Histopathological Image Analysis Using Model-Based Intermediate Representations and Color Texture: Follicular Lymphoma Grading

**Olcay Sertel · Jun Kong · Umit V. Catalyurek ·
Gerard Lozanski · Joel H. Saltz · Metin N. Gurcan**

Abstract Follicular lymphoma (FL) is a cancer of lymph system and it is the second most common lymphoid malignancy in the western world. Currently, the risk stratification of FL relies on histological grading method, where pathologists evaluate hematoxilin and eosin (H&E) stained tissue sections under a microscope as recommended by the World Health Organization. This manual method requires intensive labor in nature. Due to the sampling bias, it also suffers from inter- and intra-reader variability and poor reproducibility. We are developing a computer-assisted system to provide quantitative assessment of FL images for more consistent evaluation of FL. In this study, we proposed a statistical framework to classify FL images based on their histological grades. We introduced model-based intermediate representation (MBIR) of cytological components that enables higher level semantic description of tissue characteristics. Moreover, we introduced a novel color-texture analysis approach that combines the MBIR with low level texture features, which capture tissue characteristics at pixel level. Experimental results on real follicular lymphoma images demonstrate that the combined feature space improved the accuracy of the system significantly. The implemented system can identify the most aggressive FL (grade III) with 98.9% sensitivity and 98.7% specificity and the overall classification accuracy of the system is 85.5%.

Keywords Histopathological image analysis ·
Model-based intermediate representation ·
Color texture analysis · Follicular lymphoma

O. Sertel (✉) · J. Kong · U. V. Catalyurek
Department of Electrical and Computer Engineering,
The Ohio State University, 2015 Neil Ave,
Columbus, OH 43210, USA
e-mail: sertelo@ece.osu.edu

J. Kong
e-mail: kongj@ece.osu.edu

O. Sertel · J. Kong · U. V. Catalyurek ·
J. H. Saltz · M. N. Gurcan
Department of Biomedical Informatics,
The Ohio State University, 333 W Tenth Ave,
Columbus, OH 43210, USA

O. Sertel
e-mail: osertel@bmi.osu.edu

J. Kong
e-mail: kongj@bmi.osu.edu

U. V. Catalyurek
e-mail: umit@bmi.osu.edu

J. H. Saltz
e-mail: saltz@bmi.osu.edu

M. N. Gurcan
e-mail: gurcan@bmi.osu.edu

G. Lozanski
Department of Pathology, The Ohio State University,
1645 Neil Ave, Columbus, OH 43210, USA
e-mail: Gerard.Lozanski@osumc.edu

1 Introduction

Follicular lymphoma (FL) is a cancer of lymph system and it is the second most common lymphoid malignancy in the western world. FL is a mature B lymphocyte malignancy of follicular center cell origin. Diagnosis of FL is based on specific morphologic, immunophenotypic and cytogenetic findings in lymph node/tissue biopsy specimens.

FL has a highly variable clinical course ranging from an indolent to a highly aggressive disease. Patients with

indolent disease often live for many decades and may never require therapy, while the patients with aggressive FL have short survival if not treated early with aggressive chemotherapy. It is important to note that in contrast to aggressive FL, the indolent FL patients do not benefit from early chemotherapy and that treatment should be avoided in these patients to prevent serious side effects. This variable clinical presentation requires an accurate risk stratification of FL samples as a guidance for oncologist in making decisions on timing and type of therapy. As a result, it can contribute to reducing the likelihood of making under and over treatments.

Currently, the most commonly used FL risk stratification method is histological grading (HG) system adopted by the World Health Organization [1]. The HG method is based on average count of large malignant cells called centroblasts (CB) per standard microscopic high power field (HPF) defined as 0.159 mm^2. Follicular lymphoma cases are stratified into three histological grades: Grade I (0–5 CB/HPF), grade II (6–15 CB/HPF) and grade III(> 15CB/HPF). Grades I and II are considered low risk category while grade III is considered high risk category. In this method the average centroblast count per HPF is based on CB count in ten random HPFs representing malignant follicles. The CB count is performed manually by the pathologist using an optical microscope and hematoxilin and eosin (H&E) stained tissue section(s). Since this is a highly subjective method, the results show well documented inter- and intra-observer variability [2, 3] for the various grades of FL even among the experts [4]. Moreover, since this method, for practical reasons, uses only ten high power fields for CB count, the results for specimens with high tumor heterogeneity are vulnerable to sampling bias. This poor reliability and reproducibility of FL histological grading may lead to inappropriate clinical decisions on timing and type of therapy and result in under or over treatment for the individual FL patient with many serious clinical consequences. Using computerized image analysis, it is possible to extract more objective and accurate prognostic clues, which may not be easily observed by qualitative analysis performed by pathologists. Besides, instead of evaluating only representative regions, a computerized system can process the whole-slide and prevent the sampling bias.

As reported by Meijer et al., the roots of image analysis for a more objective and reproducible prognosis date back to seventeenth century [5]. Being amazingly precise, Leeuwenhoek had developed a system to measure the size of human erytrocytes using sand grain and hairs from his head. However, the real acceleration in histopathological image analysis is due to the recent developments in whole-slide scanners. Whole slide scanners allow digitization of whole microscope slides at high magnifications up to 40× and provide very high resolution images. Recently, several image analysis approaches have been proposed for diverse types of cancer such as prostate [6], breast [7], brain [8] and neuroblastoma [9, 10]. These methods commonly exploit texture, color or morphological properties of the tissue and propose quantitative methods to differentiate different histological grades.

In this study, our goal is to develop a computer-aided prognosis (CAP) system that will assist the pathologists in the grading of FL. The flowchart of the proposed system is given in Fig. 1.

We propose a novel approach that semantically describes histology images using model based intermediate representation (MBIR) and incorporates low level color texture analysis. In this approach, we first identify basic cytological components in the image and model the connected components of such regions using ellipses. An extensive set of features can be constructed from this intermediate representation to characterize the tissue. Using this representation, we measure the

Figure 1 The flowchart of the proposed computerized image analysis system for FL prognosis.
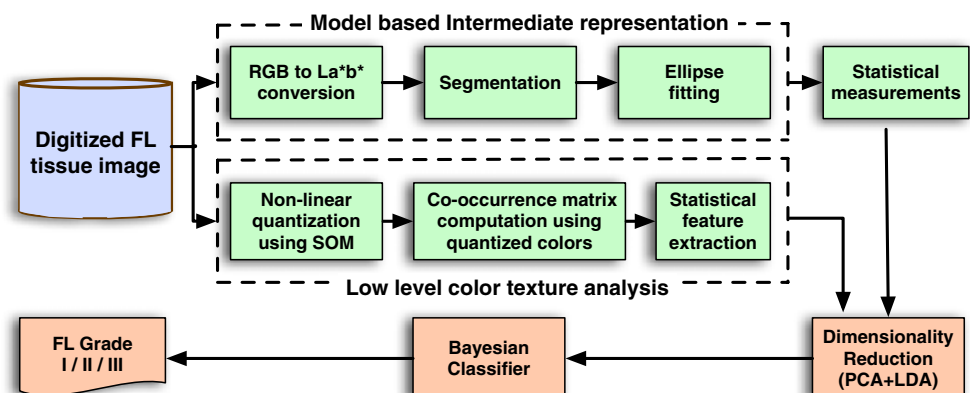
relative amount and spatial distribution of these cytological components. We observe that the spatial distribution of these regions vary considerably between different histological grades and using MBIR provides a convenient way to quantify our observations. Although this approach provides reasonable results especially identifying the most aggressive grade of FL, it is relatively less successful in classification of low grades. The tissue samples of these grades are better characterized by low level color texture features. Since gray-level features or other color texture features could not adequately model the microscopic tissue image content, we developed a non-linear color quantification based color texture feature constructed method. Due to the staining of the tissue samples, the resulting digitized FL images have considerably limited dynamic ranges in the color spectrum. Taking this fact into account, we propose the use of a non-linear color quantization using self-organizing maps (SOM). We used the quantized image to construct the co-occurrence matrix that is used to compute low level color texture features [12]. By combining the statistical features constructed from the MBIR with the low level color texture features, the classification performance of the system is improved significantly.

The rest of the paper is organized as follows. Section 2 describes the image dataset used in this study; feature construction and extraction methods for the grading of FL images as well as the statistical classification. Section 3 presents the experimental results. Finally, in Section 4, we conclude our results and point out the future research directions.

2 Feature Construction and Classification

We formulated the problem of grading FL images as a statistical pattern recognition problem. Using a number of image samples evaluated in consensus by five pathologists, we constructed a set of discriminative features and performed classification. We first segmented the image into basic cytological components based on the color information using an unsupervised clustering approach. This is followed by feature construction using MBIR and the color texture analysis. We used the combination of the principal components analysis (PCA) and the linear discriminant analysis (LDA) methods to reduce the dimensionality of the feature space. This step is followed by a Bayesian classifier based on maximum *a posteriori* decision rule.

In this section, we describe the dataset and the methods we used for automated grading of FL through image analysis.

2.1 Image Dataset

The input FL images to our system are H&E-stained tissue slides digitized using a Scope XT digitizer (Aperio, San Diego, CA, USA) at 40× magnification. Tissue slides are collected from the Department of Pathology, The Ohio State University in accordance with an IRB (Institutional Review Board) approved protocol. The image dataset consists of 17 whole-slide FL cases. A consensus of five hematopathologists evaluated the grades of these samples. Six of the samples were identified as grade I, eight were identified as grade II and three were identified as grade III.

The whole-slide images used in this study are representative samples and cover most of tissue variations among all three histological grades. In order to train the computerized system and test its accuracy, we asked three experienced pathologists to extract ten follicle regions that are equivalent of one HPF from all these whole-slide cases. In the rest of the manuscript, we will refer to these three sets of images provided by distinct pathologists as Sets 1, 2 and 3, respectively. In each set of images, there are 170 images extracted from 17 whole-slide images. This constructed a dataset of 510 images with a total of 180 grade I, 240 grade II and 90 grade III images each of which has a spatial resolution of $2{,}165 \times 1{,}365$ pixels that is equivalent to one microscopic HPF. For our training and testing of the computerized image analysis system, we used one set of images as training data and the remaining two sets of images as testing data and we repeat this three times each time using a different set of images.

2.2 Image Segmentation

There are five major cytological components in the FL tissue: nuclei, cytoplasm, extra-cellular material, red blood cells (RBC) and background regions. Having nuclei and cytoplasm regions dyed with hues of blue and purple, extra-cellular material dyed with hues of pink and RBCs dyed with hues of red, H&E-stained FL images provides useful visual clues for segmentation. In addition to these components, there are also background regions that do not correspond to any tissue component. With this *a priori* knowledge on FL images, we performed the segmentation using K-means clustering algorithm to identify these cytological components.

Instead of using the red–green–blue (RGB) color space, we converted the image to the La*b* color space, a perceptually uniform color space developed by Commission Internationale d'Eclairage (CIE). Perceptually uniform means that the same amount of change in color values produces the same amount of perceptual

difference of visual importance. This property of the La*b* color space allows us to use the Euclidean distance in comparing the colors [13]. Moreover, the La*b* color space separates the luminance and the chrominance information such that L channels corresponds to illumination and a* and b* channels correspond to color opponent dimensions. Hence, the feature vector for each pixel contains intensity and color information separately.

We performed the clustering in the La*b* color space using the *K*-means algorithm. RBCs and background regions show relatively uniform patterns; thus they are segmented by thresholding the intensity values in the RGB color space. The segmentation of RBCs is relatively straightforward by using a simple threshold operation in the RGB color space; therefore, to reduce the cost of the iterations and accelerate the convergence of the algorithm, we identified the RBC regions prior to the *K*-means clustering step and excluded such pixels. A pixelwise thresholding is used to determine the RBC regions as follows:

$$\text{RBC}(i,j) = \begin{cases} 1 \text{ if } \frac{r(i,j)}{r(i,j)+b(i,j)+c(i,j)} \geq \tau; \\ 0 \text{ otherwise}, \end{cases} \quad (1)$$

where *i*, *j* indicate the pixel coordinates; and r, g, and b are the corresponding red, green and blue color channels of the image in the RGB color space. The threshold τ is experimentally chosen as 0.45 after examining the histogram values of several representative regions.

Similarly, the background pixels were identified using a threshold as follows:

$$bckgrnd(i,j) = \begin{cases} 1 \text{ if } (r(i,j) \geq \xi) \& (g(i,j) \geq \xi) \& (b(i,j) \geq \xi); \\ 0 \text{ otherwise}, \end{cases} \quad (2)$$

where ξ was determined to be 200.

Due to the possible variations in the color spectrum of whole-slide images, we selected an unsupervised segmentation method as opposed to a supervised segmentation method. Since we are not limiting our system to training samples, using an unsupervised method, the segmentation step will be more robust to intensity variations. *K*-means [14] is an unsupervised learning method. It is an iterative process that divides samples into *k* partitions based on their attributes (e.g., La*b* color values) by minimizing a cost function:

$$J = \sum_{j=1}^{K} \sum_{i=1}^{N_j} \left| x_i^j - \mu_j \right|^2, \quad (3)$$

where *k* is the number of components to be identified, N_j is the number of samples in cluster *j*, x_i^j is the ith feature data ($i \in \{1, 2, \cdots N_j\}$) in the jth cluster, and μ_j is the centroid of the cluster *j*. Incorporating our prior knowledge of the cytological components in the FL slides, we set the number of clusters to be three that correspond to nuclei, cytoplasm and extra-cellular material. Figure 2, shows sample H&E-stained FL images associated with three different histological grades and

Figure 2 First row shows sample H&E-stained FL images associated with grades I, II and III from left to right, respectively. The corresponding segmentation results are shown in the *second row*. In these color labeled images, *blue* corresponds to nuclei, *cyan* to cytoplasm, *yellow* to extra-cellular material and *red* and *grey* to background and RBCs, respectively.
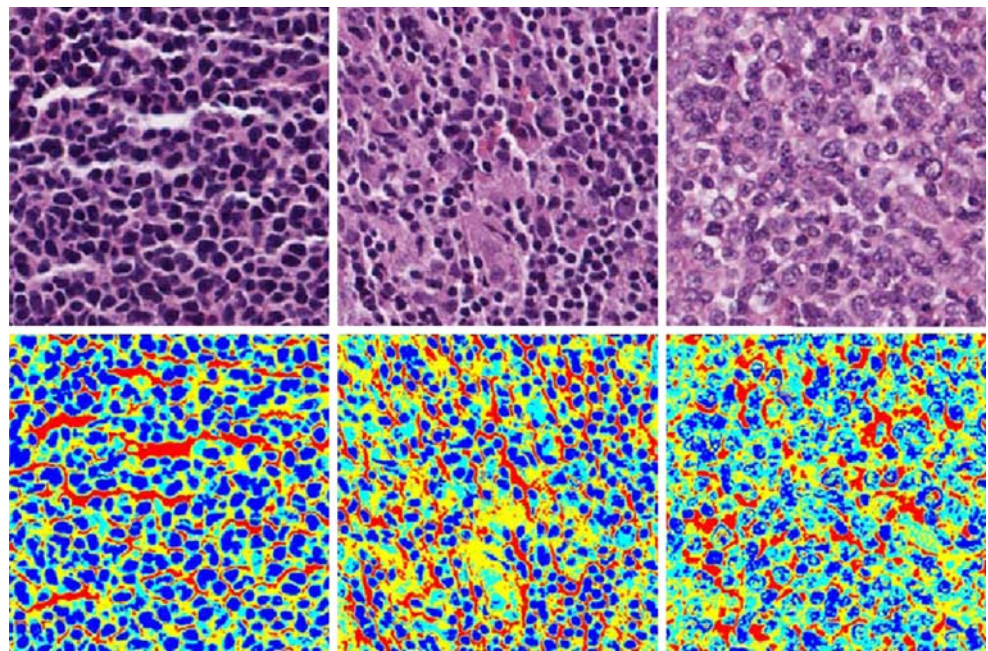
their segmentation results using the *K*-means algorithm as explained above.

2.3 Model Based Intermediate Representation (MBIR)

We observed that the spatial distribution of cytological components vary considerably with different histological grades. This is mostly due to the fact that there are larger number of CBs in higher grade FL samples. CBs are characterized by larger size, vesicular chromatin, accentuated nuclear membrane and one to three prominent nuclei as opposed to centrocytes characterized by disposed chromatin, inconspicuous nucleoli and scant cytoplasm. Therefore, images of higher histological grades have less homogeneous organization of nuclei and cytoplasm components with respect to their relative spatial distributions when compared to lower grade samples in which those regions are more compact and evenly distributed. In an attempt to capture this information, we introduced MBIR to analyze the spatial distribution of nuclei and cytoplasm components. Using the segmentation approach described in Section 2.2, we represented each nuclei and cytoplasm component using ellipses and computed a set of features to differentiate histological grades of FL.

The use of high level representations has been studied for content based image retrieval applications in remote sensing and medicine. Aksoy et al. proposed a Bayesian framework to construct a visual grammar to incorporate low-level features with high-level semantics [15]. They first segmented the images using the spectral and texture information and decomposed the image scene into prototype regions. Subsequently, they modeled the spatial relationships and interactions between these regions and performed the land cover classification. In a more recent study, Aksoy discussed several representations such as smoothed polygons, convex hull, grid representation and minimum bounding rectangle at different levels of complexity to simplify the computation of spatial relationships between regions [16]. In [17] authors proposed the use of grid-based layouts for semantic content analysis for histological images. Their system combines low-level image processing technology (i.e., segmentation) with high-level semantic analysis of medical image content similar to our abstraction using MBIR that corresponds to different cytological components.

Different from the region representation methods described in [16], we use ellipses to represent different cytological regions. Ellipses are accurate representations of 2D shapes and their computation is relatively efficient. Besides we can easily construct features using

their morphology such as the lengths of major and minor axis, and area as well as using their relative spatial distribution such as the number of neighbors and the minimum and average distances among neighbors. Below, we explain how we obtain the MBIR from the segmented images and construct morphological and topological features to be used differentiating histological grades.

After the segmentation process, we constructed the MBIR of the cytological regions in the tissue for further analysis. We only considered the nuclei and cytoplasm components since the variations in the rest of the components are random and intuitively do not provide any discriminative information. Segmentation step generates the binary representation of both components. Using the binary representation, we first applied a morphological pre-processing step to remove noisy pixels. We first removed insignificant regions by an area threshold (≈ 0.6 μm^2). This is followed by separating the touching components from each other for a more precise and accurate representation. For nuclei components, we used the watershed transform, a mathematical morphology based on partitioning [18], to separate touching nuclei. For the cytoplasm components, we incorporated our prior knowledge and associated each cytoplasm component to the closest nuclei component. We used the distance transform, which computes, for every cytoplasm pixel, the distance to the nearest nuclei pixel. By regrouping the connected components, we obtained the separated labels. It should be noted that there will also be some cytoplasm regions that are not associated with any nuclei because the corresponding nuclei of the particular cell associated with the cytoplasm component may be too scant, or the size of the nuclei component can be very small so that it has been discarded in the pre-processing step. Such cytoplasm components are still retained since they are important for the feature construction step.

Next, we fit an ellipse to represent each connected nuclei and cytoplasm region. We used the least squares ellipse fitting approach introduced in [19]. The authors proposed a closed form solution, which leads to a simple and robust solution with an efficient computational cost. Using this approach, we represented each connected region with a few parameters, i.e., the *x* and *y* coordinates of the center point, the length of major and minor axes and the angle between the x-axis and the major axis. Figure 3 shows the construction of MBIR for the nuclei component segmented in a sample grade II image. Similarly, we obtained the MBIR for cytoplasm components in the FL images. Figure 4 shows the MBIR for typical images associated with three different histological grades in each row.

a Original image **b** Segmented image

c Nuclei regions **d** MBIR representation
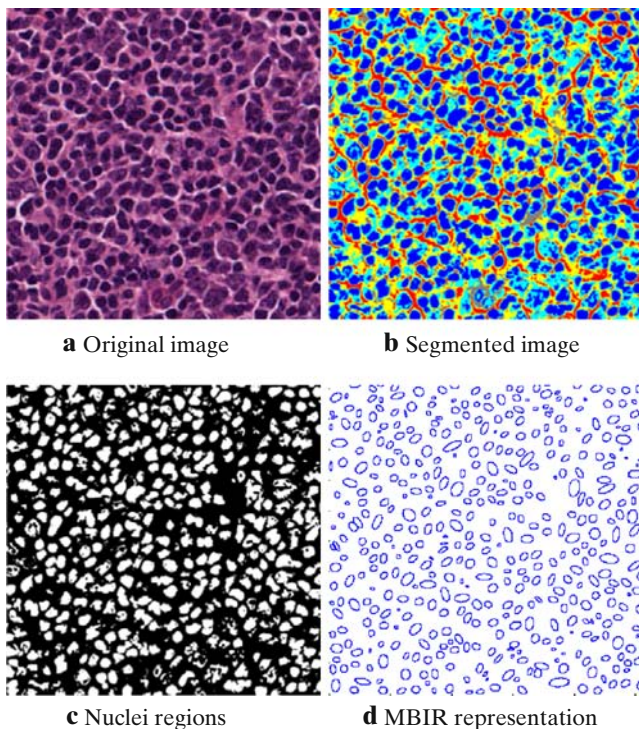
Figure 3 Construction of the ellipse representation of the nuclei component on a sample FL image associated with grade II.

Using the MBIR that correspond to nuclei and cytoplasm components, we introduced statistical measurements to capture the morphology and topology of these regions. Tables 1 and 2 give the list of features constructed using the MBIR. The morphological features attempts to capture the morphological differences in different grades, and consist of the length of major and minor axis and the area of ellipses. On the other hand, topological features capture the variations in the spatial distribution of these components. We computed the number of neighboring ellipses, the minimum and the average distances to these neighbors for each component. We defined the neighborhood as a circular region with a radius of 20 pixels ($\approx 5\,\mu m^2$) around the center of each ellipse and counted other regions whose center is in this area as neighbors. We also treated nuclei and cytoplasm components independently and computed the neighboring nuclei and cytoplasm components for each nuclei and cytoplasm regions separately. Figure 5 illustrates the feature construction step using MBIR.

After obtaining all these measurements for each component, we had the representation of each image in terms of distributions of these measurements (e.g. L_{major}, $\mathcal{D}_{\mathrm{nuc-nuc}}^{\mathrm{min}}$, $\mathcal{D}_{\mathrm{cyt-cyt}}^{\mathrm{avg}}$, etc.) associated with each cytological component as shown in Fig. 5b. First, we obtain an approximate probability mass function (pmf) for each feature. Using the pmfs, we compute sample mean

($\overline{\mu}$), standard deviation (σ^2), skewness (γ_1), kurtosis (γ_2) and entropy (E) and used these values to perform subsequent analysis for grading.

Table 2 summarized the topological features and indicates which components are used for these measurements using subscript notation such that $\mathcal{N}_{\mathrm{nuc-cyt}}$ indicates the number of neighbor cytoplasm regions around each nuclei. Similarly, $\mathcal{D}_{\mathrm{cyt-cyt}}^{\mathrm{avg}}$ indicates the average distance between the neighboring cytoplasm components for each cytoplasm region. The base component for which these measurements are computed is also given in parenthesis.

2.4 Color Texture Analysis Using Self-Organizing Maps (SOM)

In addition to the morphological and topological features computed using MBIR, we also introduced a novel color texture analysis approach using the self-
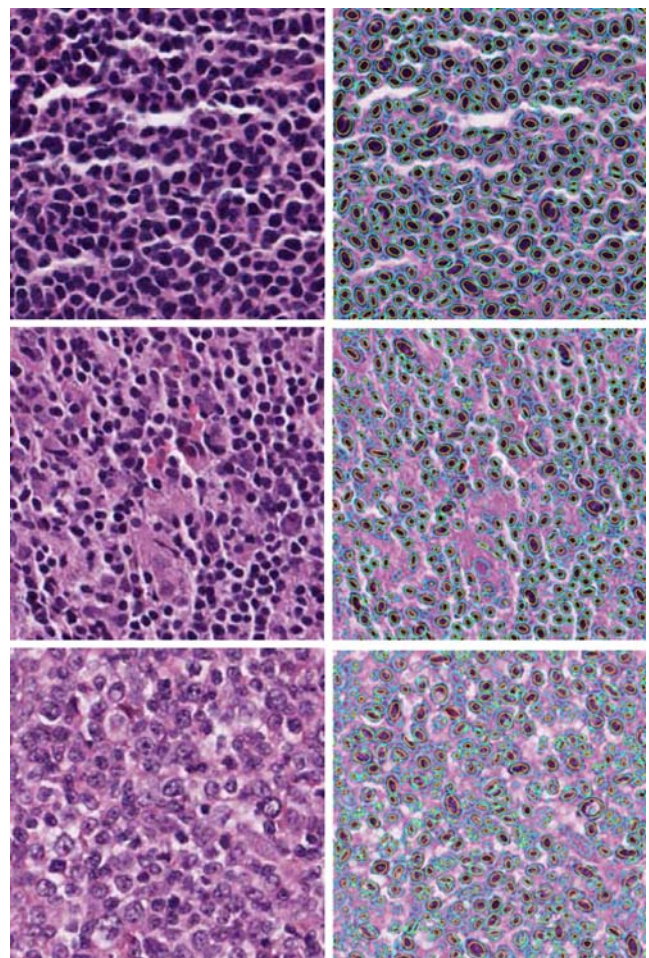
Figure 4 Representation of different cytological regions in the images associated with different histological grades using ellipses. *Yellow* and *cyan* colored ellipses represent regions that correspond to nuclei and cytoplasm regions, respectively.

Table 1 Measurements based on the morphology of for nuclei and cytoplasm components using MBIR.

|	
L_{major}	Major axis length
L_{minor}	Minor axis length
A	Area

organizing maps (SOM). We exploited the fact that the H&E-stained images have a considerably limited dynamic range in the color spectrum. Hence, we propose to apply a non-linear quantization using SOM and construct the co-occurrence features from the quantized image.

Previously, Shim and Choi proposed the use of a simplified color co-occurrence matrix, which represents the number of color pairs between adjacent pixels to index the images for content based retrieval applications [20]. In another study, Arvis et al. compared three different ways of incorporating color and texture information using the co-occurrence matrices [21]. First they considered using the grey scale texture features jointly with the color features at the time of the classification. Secondly, they constructed the grey scale co-occurrence texture features on a previously quantized color image using uniform quantization. Finally, they introduced the multi-spectral extension, where they computed several co-occurrence matrices between and within the color bands. With the multi-spectral approach being the most accurate in classifying texture images, the authors concluded that the color-texture approach increases the performance remarkably. For our application of FL grading, the proposed approach is more intuitive than the uniform quantization-based methods because

Table 2 Measurements based on the spatial distributions of nuclei and cytoplasm components using MBIR.

$\mathcal{N}_{nuc-nuc}$	# neighboring nuclei components (nuclei)
$\mathcal{N}_{nuc-cyt}$	# neighboring cytoplasm components (nuclei)
$\mathcal{N}_{cyt-cyt}$	# neighboring cytoplasm components (cytoplasm)
$\mathcal{D}_{nuc-nuc}^{min}$	Closest distance to a nuclei component (nuclei)
$\mathcal{D}_{nuc-cyt}^{min}$	Closest distance to a cytoplasm component (nuclei)
$\mathcal{D}_{cyt-cyt}^{min}$	Closest distance to a cytoplasm component (cytoplasm)
$\mathcal{D}_{nuc-nuc}^{avg}$	Average distance to neighboring nuclei components (nuclei)
$\mathcal{D}_{nuc-cyt}^{avg}$	Average distance to neighboring cytoplasm components (nuclei)
$\mathcal{D}_{cyt-cyt}^{avg}$	Average distance to neighboring cytoplasm components (cytoplasm)

These features summarize the topology of the components of the tissue.

a Feature construction using MBIR **b** Histograms
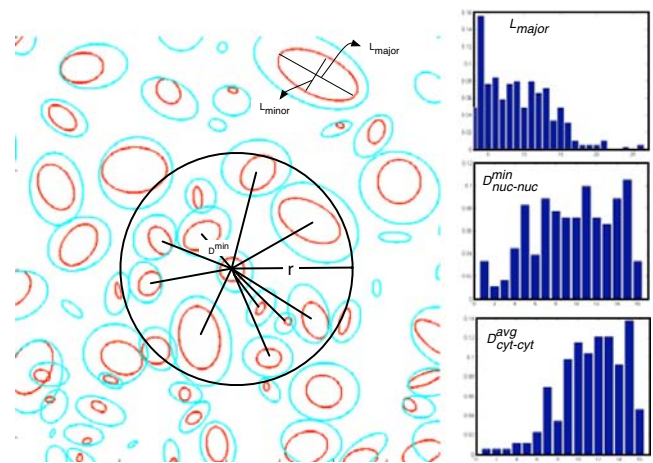
Figure 5 Construction of the morphological and topological features using MBIR and sample histograms for some of the features given in Tables 1 and 2.

the H&E images have considerably limited color spectrum, i.e., there are very few dominant colors (hues of blue, purple and pink) as shown in sample images in Fig. 2. Therefore, instead of using a uniform quantization, we can enhance the image by using a non-linear quantization.

In order to construct the co-occurrence matrix, we first utilized SOM method for quantizing image colors in a non-linear way. SOM is an artificial neural network that is trained using unsupervised learning [22]. Suppose, C is the set of all colors in an image in the RGB color space. Let C be denoted by $C = \{c_j, j = 1, 2, \cdots, N\}$ where N is the number of colors in the image. Color quantization is the process of choosing K colors from C to construct the quantized color space, $\overline{C}$ where $\overline{C} = \{\overline{c}_j, j = 1, 2, \cdots, K\}$, where $K \leq N$.

SOM is an iterative vector quantization method based on competitive learning, where the quantization vector consisting of neurons is initialized and each neuron is selectively tuned. In addition to the random initialization step, there are two basic steps involved in the algorithm at each iteration step: similarity matching and updating.

- **Initialization:** Assign random values for the initial weight vector, $\overline{C}(0)$. Here, the only constraint is to prevent any two entries having the same values.
- **Similarity Matching:** For every color, C_i in the image, compute the best matching neuron, $q(\overline{C})$ using the Euclidean distance as follows:

$$q(\overline{C}_{i*}) = \underset{j}{\arg\min} \left| C_i - \overline{C}_j \right|, j = 1, 2, \cdots, K \qquad (4)$$

– **Updating:** The weight vectors of all the neurons are updated as follows:

$$\overline{C}_j(n+1) = \begin{cases} \overline{C}_j(n) + \eta(n)\left[C_{i^\star} - \overline{C}_j(n)\right] & j \in A_{q(\overline{C})}(n); \\ \overline{C}_j(n) & \text{otherwise,} \end{cases}$$
(5)

where n is the iteration number, $\eta(n)$ is the learning rate parameter which decreases monotonically with the number of iterations, $A_{q(\overline{C})}$ is the neighborhood function whose shape is Gaussian and similarly decreases monotonically with the number of iterations and $\overline{C}_j$ is the best matching neuron. The iterations stop after a predefined number of iterations or until there is no significant change in the color map $\overline{C}$.

As a result of the neighborhood function, the final quantization vector is always ordered, i.e., neurons are organized into a meaningful order in which similar neurons in color are closer to each other in the grid than the ones with different colors. The ordered quantization vector is a desired property for our application since we will be using the labels of the pixels that map the colors in the image to this quantized vector. Figure 6 shows the resulting quantized images using uniform gray level quantization, uniform color quantization and non-uniform color-quantization using the SOM all in pseudo-colored format. In all three cases, the resulting images have 64 levels. The quantization vector computed by the SOM is also given on the top of the third column. It can be seen that the uniform color quantiza-

tion based method gives the least precise result in terms of image distortion. This could be explained by the correlation between the color channels. However, this drawback is overcome using the SOM for quantization. In addition, the quantized image obtained using the SOM provided a more enhanced image compared to the quantized grey level image, thus resulting in more descriptive texture information [12].

Co-occurrence matrices are one of the most commonly used representation in texture classification [11]. Let $\{I(x, y), 0 \leq x \leq N - 1, 0 \leq y \leq N - 1\}$ denotes a grey scale image I, we can construct the grey-level co-occurrence matrix, P_d, that gives the joint distribution of pair of pixels with a given spatial relationship determined by the displacement vector $d(dx, dy)$ as follows:

$$p_d(i, j) = |\{((r, s), (t, v)) : I(r, s) = i, I(t, v) = j\}|,$$
(6)

where $(r, s), (t, v) \in N \times N$, $(t, v) = (r+dx, s+dy)$, and $|\cdot|$ is the cardinality of a set. Subsequently, a set of features is constructed from this matrix to describe the texture. In addition to the features constructed by analysing the regions of cytological components using ellipse representation, we constructed low level color texture features using color co-occurrence matrices. In Table 3 we summarize these low level color texture features extracted from the co-occurrence matrix constructed directly from the color information using the proposed non-linear quantization method.

Figure 6 Quantized images using uniform gray level quantization, uniform color quantization and non-uniform color-quantization using SOM. All resulting images are shown using pseudo-coloring with the color map shown at *lower-left corner*.
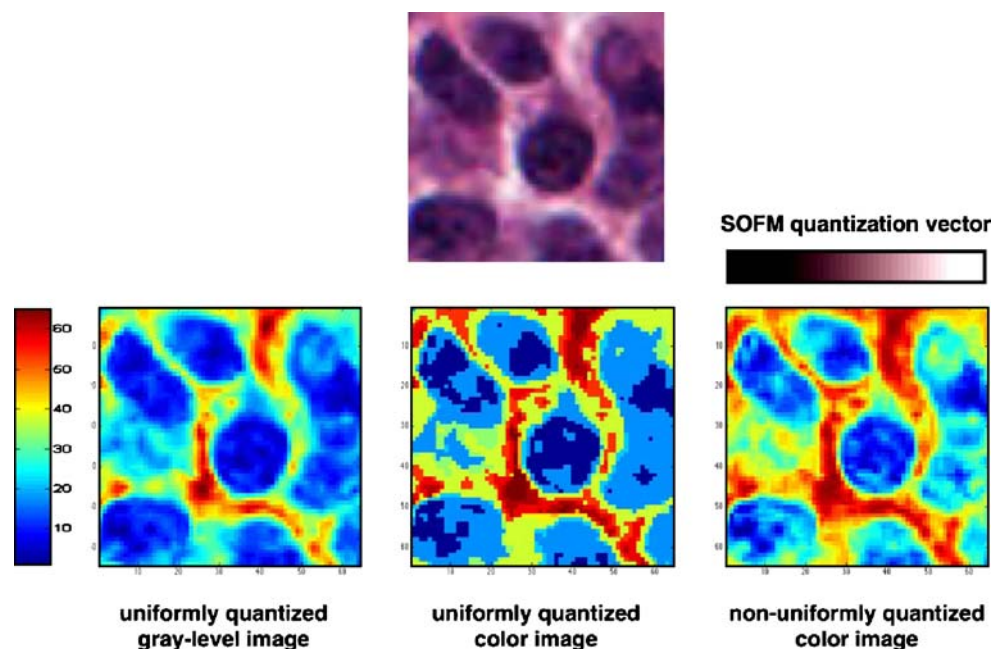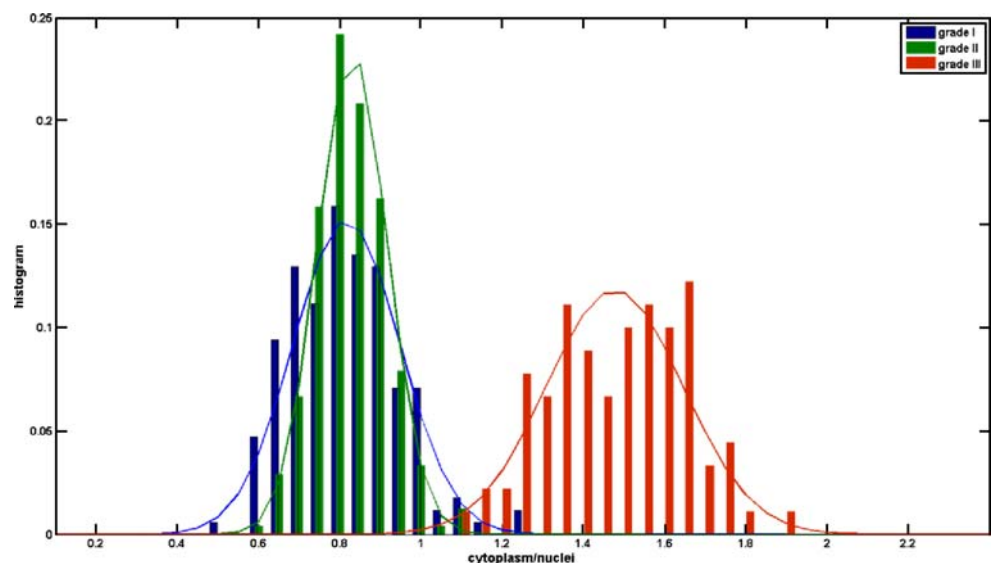
Table 3 Low level color texture features extracted from color co-occurrence matrix representation.

Homogeneity	The degree of homogeneity
Energy	The measure of uniformity
Contrast	The range of color distribution
Correlation	The correlation between pixel values and its neighborhood
Entropy	The measure of randomness

3 Experimental Results

In this section, we present the classification performances of the proposed study using the MBIR and color texture approaches mentioned in the previous section. Our first goal is differentiating high grades (grade III) from the low and intermediate grades (grade I and II). Due to the fact that higher grades contain higher number of large malignant cells, we expected to observe differences between the amount of relative cytological components (e.g. nuclei and cytoplasm). This is because larger cells (i.e., CBs) have open chromatin with smaller nuclei and larger cytoplasm as compared to the compact cells (i.e., centrocytes) with larger nuclei and scant cytoplasm. Therefore, ratio of the amount of cytoplasm to the amount of nuclei, ζ, vary considerably between low and high grades since the proportion of large to small cells increases from low to higher grades. We quantify this observation from the digitized FL images based on our segmentation method as follows:

$$\zeta = \frac{\sum_{i,j} \chi}{\sum_{i,j} \nu} \qquad (7)$$

where i, j indicate the pixel coordinates and χ, ν are the cytoplasm and nuclei pixels, respectively. Accordingly, we constructed this feature from the images in our dataset and analyzed the results. Figure 7 shows the distribution of FL cases among different histological grades using this ratio feature. The histograms are plotted together with the estimated normal distributions for samples of grades I, II and III, in blue, green and red colors, respectively.

Subsequently, we used the maximum a posteriori decision rule for the classification as follows:

$$W = \underset{j}{\operatorname{argmax}} \; p(x|\omega_j) P(\omega_j) \qquad (8)$$

where $P(\omega_j)$ is the prior probability of $j^t h$ class ($j \in \{1, 2 \cdots C\}$) and $p(x|w_j)$ is the class conditional probability of the sample x to be classified. In our case of classifying low and high grades of FL, the number of classes, $C = 2$. In Table 4 we summarize the classification results in terms of training and testing accuracies. Each time we used one set of images provided by a particular pathologist as training samples and the remaining two sets of images provided by the other two pathologists as testing samples. Thus, we have 170 training and 340 test images for every testing instance. Training accuracies are computed using a k-fold cross validation, where $k = 5$. The sensitivity and the specificity of the system identifying grade III samples are 98.7% and 98.9%, where we computed these values by taking the average of true positive, true negative, false positive and false negative classifications using different configurations of training and testing sets. Although this feature was found to be very discriminative in separating high grades (III) from lower grades (I and

Figure 7 The distribution of ζ, the ratio of the amount of cytoplasm to the amount of nuclei, among the FL cases associated with three different histological grades.

Table 4 Classification accuracies (in percentage) on training and testing samples using ζ, the amount of cytoplasm to nuclei to differentiate low and intermediate grade samples from high grade samples.

	Train: Set 1 Test: Set 2, 3	Train: Set 2 Test: Set 1, 3	Train: Set 3 Test: Set 1, 2
Training accuracy (%)	99.4	97.1	98.8
(Grade I, II- III)	(99.3–100)	(96.4–100)	(98.5–100)
Testing accuracy (%)	98.4	98.8	98.8
(Grade I, II–III)	(98.5–98.3)	(98.5–100)	(98.9–98.3)

Columns represent different test instances, where each time images provided by one pathologist used as training and rest as testing. Set 1, 2 and 3 refers to the set of images provided by individual pathologists. The accuracies in parenthesis correspond to classification accuracies of individual classes (i.e., low & intermediate grades vs. high grade).

II), it is not sufficient to discriminate grade I and II samples.

Next, we analyzed the the morphological and topological features computed from the MBIR of cytological components (i.e., nuclei and cytoplasm). These features consist of the sample mean, standard deviation, skewness, kurtosis and entropy of the estimated pmfs of 15 different measurements summarized in Tables 1 and 2 providing a feature vector of length 75 elements. Therefore, we first applied dimensionality reduction step prior to the actual classification. We considered using the principal components analysis (PCA) and linear discriminant analysis (LDA) methods for this purpose [23].

Without loss of generality, the PCA, which is an unsupervised method, provides a compact representation by projecting the data onto a new space, such that the first principal component is associated with the largest eigen-values of the covariance matrix. The largest eigen-value corresponds to the direction where data have the largest variance while the smallest one corresponds to the smallest variance. Hence, we can

select the projection matrix composed of the first few eigen-vectors associated with the largest eigen-values. By doing so, we can reduce the dimensionality of our feature space yet preserve the most of the variance.

On the other hand, the LDA is a supervised method; thus requires class labels for each data sample. It maps the data onto a lower dimensional subspace that best discriminates data. The goal is to find the mapping, where the sum of distances between samples in different classes is maximized; while the sum of distances between samples in same classes is minimized. These measures are captured, using the between-class scatter matrix, S_b and the within class scatter matrix, S_w [24]. The resulting projection is given by the leading eigen-vectors of matrix $S_w^{-1} S_b$. For a classification problem of C classes, this matrix has at most $C - 1$ nonzero eigen-values, defining a feature subspace of $C - 1$ dimensions. There is no certain conclusion stating whether PCA or LDA is superior to the other. Depending on the distribution of data, either one can perform better. A comprehensive discussion of PCA and LDA can be found in [25].

Table 5 Classification accuracies (in percentage) is reported using the Bayesian classifier and KNN over the feature set constructed using MBIR.

	Train: Set 1 Test: Set 2, 3	Train: Set 2 Test: Set 1, 3	Train: Set 3 Test: Set 1, 2
PCA			
Bayesian	77.3	78.8	75.0
	(79.1–73.1–85.0)	(60.0–85.0–86.7)	(69.2–80.6–71.7)
KNN	76.7	78.1	74.4
	(71.5–73.8–93.8)	(65.4–80.6–95.0)	(63.8–74.2–96.3)
PCA + LDA			
Bayesian	78.2	78.2	73.2
	(73.6–74.4–96.7)	(58.2–84.4–98.3)	(58.3–75.6–96.7)
KNN	76.7	78.1	72.6
	(72.7–73.7–91.6)	(53.6–88.1–96.6)	(56.7–78.7–88.3)

Each column represents a different testing instance, where we used samples provided by one pathologist (indicated by either Set 1, 2 or 3) as training samples and rest as testing samples. Two different dimensionality reduction methods are used that are PCA and PCA followed by LDA. The accuracies in parenthesis correspond to classification accuracies of individual grades.

Table 6 Classification accuracies (in percentage) using the extended feature space together with the low level color texture features.

	Train: Set 1 Test: Set 2, 3	Train: Set 2 Test: Set 1, 3	Train: Set 3 Test: Set 1, 2
PCA			
Bayesian	81.2	83.6	79.1
	(77.3–81.3–88.3)	(72.7–90.6–85.0)	(70.8–88.8–70.0)
KNN	80.1	82.3	77.4
	(73.4–78.6–96.3)	(78.1–81.5–92.3)	(69.3–78.5–91.1)
PCA + LDA			
Bayesian	83.3	86.7	80.6
	(79.1–80.6–98.3)	(79.1–87.5–98.3)	(66.7–86.3–93.3)
KNN	82.1	86.3	78.2
	(81.8–77.5–95.0)	(78.1–87.5–98.3)	(55.8–92.5–85.0)

Similarly, two different dimensionality reduction methods are used that are PCA and PCA followed by LDA.

In Table 5, we report the classification results using the features constructed from MBIR. We used two different methods for dimensionality reduction that are PCA and PCA+LDA. We kept the first few principal components, whose corresponding eigen-values sum up to 92% of the total variance after applying PCA. We also performed the classification using two different classifiers Bayesian and *K*-nearest neighbor (KNN) classifiers. The Bayesian classifier maximizes the posterior probability, which is a function of likelihood and prior probability, as in Eq. 8. The KNN classifies each pattern by a majority vote of its neighbors, with the pattern being assigned to the class that is most common amongst its *k*-nearest neighbors. Each classifier has its

own benefits and drawbacks. Bayesian classifier comes at low computational cost, but it assumes a Gaussian data distribution. On the other hand, the KNN classifier does not assume any particular distribution and it is based on non-parametric density. However, the computational cost is higher since it requires the computation of pairwise distances to every training pattern to determine the nearest neighbors. We report the results over three different testing sets such that for each experiment we used images provided by one pathologist for training and the rest for independent testing and reported the corresponding classification accuracies. The best classification performances using the MBIR approach are indicated in boldface in Table 5. As can

Figure 8 Decision boundaries determined by the Bayesian classifier.
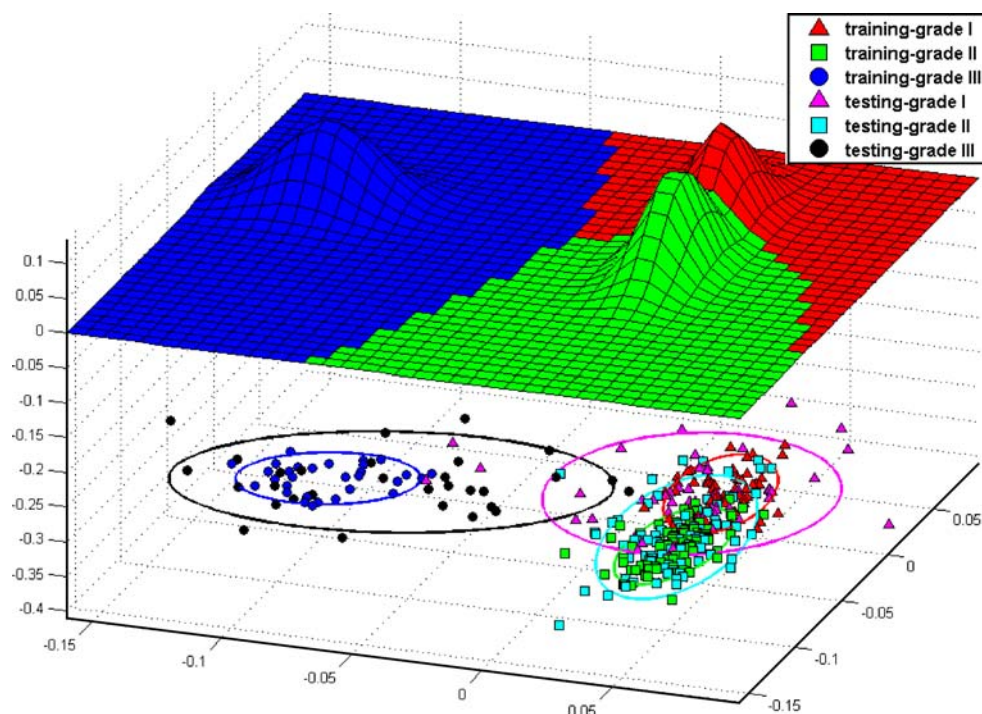
be seen from these results, the MBIR provided remarkable discrimination power parallel to our observations with the best overall classification accuracy of %78.8. As a dimensionality reduction method, PCA provided better results than PCA+LDA in most of the experiments using the MBIR features. Also, the accuracy of the Bayesian classifier was higher than the KNN classifier in every experiment.

In addition to MBIR features, we incorporated low level texture features into our system. We used the conventional co-occurrence matrix approach to construct statistical texture features. As explained in Section 2.4, we introduced the use of SOM to apply non-linear color quantization and constructed the color co-occurrence matrices to exploit specific color content of H&E-stained FL images. We extracted the homogeneity, energy, contrast, correlation and entropy features using the color co-occurrence matrix representation. These low-level color texture features are then combined with the MBIR features and we similarly applied the same analysis using dimensionality reduction and classification. The classification accuracies over the extended feature space are reported in Table 6. With the low level texture features, the classification accuracy especially in identifying grades I and II is improved considerably and as a result, the overall classification accuracy is increased with the best overall classification performance of %86.7. In this extended feature space, the PCA+LDA approach outperformed the PCA only dimensionality reduction method, however, the performances are not statistically different from one another.

In addition to the high classification accuracy, our approach was successful in predicting the underlying distribution of the testing samples with a small deviation. This shows that the extended feature space is capable of robustly capturing the discriminative features among histological grades of FL. In Fig. 8 we show the estimated probability distributions of samples of each grade and the Bayesian decision boundaries.

4 Conclusions

This study presents an automated system for the quantitative analysis of the histopathological images of FL. We introduced a model-based intermediate representation (MBIR) that allows describing the cytological components of the tissue in a quantitative way. This representation is very rich and an extensive set of features can be extracted from this intermediate representation. Using MBIR, we constructed features that characterizes the morphology and topology of nuclei and cytoplasm components in the tissue and allows dis-

crimination of histological grades for FL. Additionally, we introduced a novel color texture analysis method. Based on the fact that the H&E stained tissue images have a considerably limited dynamic range in the color spectrum, we proposed the use of a non-linear quantization of color images using SOM prior to co-occurrence calculation. This is also applicable to other natural images with a priori limited color spectrum.

Although we achieved reasonable performances using the MBIR features, the best overall accuracy was obtained by fusing this higher level representation of cytological components with low level color texture featues. The overall classification accuracy of the system in identifying histological grades of I, II and II was 74.9%, 84.6% and 95.0%, respectively with a remarkable overall classification accuracy of 83.7%.

The histopathological grading process for FL prognosis can be improved using the proposed quantitative tools in order to minimize the variabilities between individual readers. Additionally, our computerized system may help preventing the sampling bias problem since it is practical to evaluate whole-slide images in reasonable times with the help of computers.

In our future work, we will detect and segment individual CBs to further improve the performance of the classification of lower grades of FL. A semantic graph representation of MBIR will be enhanced with graph theoretical measures to provide another rich set of features to characterize the topology and morphology of the tissue. Additionally, we will extend our MBIR framework to support semantic queries that will allow extracting inferences on the cytological content of tissues. Considering the large-scale of these images, computer-aided characterization of semantic representation can provide unprecedented ways of characterizing and analyzing these type of images.

References

1. Jaffe, E. S., Harris, N. L., Stein, H., & Vardiman, J. W. (2001). *Tumours of haematopoietic and lymphoid tissues*. Lyon: IRAC Press.
2. Metter, G. E., Nathwani, B. N., Burke, J. S., Winberg, C. C., Mann, R. B., Barcos, M., et al. (1985). Morphological sub-

classification of follicular lymphoma: Variability of diagnoses among hematopathologists, a collaborative study between the repository center and pathology panel for lymphoma clinical studies. *Journal of Clinical Oncology, 3*, 25–38.

3. Dick, F., Van Lier, S., Banks, P., Frizzera, G., Witrak, G., Gibson, R., et al. (1987). Use of the working formulation for non-Hodgkin's lymphoma in epidemiological studies: Agreement between reported diagnoses and a panel of experienced pathologists. *Journal of National Cancer Institue, 78*, 1137–1144.

4. The Non-Hodgkin Lymphoma Classification Project. (1997). A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin lymphoma. *Blood, 89*, 3909–3918.

5. Meijer, G. A., Belien, J. A. M., Diest, P. J., & Baak, J. P. A. (1997). Image analysis in clinical pathology. *Journal of Clinical Pathology, 50*, 365–370.

6. Tabesh, A., Teverovskiy, M., Pang, H., Kumar, V. P., Verbel, D., Kotsianti, A., et al. (2007). Multifeature prostate cancer diagnosis and Gleason grading of histological images. *IEEE Transaction on Medical Imaging, 26*, 1366–1378.

7. Petushi, S., Garcia, F. U., Haber, M. H., Katsinis, C., & Tozeren, A. (2006). Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Medical Imaging, 6*, 14.

8. Demir, C., Gultekin, S. H., & Yener, B. (2005). Augmented cell-graphs for automated cancer diagnosis. *Bioinformatics, 21*, 7–12.

9. Gurcan, M. N., Pan, T., Shimada, H., & Saltz, J. H. (2006). Image analysis for neuroblastoma classification: Segmentation of cell nuclei. In *Int. conf. of the IEEE engineering in medicine and biology society (EMBC)*. New York, NY.

10. Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H., & Gurcan, M. N. (2007). Computer-aided grading of neuroblastic differentiation: Muti-resolution and multi-classifier approach. In *Proc. of the int. conf. on image processing (ICIP)*. (pp. 525–528). San Antonio, TX.

11. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transaction on Systems, Man, and Cybernetics, 3*, 610–621.

12. Sertel, O., Kong, J., Lozanski, G., Shanaah, A., Catalyurek, U., Saltz, J., et al. (2008). Texture classification using nonlinear color quantization: Application to histopathological image analysis. In *Proc. of the IEEE int. conf. on acoustics, speech, and signal processing (ICASSP)* (pp. 597–600). Las Vegas, NV.

13. Paschos, G. (2001). Perceptually uniform color spaces for color texture analysis: an empirical evaluation. *IEEE Transaction on Image Processing, 10*, 932–937.

14. MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. In *Proc. of 5th Berkeley symp. on mathematical statistics and probability*. (vol. 1, pp. 281–297) Berkeley, Univ. of California Press.

15. Aksoy, S., Koperski, K., Tusk, C., Marchisio, G., & Tilton, J. C. (2005). Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Transactions on Geoscience and Remote Sensing, 43*, 581–589.

16. Aksoy, S. (2006). Modeling of remote sensing image content using attributed relational graphs. *Lecture Notes in Computer Science, 4109*, 475–483.

17. Tang, H. L., Hanka, R., & Ip, H. H. S. (2003). Histological image retrieval based on semantic content analysis. *IEEE Transaction on Information Technology in Biomedicine, 7*, 26–36.

18. Meyer, F. (1994). Topographic distance and watershed lines. *Signal Processing, 38*, 113–125.

19. Halir, R., & Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In *Proc. int. conf. in central Europe on computer graphics, visualization and interactive digital media* (pp. 125–132).

20. Shim, S., & Choi, T. (2003). Image indexing by modified color co-occurrence matrix. In *Proc. of IEEE int. conf. on acoustics, speech, and signal processing (ICASSP)*.

21. Arvis, V., Debain, C., Berducat, M., & Benassi, A. (2004). Generalization of the co-occurrence matrix for color images: Application to color texture classification. *Image Analysis and Stereology, 23*(1), 63–72.

22. Kohonen, T. (1997). *Self-organizing maps*. Heidelberg: Springer.

23. Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. Academic.

24. Duda, R. O., Hart, P. E., Stork, G. (2001). *Pattern classification*. Wiley.

25. Martinez, A. M., & Kak, A. C. (2001). PCA versus LDA. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 23*, 228–233.

Olcay Sertel received his BSc. degree from Yildiz Technical University, Istanbul, Turkey, in 2004 and MSc. degree from Yeditepe University, Istanbul, Turkey, in 2006, both in computer engineering. He is currently a Ph.D. student at the Dept. of Electrical and Computer Engineering and a research assistant at the Dept. of Biomedical Informatics, The Ohio State University, Columbus, OH. His research interests include computer vision, image analysis and understanding, and pattern recognition with applications in medicine.

Jun Kong received the B.S. in Information and Control and M.S. in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and 2004. He is currently a Ph.D. student in Dept. of Electrical and Computer Engineering at Ohio State University, Columbus, Ohio. His research interests include computer vision, machine learning, and pathological image analysis.

Gerard Lozanski is an Assistant Professor in the Department of Pathology and the Medical Director of Clinical Flow Cytometry Laboratory at The Ohio State University Medical Center. He received his M.D. degree from the Medial University School, Poznan. Poland. His research interests include diagnosis, risk stratification and biology of non-Hodgkin lymphomas with a special interest in chronic lymphocytic leukemia, follicular lymphoma and T cell non-Hodgkin lymphomas.

Umit V. Catalyurek is an Associate Professor in the Department of Biomedical Informatics at The Ohio State University and has a joint faculty appointment in the Department of Electrical and Computer Engineering. His research interests include combinatorial scientific computing, grid computing, and runtime systems and algorithms for high-performance and data-intensive computing. He received his PhD, M.S. and B.S. in Computer Engineering and Information Science from Bilkent University, Turkey, in 2000, 1994 and 1992, respectively.

Joel H. Saltz is the Professor and Chair of the Department of Biomedical Informatics, Professor in the Department of Computer Science and Engineering at The Ohio State University (OSU), Davis Endowed Chair of Cancer at OSU, and a Senior Fellow of the Ohio Supercomputer Center. He received his M.D.-Ph.D. degrees in Computer Science at Duke University. Dr. Joel Saltz has developed a rich set of middleware optimization and runtime compilation methods that target irregular, adaptive, and multi-resolution applications. Dr. Saltz is also heavily involved in the development of ambitious biomedical applications for high-end computers, very large-scale storage systems, and grid environments. He has played a pioneering role in the developing of Pathology virtual slide technology and has made major contributions to informatics applications that support point-of-care testing.

Metin N. Gurcan is an Assistant Professor at the Department of Biomedical Informatics at The Ohio State University. He received his M.Sc. degree in Digital Systems Engineering from UMIST, England and his B.Sc. and Ph.D. degrees in Electrical and Electronics Engineering from Bilkent University, Turkey. He was a postdoctoral research fellow and later a research investigator in the Department of Radiology at the University of Michigan, Ann Arbor. Prior to joining The Ohio State University in October 2005, he worked as a senior researcher and product director at a high-tech company, specializing in computer-aided detection and diagnosis of cancer from radiological images. Dr. Gurcan's research interests include image analysis and understanding, computer vision with applications to medicine. He is the author of over 50 peer-reviewed publications and has a patent in computer-aided diagnosis in volumetric imagery.