A Combined Computerized Classification System for Whole-slide Neuroblastoma Histology: Model-based Structural Features

Olcay Sertel\textsuperscript{1,2}, Umit V. Catalyurek\textsuperscript{1,2}, Hiroyuki Shimada\textsuperscript{3}, and Metin N. Gurcan\textsuperscript{2}

\textsuperscript{1} Dept. of Electrical & Computer Eng., The Ohio State Univ., Columbus, OH, USA.
\textsuperscript{2} Dept. of Biomedical Informatics, The Ohio State Univ., Columbus, OH, USA.
\textsuperscript{3} Keck School of Medicine, The Univ. of Southern California, Los Angeles, CA, USA.

\textbf{Abstract.} Neuroblastoma (NB) is one of the most malignant tumors affecting infants and children. In current clinical practice, NB prognosis and further treatment planning highly relies on histopathological examination of tissue samples. The International Neuroblastoma Pathology Committee has adopted the Shimada classification system, which relies on several morphological characteristics of the tissue such as the degree of Schwannian stromal development and the grade of neuroblastic differentiation to categorize the tissue sample as either favorable or unfavorable histology. In this study, we present a combined computer-aided prognosis system that integrates these two diagnosis processes within one analysis framework. The proposed system first segments the digitized H&E-stained tissue image into eosinophilic and basophilic structures using the expectation maximization algorithm. For the classification between different tissue subtypes, in addition to conventional co-occurrence texture features, we propose a novel set of structural features that capture higher-level perceptual patterns. We evaluated the developed system over an independent set of 34 whole-slide images and achieved a classification accuracy of 94.1\% (32/34).

1 Introduction

Mainly affecting infants and children, neuroblastoma (NB) is a cancer of nervous system [1]. NB prognosis and further treatment planning highly relies on histopathological examination of tissue samples. However, the outcome of this qualitative visual examination process is subject to considerable inter- and intra-reader variations [2]. Besides it is often quite time-consuming especially when a large number of slides need to be diagnosed in practice. The International Neuroblastoma Pathology Committee adopted the Shimada classification system, which categorizes NB tumor as either favorable or unfavorable histology [3]. According to Schwannian stromal development and the grade of differentiation, the two most salient morphological indicators in the Shimada system, NB histology is categorized into four different tissue subtypes: stroma-rich (SR), differentiating (D), poorly-differentiated (PD) and undifferentiated (UD). Figure 1 (a)-(d)
Fig. 1. Sample representative region of interest images viewed at 20× magnification level; (a) Stroma-rich (SR), (b) Differentiating (D), (c) Poorly-differentiated (PD), (d) Undifferentiated (UD); and their corresponding segmentation results shown in (e)-(h), respectively. Cyan, yellow, blue and white correspond to eosinophilic and basophilic structures, RBCs and background, respectively.}

show typical region of interest images cropped from a digitized H&E-stained whole-slide tissue associated with each subtype viewed at 20× magnification.

Visual observation of patterns, shapes, sizes, relationship between cytological components, and other disease specific features can be analyzed using image analysis and pattern recognition techniques. Several research studies demonstrate remarkable accuracies in classifying tissue subtypes associated with different grades of malignancies using quantitative textural, structural, morphological and topological features for different tumors such as prostate, breast, brain and lymph system [4–7]. In our recent work, we presented computer-aided prognosis of NB, tackling the classification of Schwannian stromal development and the grade of differentiation as two separate problems [8, 9]. In this study, we introduce a new combined NB histology classification system that formulates the stroma-grade problem as an integrated pattern analysis and classification problem. Characteristics of cytological components and different tissue patterns associated with distinct tissue subtypes are represented by constructing structural and textural features. Tissue images are then classified after training a statistical classifier over an image dataset consisting of representative samples of each tissue subtype.
Fig. 2. Flowchart of the combined NB image analysis system. E-weighted and H-weighted correspond to tissue samples with majority of the structures stained with eosin or hematoxylin, respectively.

2 Methods

Flowchart of the proposed classification approach is given in Figure 2. Image analysis starts with the segmentation step, where the image is partitioned into distinct components. Based on the relative amount of eosinophilic and basophilic structures, tissue subtypes are categorized into two groups: eosinophilic weighted (SR and D) or basophilic weighted (UD and PD). Further classification is achieved using either structural or textural features. The following subsections explain these steps in more details.

2.1 Image Acquisition and Data Set

NB tissue slides used in our study are collected in accordance with an Institutional Review Board protocol. In accordance with the Children’s Oncology Group protocol, each slide was sliced at a thickness of 5µm and embedded in paraffin and stained with H&E. The slides are digitized using a Scope XT digitizer (Aperio, San Diago, CA) at 40× magnification. The resulting digitized images are very large both in resolution, (70K × 100K pixels on average) and disk size (approximately 30GB for each slide); therefore we process the whole-slide images by dividing them into smaller image tiles that are computationally tractable. Accordingly, the flowchart given in Figure 2 is applied to every image tile in the whole-slide.

2.2 Image Segmentation

H&E-stain colors the basophilic structures consisting of nuclear and cytoplasmic regions with blue-purple hue, whereas the protein rich structures consisting of extracellular regions such as stroma and neuropil are colored with hues of pink. Red blood cells (RBCs) are stained intensely red. Accordingly, we segmented the NB images into four components: eosinophilic and basophilic structures, RBCs and background. Figure 1(e)-(h) shows the corresponding segmentations of the
sample NB images associated with different tissue subtypes. As can be seen from these samples, eosinophilic structures dominate the SR and D subtypes, whereas UD and PD subtypes exhibit majority of basophilic structures that correspond to nests of neuroblastic cells.

Background and RBC regions are segmented using simple thresholding since the intensity profile of such regions is relatively stable across the data set. The segmentation of eosinophilic and basophilic structures is achieved in the 2D space computed using the principle components analysis (PCA). The projection vectors presenting the directions of maximum variance are computed by solving the eigenvalue problem:

\[
CV = \lambda V,
\]

(1)

where \(C\) is the covariance matrix of image \(I_{N \times 3}\), in red-green-blue color space, \(N\) being the number of pixels. When we sorted with the values of \(\lambda_i\) in the descending order, the first two corresponding eigenvectors are used as the projection vectors with which the mapped data set \(Y\) is computed as:

\[
Y = (I - \bar{I})V_2.
\]

(2)

where \(\bar{I}\) is the mean of the image, \(I\), and \(V_2\) are the first two eigenvectors associated with the largest two eigenvalues.

We discarded the third component, which corresponds to the direction with the least variance; hence most likely to be associated with noise. Additionally, despite the unpredictable variations in the original RGB color space due to the varying staining conditions, the principle direction of variance stays more stable. Therefore, PCA provides a more compact representation for the subsequent estimation process.

We used the expectation maximization (EM) algorithm to identify the underlying distributions, \(p(x|\omega_i)\), of the eosinophilic and basophilic components, both modeled with Gaussian distributions. The goal is to find the unknown parameters \(\{\mu_1, \mu_2, \Sigma_1, \Sigma_2\}\), where \(\mu_1\) and \(\mu_2\) are the means, \(\Sigma_1\) and \(\Sigma_2\) are the covariance matrices associated with each distribution. EM is an iterative method consisting expectation (E) and maximization (M) steps [10]. E-step, computes the of the log likelihood with respect to the current estimates, whereas M-step maximizes the expected log likelihood. Once the underlying distributions, \(p(x|\omega_i)\), are computed, each pixel is assigned to a class based on the posterior probability \(P(\omega_i|x)\).

Figure 3 shows the segmentation on a sample image region. Based on the segmentation results, we first apply an initial categorization to separate the SR and D images, from PD and UD images such that if relative amount of eosinophilic to basophilic components exceeds a threshold, \(\tau\), we assign it to eosinophilic-weighted category (i.e., SR or D), otherwise it is assigned to basophilic-weighted category (i.e., PD or UD). We experimentally determined \(\tau\) to be 1.55, which will be discussed in the experimental results section.
Fig. 3. A sample image region is shown in (a) and its corresponding segmentation is given in (b), where background, RBCs, eosinophilic (E) and basophilic (H) components are shown in white, blue, cyan and yellow colors respectively. (c) shows the estimated distributions of H and E components using ellipses in the 2D feature space after applying PCA.

2.3 Model-based Structural Features

SR subtype mostly consists of Schwannian stroma, which is characterized by the hair-like fibrin structures that exhibit local angular similarity, whereas D subtype mainly exhibits strong presence of neuropil meshwork that consists of randomly distributed neurites that do not exhibit such prominent organization. Accordingly, we propose a set of structural features that quantify the level of organization of curvilinear structures in the tissue. Similar to our recent work on automated grading of follicular lymphoma tissue images [11] using model-based intermediate representations of cytological components, this representation allows quantifying higher-level perceptual observations.

The proposed structural features consist of line primitives that represent the ridge structures in the image intensity profile. Mathematically, ridges are a set of curves whose points are local maxima in at least one dimension. In the domain of computer vision and image processing, ridge-related representations are widely being used to represent elongated objects such as vessels and neurites [12]. Similarly, we use intensity ridges to represent the curvilinear patterns in the eosinophilic structures.

In order to detect ridge structures in the image, we construct the 2D Hessian matrix for each pixel. Hessian matrix consists of the second partial derivatives of the image:

\[
H = \begin{bmatrix}
\frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\
\frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2}
\end{bmatrix}.
\]  

where the image derivatives are obtained by convolving the image with corresponding derivatives of a Gaussian kernel with scale \(\sigma\). Let \(\lambda_1\) and \(\lambda_2\) (\(|\lambda_1| > |\lambda_2|\)) be the eigenvalues of \(H\) and \(e_1, e_2\) their corresponding eigenvectors, respectively. Then the likelihood of ridgeness, \(R\), can be defined as follows:
\[ R = \frac{\lambda_1}{\lambda_2}, \]  

(4)

Furthermore, the orientation of the ridge at a particular point is also given by the eigenvector \( e_2 \), associated with the minor eigenvalue.

Accordingly, we compute the ridge-like structures in the eosinophilic regions using \( \sigma = 2 \) for the scale parameter of the Gaussian smoothing kernel to compute the Hessian matrix, \( H \) and a threshold value \( R > 2 \). We further group the ridge structures with similar orientations and represent their connected components using line segments. Such representation using line segments provides a mathematically tractable model to quantify the local angular similarity of fibrin structures; hence allow separation between Schwannian stroma and neurophil regions.

We can exploit the structural differences to differentiate different components. For instance, Figure 4 (a) and (d) show sample images associated with SR and D tissue subtypes, respectively. Figure 4 (b) and (e) show the corresponding binary representation of the ridge structures and Figure 4 (c) and (f) show the corresponding line segment representations. As can be seen, SR subtype exhibits longer line segments and considerable angular similarity in local neighborhoods. In contrast, D subtype exhibits shorter line segments with relatively random orientations.

We extracted a set of features from the line segment representation to capture the local and global perceptual organization differences between the SR and D tissue subtypes. Let \( \Psi = \psi_i \) be the corresponding line segment representation of the ridges in the image, where \( n = ||\Psi|| \) is the total number of line segments extracted from the image. Each feature has associated with it a set of descriptors, \( \psi_i = \{ x_i, l_i, \phi_i \} \), where \( x_i = (x_i, y_i) \) is the line centroid, \( l_i \) is the line length and \( \phi_i \) is the orientation of the line with respect to the \( x \)-axis. Table 1 gives the proposed features computed based on these descriptors, where \( \mu_l \) and \( E_l \) are the mean and entropy of the line segment lengths and are computed as follows:

\[ \mu_l = \frac{1}{n} \sum_i l_i \]  

(5)

\[ E_l = -\sum_i [h_l(i) \log_2(h_l(i))] \]  

(6)

where \( h_l \) is an approximate probability mass function (pmf) of the histogram of line lengths \( l_i \). Due to dominant fibrin structures, SR tissue subtype exhibit longer and more variable length line segments when compared to neuropil tissue. \( N_l \) and \( N.A_l \) captures the number of neighboring line segments and the number of line segments with similar orientations in this local neighborhood for each line segment, \( \psi_i \) as follows:

\[ N_l = \sum_j U(\epsilon_d - ||x_i - x_j||) \]  

(7)

\[ N.A_l = \sum_j U(\epsilon_d - ||x_i - x_j||)U(\epsilon_\phi - ||\phi_i - \phi_j||) \]  

(8)
where $j = \{1, \ldots, n\}$, the spatial and angular similarity thresholds are determined experimentally as $\epsilon_d = 40$ and $\epsilon_\phi = \frac{\pi}{8}$, and $U(t - t_0)$ is the unit step function as follows:

$$U(t - t_0) = \begin{cases} 1, & t \geq t_0 \\ 0, & t < t_0 \end{cases}$$ (9)

We constructed $\mu_N$, $\mu_NA$, sample means of $N$, $NA$ and the ratio $\frac{\mu_NA}{\mu_N}$ to define three more features as listed in Table 1 to characterize the perceptual organization of curvilinear structures.

### 2.4 Textural Features

The texture of cytological components in PD and UD subtypes bear distinct patterns [9]. In PD subtype most of these cells show some differentiation with clearly recognizable neuropil content, whereas UD subtype exhibits NB cells with thin cytoplasm and “salt and pepper” appearance of chromatin, and none to very little neuropil content. In order to capture texture variations, we construct
Table 1. List of the structural features that characterizes the organization of line patterns.

<table>
<thead>
<tr>
<th>Feature Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_l$</td>
</tr>
<tr>
<td>$E_l$</td>
</tr>
<tr>
<td>$\mu_N$</td>
</tr>
<tr>
<td>$\mu_{NA}$</td>
</tr>
<tr>
<td>$\mu_{NA}$</td>
</tr>
</tbody>
</table>

Table 2. List of the textural features derived from the image.

<table>
<thead>
<tr>
<th>Feature Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_R, \sigma_R$</td>
</tr>
<tr>
<td>$E_C, H_C$</td>
</tr>
</tbody>
</table>

features from both eosinophilic and basophilic regions, separately. These features are summarized in Table 2.

A co-occurrence matrix is defined as a second-order histogram of pairwise pixel values with respect to a given spatial relationship $C(i,j)\{d,\theta\}$. After normalization, $C_{d,\theta}(i,j)$ becomes an estimate of the joint pmf of two pixels $\{X_1, X_2\}$ having co-occurring value $i$ and $j$, and constrained by the spatial relationship:

$$
\begin{bmatrix}
  x_{11} \\
  x_{12}
\end{bmatrix} =
\begin{bmatrix}
  x_{21} \\
  x_{22}
\end{bmatrix} +
\begin{bmatrix}
  d \cos(\theta) \\
  d \sin(\theta)
\end{bmatrix} (10)
$$

where $d$ is the displacement distance and $\theta$ is the orientation of the spatial pairs. In our application we set the distance as $d = 3$, the approximate scale of the salient texture patterns. Additionally, histograms associated with eight directions from images quantized with 16 gray levels are averaged as a way to represent images invariant to rotation:

$$
C_{d}(i,j) = \frac{1}{||\theta||} \sum_{\theta \in \Theta} C_{d,\theta}(i,j). (11)
$$

The entropy and homogeneity features constructed from the co-occurrence histograms are as follows:

$$
E_C = \sum_i \sum_j \hat{C}_d(i,j) \log_2 \hat{C}_d(i,j) (12)
$$

$$
H_C = \sum_i \sum_j \frac{\hat{C}_d(i,j)}{1 + (i - j)^2}. (13)
$$

2.5 Classification

Before the classification, we first applied a dimensionality reduction step using the linear discriminant analysis (LDA) method [13]. LDA maximizes the Fisher-Rao criterion defined as the ratio of the sum of inter-class distances to that
of intra-class distances associated with a set of training data. The resulting transformation to the desired subspace, where the class-seperability criteria is maximized, is obtained by solving the generalized eigenvalue problem:

$$S_b U = \lambda S_w U$$  \hspace{1cm} (14)

where $S_b$ and $S_w$ are the between- and within-class scatter matrices [10]. In the resulting feature space, the classification is achieved using the maximum a posteriori (MAP) decision rule for the classification as follows:

$$\omega = \arg \max_j p(x|\omega_j) P(\omega_j).$$  \hspace{1cm} (15)

where $P(\omega_j)$ is the prior probability of $j^{th}$ class and $p(x|\omega_j)$ is the class conditional probability.

## 3 Experimental Results

As mentioned in Section 2, the proposed combined system consists of a two-step procedure such that in the first step, the classification between the tissue samples rich in eosinophilic structures (i.e., SR and D) and the tissue samples rich in basophilic structures (i.e., PD and UD) is achieved by comparing the amount of such regions based on the outcome of the segmentation step. In Figure 5, we show the histograms of the amount of eosinophilic and basophilic structures in both categories and their estimated distributions as well as the determined classification boundary. The threshold was determined as $\tau = 1.55$ that maximizes the classification accuracy.

The training data set consists of 600 image tiles, 150 for each subclass with 1024×1024 pixel resolutions each. These are sampled from 12 digitized whole-slide NB tissue samples and represent the typical tissue structures of the underlying histopathological subtype. To evaluate the generalization of the developed system, we tested it on an independent clinical test set consisting of 34 whole-slide images. This independent testing data set is comprised of ten SR, seven D, ten PD and seven UD whole-slide NB images. On average, a digitized whole-slide image consists of approximately 7,000 image tiles of 1024×1024 pixels, providing a sufficient amount of independent samples to evaluate the outcomes of the proposed classification approach. Due to their large sizes, the whole-slide images were processed using a distributed computational infrastructure, which divides the image into smaller image tiles, processes them independently on several computational nodes on a cluster environment, and stitches the classification labels together to create the final classification label. The final classification was determined based on the majority count of image tiles associated with one of the subtypes. For evaluation purposes, we did not include any samples having high levels of heterogeneity since their evaluation requires extensive labor to generate manual ground-truth information.
The histograms and estimated distributions of $\xi_{\text{eos}}/\xi_{\text{bas}}$ ratio of SR/D and PD/UD categories.

Figure 6 (a)-(d) shows thumbnails of four whole-slide images associated with SR, D, PD and UD subtypes, respectively. Figure 6 (e)-(h) shows the corresponding “classification maps,” where colors magenta, yellow, cyan and blue correspond to image tiles classified as SR, D, PD and UD, respectively. In these resulting “classification maps”, color of each pixel represents the computerized classification label assigned to a $1024 \times 1024$ image tile located at that specific position.

The proposed system was able to correctly classify 32 out of 34 whole slide cases with an overall classification accuracy of 94.12%. The classification results of all categories are summarized in Table 3. The only two misclassified slides are one D case classified as SR and one PD case classified as UD. In fact, after a qualitative visual inspection, these two cases were found challenging since they bear similar patterns with the tissue subtype they assigned by the computerized system. Table 3 also presents the classification accuracies of the same dataset by the previously developed individual systems [8, 9]. It has been shown that the proposed combined system not only integrates the two components into one, but also outperformed the previously developed sequential computerized systems by significantly increasing the classification accuracy. Furthermore, the combined system also provides approximately 65% computational savings on average. In this proposed system, the computation time of a moderate size whole-slide image took approximately 3,000 sec. on average on a 16-node configuration.
Table 3. Classification accuracies of the proposed combined system and the previous individual systems.

<table>
<thead>
<tr>
<th></th>
<th>SR</th>
<th>D</th>
<th>PD</th>
<th>UD</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>proposed combined system</td>
<td>100%</td>
<td>85.7%</td>
<td>90.0%</td>
<td>100%</td>
<td>94.1%</td>
</tr>
<tr>
<td>previous individual systems [8, 9]</td>
<td>80.0%</td>
<td>84.6%</td>
<td>90.0%</td>
<td>90.0%</td>
<td>85.4%</td>
</tr>
</tbody>
</table>

4 Conclusions

We presented a combined computerized NB image analysis approach for identifying the degree of Schwannian stroma development and the neuroblastic grade of differentiation. Different from the previously developed NB image analysis approaches, we treated the problem as a single multi-class pattern classification problem. In addition to the conventional textural features used to discriminate PD and UD subtypes, we introduced a novel way of constructing structural features to capture the high-level perceptual organization of curvilinear structures in SR and D subtypes. After applying a linear dimensionality reduction to the constructed feature space, the class conditional distributions were composed over a training set of 600 image tiles and classification was achieved using the MAP decision rule. The proposed system was tested on an independent set of 34 whole-slide images each consisting of several thousands of image tiles and
provided a remarkable classification accuracy of 94.12%, considerably improving
the previously developed sequential systems.

5 Acknowledgments

This work is supported in part by Children’s Neuroblastoma Cancer Foundation
Young Investigator Award to Metin Gurcan, and the “2009 OSUMC Research
Day Travel Award” to Olcay Sertel.

References

2. Teao, L., Khayat, R., Qualman, S., Reaman, G., Parham, D.: The problem and
promise of central pathology review. Pediatric and Developmental Pathology 10(2)
(199-207)
ology and morphologic criteria of neuroblastic tumors: Recommendation by the int.
4. Tuzel, O., Yang, L., Meer, P., Foran, D.: Classification of hematologic malignancies
5. Tabesh, A., Teverovskiy, M., Pang, H., Kumar, V., Verbel, D., Kotsianti, A., Saïdi,
O.: Multifeature prostate cancer diagnosis and gleason grading of histological
grading of breast histology using spectral clustering with textural and architectural
aided prognosis of neuroblastoma on whole-slide images: Classification of stromal
evaluation of neuroblastoma on whole-slide histology images: Classifying grade of
tion (2001)
thological image analysis using model-based intermediate representations and
Trans. on Pattern Analysis and Machine Intelligence 22(1) (2000) 4-37