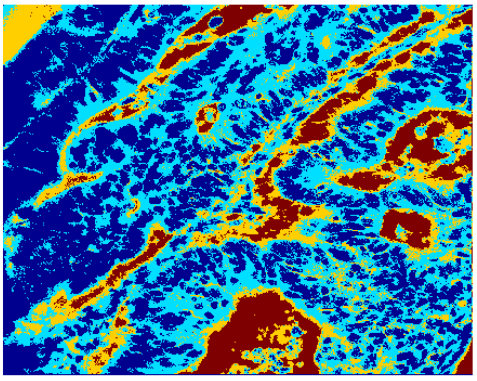
3.2 Feature Extraction

3.2.1 Binarised Images

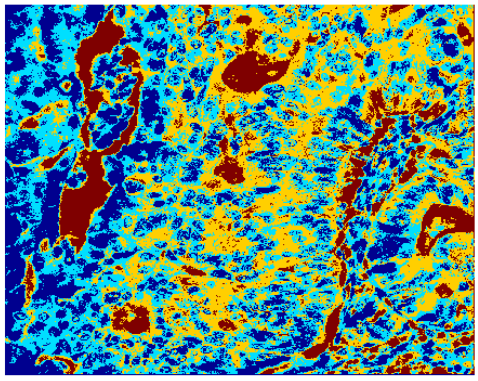
In order for the pattern recognition process to be tractable it is necessary to represent patterns into some mathematical or analytical model. The model should



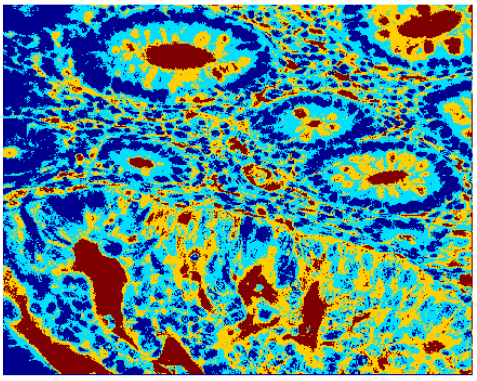
(a) Benign cells



(b) Malignant cells



(c) Malignant cells



(d) Normal/Malignant cells

Figure 3: Segmentation Results

convert patterns into features or measurable values, which are condensed representations of the patterns, containing only salient information [19]. Features contain the characteristics of a pattern to make them comparable to standard templates making the pattern classification possible. The extraction of good features from these pattern models and the selection from them of the ones with the most discriminatory power are the basis for the success of the classification process. In this work morphological texture features, extracted from the segmented images of a hyperspectral data cube for a biopsy slide of colon tissue cells, are used for the classification of the tissue cells.

Morphological features, which describe the shape, size, orientation and other geometrical attributes of the cellular components, are extracted to discriminate between two classes of input data. The segmented image is first split into four binarised image in accordance with the four cellular components. In each binary image, the corresponding cellular components i.e. nuclei, cytoplasm, gland secretions and stroma of lamina propria have binary value equal to 1. In the Figure 4, binarised images for each of benign and malignant tissues are presented and there is a notable difference in the morphology of cellular shapes.

3.2.2 Features Definition

Morphological features depending on the shape, size and orientation of cellular parts are extracted and a combined feature vector is used for classification. Experiments are performed using different combinations of ten morphological features for the representation of the feature vector. A brief definition of each feature is given below:

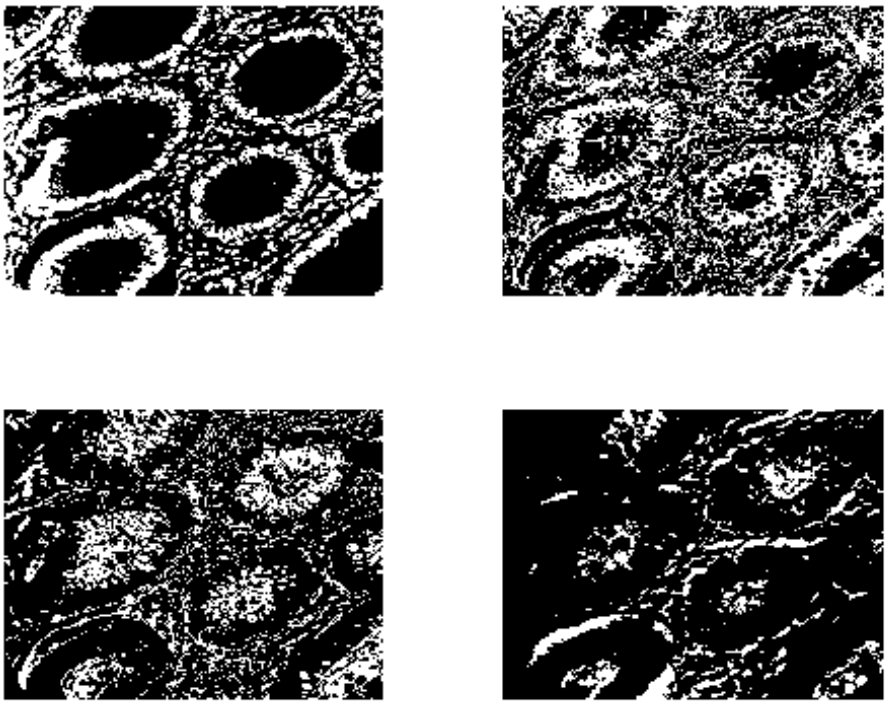
- 'EulerNumber'; equal to the number of objects (a set of object pixels such that any object in the set is in the 8 or 4 neighborhood of at least one object pixel of the same set) in the patch minus the number of holes (a set of background pixels) in those objects. Euler number can be computed from the run length representation of an image [4]. If for each run r, $k(r)$ be the number of runs on the preceding row to which r is adjacent, than Euler number can be expressed as :

  $$E = \sum_{r} (1 - k(r))$$

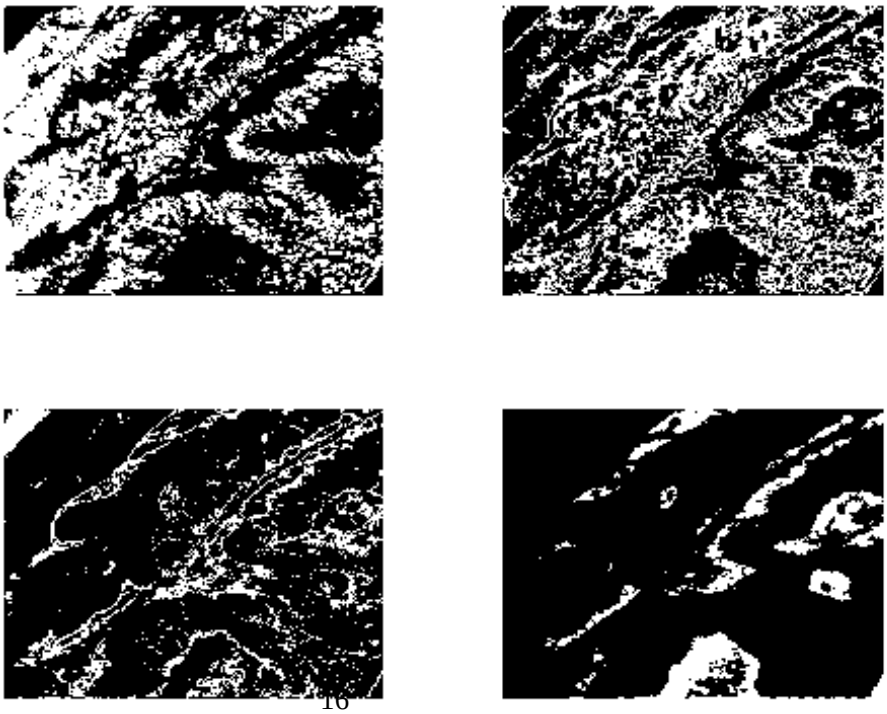
  The number of connected components C in the image is computed by applying a standard graph algorithms. The number of holes H is then computed from the Euler's formula of a planar graph as $H = 1 + m - v$, where m is number of edges and v is number of nodes in the graph. Hence Euler number is calculated as : $E = C - H$ where C = No of contiguous parts and H = No. of holes

- 'Area'; the actual number of on pixels (equal to 1) in the patch. For an $(N \times N)$ binary pixel matrix with off pixels (equal to 0) B, Area is

  $$Area = (N - B)^2$$



(a) Normal Binary Cells



(b) Malignant Binary Cells

Figure 4: Normal and Malignant Binary Images

- 'MajorAxisLength'; the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the patch.

$$Ellipse : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

with $a > b$. Now centroid of an ellipse is defined as ;

$$xx = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$yy = \frac{\sum_{i=1}^{N} y_i}{N}$$

$$x_i = x_i - xx$$

$$y_i = -(y_i - yy)$$

$$u_{xx} = \frac{\sum x_i^2}{N} + \frac{1}{12}$$

$$u_{yy} = \frac{\sum y_i^2}{N} + \frac{1}{12}$$

$$u_{xy} = \frac{\sum x_i y_i}{N}$$

$$c = \sqrt{(u_{xx} - u_{yy})^2 + 4u_{xy}^2}$$

$$a = 2\sqrt{2}\sqrt{u_{xx} + u_{yy} + c}$$

where a is major axis length.

- 'MinorAxisLength'; the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the patch.

$$b = 2\sqrt{2}\sqrt{u_{xx} + u_{yy} - c}$$

where b is minor axis length.

- 'Eccentricity' ; the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length.

$$e = 2\frac{\sqrt{(\frac{a^2}{4} - \frac{b^2}{4})}}{a}$$

$$Eccentricity = \frac{\sqrt{(a^2 - b^2)}}{a}$$

- 'Orientation'; the angle (in degrees) between the minor-axis and the major axis of the ellipse that has the same second-moments as the patch.

$$\tan\theta = \frac{u_{yy} - u_{xx} + \sqrt{(u_{yy} - u_{yy})^2 + 4u_{xy}^2}}{2u_{xy}}$$

$$r^2 = \frac{a^2 b^2}{a^2 \sin^2\theta + b^2 \cos^2\theta}$$

- 'EquivDiameter'; the diameter of a circle with the same area as the patch.

$$EquivDiameter = \sqrt{\frac{4 \times Area}{\pi}}$$

- 'Extent'; the proportion of the pixels in the bounding box that are also in the patch.

$$Extent = \frac{Area}{Area\,of\,the\,bounding\,box}$$

- 'ConvexArea'; the number of pixels in Convexhull (smallest convex polygon that can contain the patch).

- 'Solidity'; the proportion of the pixels in the convex hull that are also in the patch.

$$Solidity = \frac{Area}{ConvexArea}$$

3.2.3 Multiscale Features

Each binarised image is than divided into equal sized patches, and each patch has a size equal to 16×16. For a maximum of ten morphological feature values, each patch has a feature vector of 40 scalar values, for single scaling, to represent the shape of the patch. The input image has a size 1024×1024, so with a patch size of 16×16, total number of patches in a slide equals to 4096. Training data contains 25 percent of these patches and remaining 75 percent patches are used for testing of data. Image is divided into patches because some of the patches may represent normal cells and other patches may come from malignant cells, even from the same slide. This feature vector is used as an input for training and testing of the two classifiers which are used for the discrimination of the classes into benign and normal tissue cells.

Multiscale feature extraction is performed to obtain global information of the patches. Features are calculated for each patch at some scale 16×16, patch size is doubled and feature values are re-calculated. This process can be repeated upto 5 times. Fusion of feature values is done by simple concatenation of the features. A concatenated single feature vector for each patch containing values for multi-size patches represents the morphology of glandular nuclei. For ten morphological features, feature vector of 200 values is used as an input for the classifiers.

3.3 Co-occurnence Features

The co-coourrence approach is based on the grey level spatial dependence. Co-occurrence matrix is computed by second-order joint conditional probability density function $f(i, j|d, \theta)$. Each $f(i, j|d, \theta)$ is computed by counting all pairs of pixels separated by distance d having grey levels i and j, in the given direction θ. The angular displacement θ usually takes on the range of values from $\theta = 0, 45, 90, 135$ degrees. The co-occurrence matrix captures a significant amount of textural information. The diagonal values for a coarse texture are high while for a fine texture these diagonal values are scattered. To obtain rotataion invariant features the co-occurrence matrices obtained from the different directions are accumulated. The three set of attributes used in our experiments are Energy, Inertia and Local Homogeneity.

$$E = \sum_i \sum_j [f(i, j|d, \theta)]^2$$

$$I = \sum_i \sum_j [(i - j)^2 f(i, j|d, \theta)]$$

$$LH = \sum_i \sum_j \frac{(i, j|d, \theta)}{1 + (i + j)^2}$$

4 Experiments

4.1 Experiments with combined training/test data

The experimental setup consists of a unique tuned light source based on a digital mirror device (DMD), a Nikon Biophot microscope and a CCD camera. Two different biopsy slides containing several microdots, where each microdot is from a distinct patient, is prepared. Then each slide is illuminated with a tuned light source (capable of emitting any combination of light frequencies in the range of 450-850 nm), followed by magnification to 400 X. Thus several images, each image using a different combination of light frequencies, are produced [6].

The first set of experiments is carried out with mixed training/test data. Some patches from the same slide are used for training and remaining patches from that slide are used for testing. Hyperspectral image data cubes, equivalent to the number of input biopsy slides, are produced using hyperspectral analysis for twenty eight subbands of light. The dimension of each cube is $1024 \times 1024 \times 28$. Performing FlexICA and k-means clustering for segmentation, we get images having 1024×1024 dimensions per slide. For single scale experiments, each image is divided into 4096 patches of 16×16 dimensions per patch. Morphological operation is performed on the patches for extraction of feature vectors using different combinations of ten scalar morphological properties. Two experiments are carried out, one for PCA and other for LDA. The data (patches of all slides randomly mixed) is divided into training set (about one quarter of the patches) and test set

Morphological Features				
	Patch	**Features**	**Method**	**Accuracy (%)**
Single	16x16	1 (ENo)	PCA	51.0
Single	16x16	2 (ENo,CA)	PCA	51.4
Multi	16x16,32x32,..	5 (ENo,CA,A,E,O)	PCA	75.0
Single	16x16	2 (ENo,CA)	LDA	55.5
Single	16x16	3 (ENo,CA,A)	LDA	55.7
Single	16x16	5 (ENo,CA,A,E,O)	LDA	56.4
Single	16x16	10 (5+5 Elliptical)	LDA	56.4
Single	32x32	2 (ENo,CA)	LDA	60.1
Single	32x32	5 (ENo,CA,A,E,O)	LDA	61.3
Single	32x32	10 (5+5 ELliptical)	LDA	61.3
Multi	16x16,32x32,..	5 (ENo,CA,A,E,O)	LDA	84.0

Table 1: Table for Morphological Attributes

(remaining three quarters of patches). We have applied two fundamental classifiers which are easy to implement and fast in computation but the outcome is sacrifice of the performance of the algorithm. However, it provides significant contribution to build the basis of problem solution and future work will focus on improvement of classifiers which will help to achieve better discrimination results. In the second experiment, multiscale feature extraction is performed. Feature values are initially calculated for base patch size 16×16, patch size is then doubled and feature values are re-calculated. This process continues for at least upto five scales. Fusion of the features, for different scales, is done by simple concatenation. Thus largest feature vector for five scales and using ten morphological parameters has dimensions of 200 fetures values.

Fifteen eigenvectors, from largest eigen values to smaller, are selected for PCA and it gives overall classification rate of 75 percent. LDA is performed in a modular form [8] in which each slide of image represents a separate class. Thus two classes of benign and malignant cells are divided into eleven subclasses and taking top ten vectors, we get on the average 84 percent classification rate. The results are encouraging in terms of the computational speed and ease of implementation. The accuracy in the table refers to the percentage of correctly classified patches.

The ROC curves alongwith the AUCH (Area under convex hull) are shown in Figure 5. ROC curve shows the performance of a particular classifier in terms of true positive and false positive rates. True positive rate represents the number of correct positive cases divided by the total number of positive cases, whereas, false positive rate is the number of negative cases predicted as positive cases, divided by the total number of negative cases. A graph which represents TPR and FPR and which is plotted for several threshold values, is called Receiver operating characteristic (ROC) curve. ROC curve is a performance measure to summarise

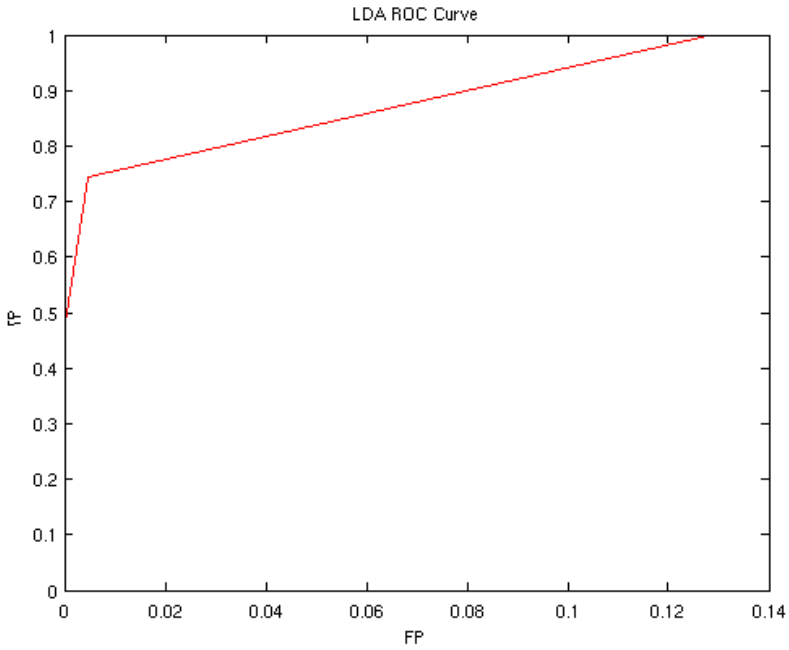
the flexibility of classifier under different operating conditions [14]. ROC curve is obtained by tuning the PCA and LDA for different thresholds. ROC curve gives the optimum threshold which is defined in terms of Euclidean distance of each test patch with all training patches. At the start point, there are some True Positives but no False Positives. As threshold increases, there is increase in positive alarms but simultaneously false positives begin to rise. When threshold becomes maximum than every case is claimed positive and it represents top right hand corner. Area under convex hull (AUCH) shows the performance of classifier with the increase in the number of discriminant features. As the area under the curve increases, it will approximate to ideal AUCH curve (top left hand corner), resulting with better classification rate. The AUCH for LDA is larger than AUCH for PCA which shows better performance of LDA as compared to PCA.

4.2 Experiments with Leave one out data

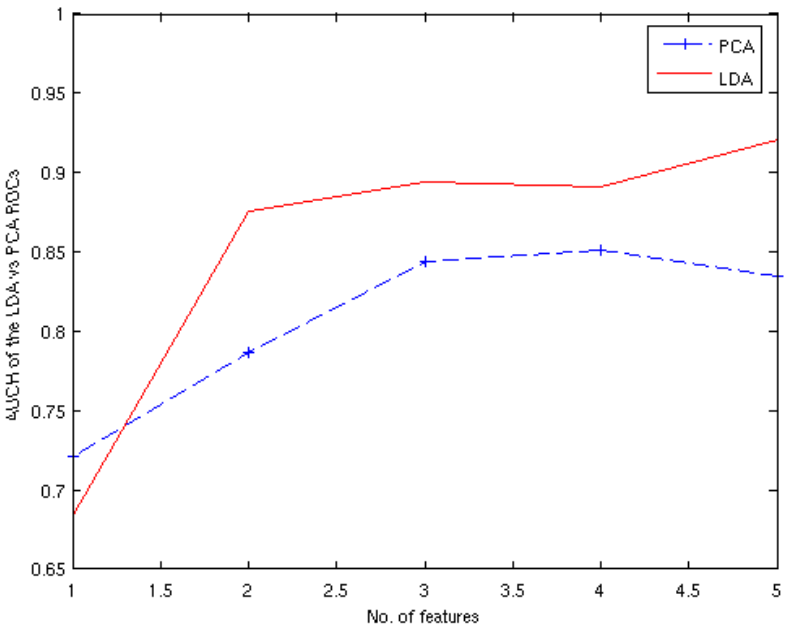
The second set of experiments are used with LOO (leave one out) settings and employing two different subsets. In the first setting, patches are directly fed to classifying machine without feature extraction. Each patch has window size of 64x64. Thus total feature vector dimensionality is 4096. Using top 100 eigen vectors and doing leave one out slide experimentation, the classification rate is 8 slides correctly classified out of 10 slides with a threshold of 55 percent correct patches of the slide.

In the second setting, co-occurrence matrix is computed from the block size of 64x64 for each slide. Three co-occurrence features i.e. Angular second moment (Energy), variance and homogenity are calculated while pixel distance is varied from one pixel to two pixel values. Four directional features in the direction of 0, 45, 90 and 135 are concanated together so that feature vector with 24 dimensions is used in the classifiers. The last experiment is carried out with SVMs. Polynomial kernel of 3rd degree with parameters C=1 (cost of constrain violation), epsilon=0.001 (tolerence of termination criterion), and coefficient=0 is used in this exercise. Experiments with SVM on new data is recommended in our future work, so we have used only default values of SVM without proper tuning to get some insight into SVMs. The results can be improved with optimum selection of kernel and its parameters.

The classification accuracy in GLCM-LOO is about 90 percent and 9 slides on the whole are classified correctly with a threshold of 55 percent on the patches. Directly fed patches have a little less performance as compared to the co-occurrence features from the patches. As we have only limited number of input slides, so comparison is difficult on these set of experiments. Using new data with these setting will give better comparison for the classification accuracy.



(a) LDA ROC Curve



(b) AUCH of LDA vs PCA

Figure 5: ROC & AUCH Performance Curves

Classification Accuracy (%)					
	Eigen Labels	**Fisher Labels**	**Co-PCA**	**Co-LDA**	**Co-SVM**
B01	60.94	68.55	65.71	66.20	37.14
B03	60.55	65.14	61.43	60.71	71.43
B05	73.05	72.07	62.86	70.00	71.43
B11	62.89	67.19	37.14	38.57	57.14
B13	56.64	69.43	61.43	64.29	64.29
M03	63.67	70.02	51.43	55.71	45.71
M05	53.12	54.55	71.40	62.39	68.57
M07	57.81	67.58	68.57	57.86	72.86
M09	47.27	49.72	68.35	70.29	55.71
M11	51.17	61.43	61.43	60.10	72.18
Overall	7/10	8/10	8/10	9/10	8/10

Table 2: Table for Leave One Out Experiments

4.3 Conclusions

4.3.1 Main Contribution

In this report, classification of colon tissue cells is achieved using the morphology of the glandular cells of the tissue region. There is an indication that the morphology of the cells, obtained from the hyperspectral analysis of biopsy slides, has enough discriminatory power to apply a simple classifier for its efficient classification. Regular structured cell shapes with some orientations are characteristics of normal cells, whereas irregular and deformed cell shapes represent malignant tissue. The extraction of features characterising efficiently the structure of tissue cells is important for the correct cell classification. Ten features describing binary cellular and elliptical properties of textures have been extracted and subsequently used for classification. Segmentation and dimensionality reduction is performed using k-means clustering and independent component analysis. Single scale and multiscale feature extraction is employed. Single scale features depend on the patch size and with increase in patch size classification accuracy increases, but with a reduced spatial resolution for classification. Multiscale feature extraction produces a bias in the testing mechanism, as test patches carry some common global information from the training patches.

4.3.2 Future work

Future work will be carried out in the following directions.

1. Segmentation : Segmentation and dimensionality reduction will be performed more efficiently. Modified form of k-means algorithm, which can improve

the direct k-means algorithm by an order to two orders of magnitude in the total number of distance calculations and the overall time of computation.

2. Feature extraction : Morphological features represent the specified glandular structure at different scales. Extraction of spatial gray level dependence matrices features, with or without combination of statistical features will be the immediate next step. Co-occurrence approach is based on the estimation of the second-order joint conditional probability density function $f(i, j|d, \theta)$. Each $f(i, j|d, \theta)$ computed by counting all pairs of pixels separated by distance d having gray levels i and j, in the given direction θ. The angular displacement θ usually takes on the range of values : $\theta = \{0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4}\}$. The co-occurrence matrix captures a significant amount of textural information. For a coarse texture these matrices tend to have high values near the main diagonal whereas for a fine texture the values are scattered. To obtain rotation-invariant features the co-occurrence feature matrices obtained from the different directions are accumulated. Based on the probability density function, the following texture measures will be computed.

    - Angular second moment, Entropy, Difference variance, Difference entropy, Inverse difference moment
    - Contrast, Correlation, Sum average, Sum variance, Sum entropy, Sum of squares variance
    - Information measures of correlation

3. Shape Analysis : The macroarchitecture of normal of normal gladular cells have a regular tubular shape with normal cells on its boundary. Automatic shape analysis to fit contours on the gland structure will be investigated. The shape of gland strucure in malignant cells become distorted while normal gland structure has elliptical and circular type shape. There are number of methods to analyse the shape and structure of the patterns and shape analysis with the help of snakes may probably help in our case.

4. Classification : So far only linear classifiers have been used and next step is to look for more efficient classifier networks. Support Vector Machines (SVMs) will replace traditional LDA in our next batch of experiments. Kernel PCA, Kernel LDA, and their modified versions will be employed to explore non-linear boundaries between classes. Different types of kernels including polynomial, gaussian, sigmoid, traingular, cosine kernels and their optimum tuning parameters will be explored.

5. New Data : New data with improved hyperspectral techniques and high resolution is available for our experiments. It contains an additional transitional class apart from normal and malignant classes. Hyperspectral analysis has been done with very narrow bandwidths and number of bands is increased upto 128. It would be a challenge to efficiently discriminate all the three

classes and it would need efficient feature extraction and a better classifier to reduce the error rate.

References

[1] John Adams, M. Smith, and A. Gillespie. Imaging spectroscopy: Interpretation based on specral mixture analysis. *Remote Geochemical Analysis*, 1993.

[2] K. Alsabti, S. Ranka, and V. Singh. An efficient k-means clustering algorithm. *www.cise.ufl.edu.*, 1997.

[3] Belhumeur, J. Hesponha, and D. Kriegman. Eigenfaces vs fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1997.

[4] Arijit Bishnu, Bhargab Bhattacharya, and Malay Kundu et el. A pipeline architecture for computing the euler number of a binary image. *International conference on image processing (ICIP)*, 3:310–313, 2001.

[5] E. A. Cloutis. Hyperspectral geological remote sensing. *Evaluation of Analytical Techniques-International Journal of Remote Sensing*, 17:2215–2242, 1996.

[6] G. Davis, M. Maggioni, and R. Coifman et al. Spectral/spatial analysis of colon carcinoma. *Journal of Modern Pathology*, 2003.

[7] Terrence Furey, Nello Cristianini, and Nigel Duffy et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 2000.

[8] R. Gottmukul and V. K. Asari. An improved face recognition technique based on modular pca approach. *Pattern Recognition Letters*, 25:429–436, 2004.

[9] Thomas Heseltine, Nick Pears, and Jim Austin. Evaluation of image preprocessing techniques for eigenface based face recognition. *Second international conference on image and graphics SPIE*, 2002.

[10] R. S. Houlston. Molecular pathology of of colorectal cancer. *Clinical Pathology*, 2001.

[11] R. Huang, Q. Liu, and S. Ma. Solving the small sample size problem of lda. *In proce. of intl. conf. pattern recognition*, 2002.

[12] A. Hyvarinen. Survey on independant component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[13] S. Kaster, S. Buckley, and T. Haseman. Colonoscopy and barium enema in the detection of colorectal cancer. *Gastrointestinal Endoscopy*, 1995.

[14] A. Khan, A. Majid, and A. Mirza. Combination and optimization of classifier in gender classification using genetic programming. *International Journal of Knowledge Based and Intelligent Engineering Systems*, 9:1–11, 2005.

[15] David Landgrebe. Hyperspectral image data analysis as a high dimensional signal processing problem. *IEEE Signal Processing magazine*, 2002.

[16] Ryan Lilien, Hany Farid, and Bruce Donald. Probabilistic disease classification of expression-dependent proteomic data from mass spectrometry of human serum. *Journal of Computational Biology*, 2003.

[17] D. E. Mansell. Colon polyps & colon cancer. *American Cancer Society Textbook of Clinical Oncology*, 1991.

[18] T. Mattfeldt, H. Gottfried, and V. Schmidt. Classification of spatial textures in benign and cancerous glandular tissues by stereology and stochastic geometry using artificial neural networks. *Journal of Microscopy*, 2000.

[19] Office of National Statistics. Cancer statistics: Registrations, england and wales. london. *HMSO*, 1999.

[20] K. M. Rajpoot and Nasir M Rajpoot. Hyperspectral colon tissue cell classification. *SPIE Medical Imaging (MI)*, 2004.

[21] N. Rajpoot and K. Masood. Human gait recognition with 3-d wavelets & kernel based subspace projections. *International Workshop on HAREM*, 2005.

[22] Dzena Rowe, Ela Claridge, and Tariq Ismail. Analysis of multispectral images of the colon to reveal histological changes characteristic of cancer. *MIUA*, 2006.

[23] Alison Todman, Raouf Naguib, and Mark Bennen. Visual characteristics of colon images. *IEEE CCECE*, 2001.

[24] Chen-Hsiang Yeang, Sridhar Ramaswamy, and Pablo Tamayo et al. Molecular classification of multiple tumor types. *Bioinformatics*, 2001.

[25] Kun Zhang and Lai-Wan Chan. Dimension reduction based on orthogonality-a decorrelation method in ica. *ICANN/ICONIP*, 2003.

[26] Y. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. *Proceedings 3rd International Conference on Automatic Face and Gesture recognition*, pages 336–341, 1996.