Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development

O. Sertel[a,b,*], J. Kong[a,b], H. Shimada[c], U.V. Catalyurek[a,b], J.H. Saltz[b], M.N. Gurcan[b]

[a]Department of Electrical and Computer Engineering, The Ohio State University, 2015 Neil Ave, Columbus, OH 43210, USA
[b]Department of Biomedical Informatics, The Ohio State University, 3190 Graves Hall, Columbus, OH 43210, USA
[c]Department of Pathology and Laboratory Medicine, Childrens Hospital Los Angeles and The University of Southern California Keck School of Medicine, 4650 Sunset Boulevard, Los Angeles, CA 90027, USA

ARTICLE INFO

ABSTRACT

We are developing a computer-aided prognosis system for neuroblastoma (NB), a cancer of the nervous system and one of the most malignant tumors affecting children. Histopathological examination is an important stage for further treatment planning in routine clinical diagnosis of NB. According to the International Neuroblastoma Pathology Classification (the Shimada system), NB patients are classified into favorable and unfavorable histology based on the tissue morphology. In this study, we propose an image analysis system that operates on digitized H&E stained whole-slide NB tissue samples and classifies each slide as either stroma-rich or stroma-poor based on the degree of Schwannian stromal development. Our statistical framework performs the classification based on texture features extracted using co-occurrence statistics and local binary patterns. Due to the high resolution of digitized whole-slide images, we propose a multi-resolution approach that mimics the evaluation of a pathologist such that the image analysis starts from the lowest resolution and switches to higher resolutions when necessary. We employ an offline feature selection step, which determines the most discriminative features at each resolution level during the training step. A modified k-nearest neighbor classifier is used to determine the confidence level of the classification to make the decision at a particular resolution level. The proposed approach was independently tested on 43 whole-slide samples and provided an overall classification accuracy of 88.4%.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Neuroblastoma (NB) is a cancer of nerve cell origin and it commonly affects infants and children. Based on the American Cancer Society's statistics, it is by far the most common cancer in infants and the third most common type of cancer in children [1]. Every year approximately 650 patients are diagnosed with this disease in the United States. As in most cancer types, histopathological examinations are required to characterize the histology of the tumor for further treatment planning. The World Health Organization recommends the use of the International Neuroblastoma Pathology Classification (the Shimada system) for categorization of the patients into different prognostic groups [2,3]. It is an age-linked classification system based on morphological characteristics of the tissue. Fig. 1 shows a relevant summary of this classification system as a tree diagram, where the grade of neuroblastic differentiation, the degree of Schwannian stromal development, and the mitosis and karyorrhexis index are the most salient features that contribute to the final tissue classification as favorable and unfavorable histology.

The histopathological examination guides the oncologists in making decisions on timing and therapy; hence the accuracy of the classification is important to prevent making any under or over treatment. Unfortunately, the qualitative visual examination performed by pathologists under the microscope is tedious and prone to error due to several factors. First, it is not practical to examine every region of the tissue slide under the microscope at high magnifications (e.g., 40×). For NB diagnosis, pathologists typically pick some representative regions at lower magnifications (e.g., 2×, 4×) and examine only those regions. The final decision about the entire slide is based on these sampled regions. Although this approach provides accurate decisions, it may be misleading for heterogeneous tumors. Second, the resulting diagnosis varies considerably between different readers. Experience and fatigue may cause significant inter- and intra-reader variations among pathologists. A recent study by Teot et al. shows that for NB diagnosis, this variation can be up to 20%

* Corresponding author at: Department of Electrical and Computer Engineering, The Ohio State University, 2015 Neil Ave, Columbus, OH 43210, USA.
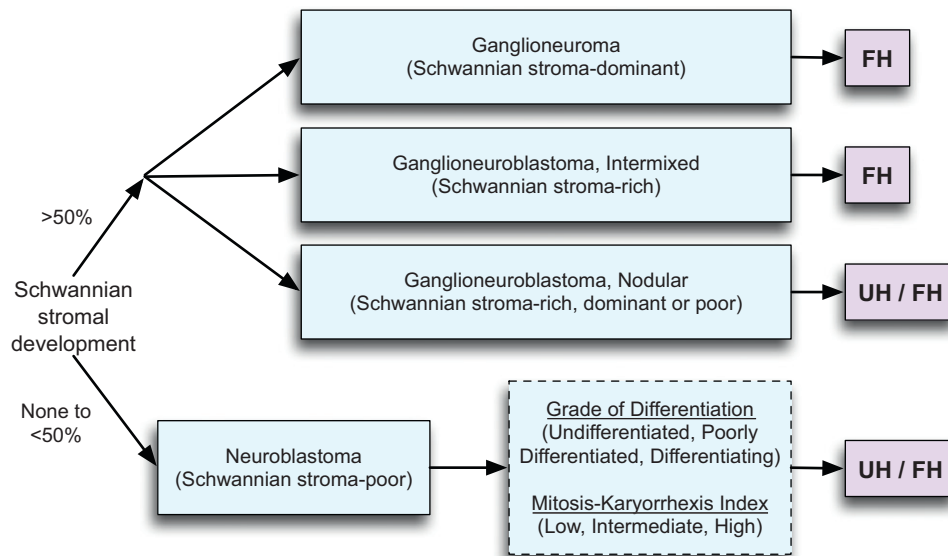   *E-mail address:* osertel@bmi.osu.edu (O. Sertel).

Fig. 1. A simplified diagram of the International Neuroblastoma Pathology Classification (the Shimada system), where UH and FH correspond to unfavorable and favorable histology, respectively.

between central and institutional reviewers [4]. There is no specific study that relates to such variations in classification of the degree of Schwannian stromal development in NB diagnosis. However, as shown in Fig. 1, this analysis is the first, hence, a critical step in NB prognosis.

To address these drawbacks, we are developing a computer-aided diagnosis (CAD) system for NB. The use of computers to assist physicians in their evaluations of medical images is not a new study. There are several commercially available CAD systems that have been proven to improve the clinical diagnosis of radiology images for several modalities such as mamagraphy and computed tomography (CT) [5,6]. However, research on the development of such similar systems for whole-slide histopathological image analysis is relatively new. This is mostly due to the challenges in both the acquisition and the processing of histopathological images that are much larger in size, as opposed to radiology images. Parallel to the recent developments in whole-slide digital scanners, research studies on quantitative histopathological image analysis have been accelerated. Providing computational tools to extract measurable features, histopathological image analysis systems help extracting more objective and more accurate diagnostic clues that might not be easily observed by qualitative analysis performed by pathologists.

Research efforts on histopathological image analysis can be categorized into two main groups based on their objectives: (1) content-based image retrieval (CBIR) and (2) CAD systems. CBIR systems aim to retrieve clinical cases from a database, containing previously diagnosed representative cases similar to a query image; hence pathologists could make use of the established knowledge in their decision making process [7,8]. On the other hand, CAD systems directly focus on making the diagnostic classifications of the tissues (e.g., malignant or benign; grade of differentiation) [9]. However, due to variations of tissue structures and different prognosis procedures in different disease types, it is impractical to develop a universal system that would work for all disease types, even if they show similar characteristics. Recently, Petushi et al. proposed an image analysis system to determine the grade of differentiation for breast cancer [10]. Their study provides automatic segmentation and classification of distinct cell nuclei that are used for identification of the grade of differentiation to indicate the degree of malignancy. Similar studies have been

conducted to automate the Gleason grading system for prostate cancer. Khouzani et al. proposed an image analysis approach using a multi-wavelet-based approach to characterize the texture of the samples associated with different grades [11]. In addition to the textural features, Doyle et al. introduced the use of architectural features based on the spatial organization of cells, as well as their morphologies [12]. Tabesh et al. incorporated color, texture, and morphometric image features at the global and histological object levels for prostate tissue grading [13].

Similarly, several research studies have been conducted on quantitative analysis of NB. Gurcan et al. proposed a cell segmentation method from H&E stained pathology slides using morphological reconstruction followed by hysteresis thresholding [14]. Most recently, Kong et al. proposed a classification approach using texture and color information to determine the grade of differentiation for NB diagnosis [15,16]. Both studies showed promising results for developing an automated framework to be used in clinical practice as assistance. However, to the best of our knowledge, there has not been any research conducted to analyze Schwannian stromal development for NB diagnosis.

In this study, our goal is to develop a CAD system to determine the degree of Schwannian stromal development as either stroma-rich or stroma-poor from digitized whole-slide NB slides. In Fig. 2 we show sample NB tissue images cropped at 40× magnification. Figs. 2(a) and (b) correspond to stroma-rich tissue samples that can be characterized by an extensive growth of Schwannian and other supporting elements. On the contrary, Figs. 2(c) and (d) demonstrate stroma-poor tissue samples that can be characterized by a diffuse growth of neuroblastic cells with various degrees of differentiation randomly distributed by thin septa of fibrovascular tissue and neurphil meshwork.

Using sophisticated computer vision and pattern recognition techniques, we introduced a multi-resolution image analysis approach to identify image regions associated with different histopathological components (i.e., stroma-rich and stroma-poor). The proposed multi-resolution approach mimics the way pathologists examine the tissue slides under the microscope such that the image analysis starts from the lowest resolution, which corresponds to the lower magnification levels in a microscope and uses
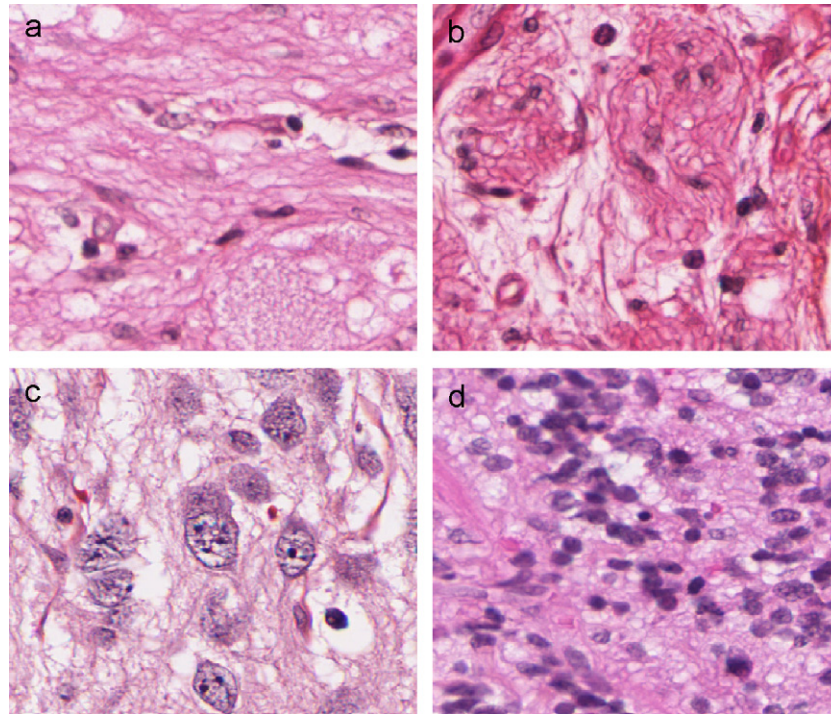
Fig. 2. Example images of (a,b) stroma-rich and (c,d) stroma-poor tissue.

the higher resolution representations for the regions where the decision for the classification requires more detailed information. We proposed a texture-based approach to differentiate Schwannian stroma-rich tissue from other cytological structures. We employed the rotation-invariant co-occurrence statistics and local binary patterns (LBP) to characterize the stroma septa with different organizations. Using representative samples, we constructed a training dataset and extracted textural features. We further employed an automated feature selection step in which the most discriminating subset of the features are determined at each resolution level to improve the classification performance. Finally, the classification has been performed using a statistical classifier.

1.1. Image dataset

In our study, we used 45 whole-slide tissue samples collected from Nationwide Children's Hospital. Tissue slides were obtained retrospectively according to an Institutional Review Board (IRB) protocol. Each slide was embedded in paraffin and was cut at a thickness of 5 μm according to commonly used Children's Oncology Group protocols. After being stained by hematoxylin and eosin (H&E), each tissue slice was fixed on a slide and was digitized using a ScanScope T2 digitizer (Aperio, San Diego, CA) at $40\times$ magnification. The digitized whole-slide tissue images typically have a spatial resolution up to $100k \times 120k$ with a disk size up to 40 GB. Therefore, at the time of processing, the whole-slide images are decomposed into smaller non-overlapping image tiles. The tiling of the whole-slide images not only made it practical to process the whole-slide images, but also allowed leveraging the parallelism in processing each image tile independently. Experimentally, we determined the image tile size as 896×896 in pixels. The average resolution of tissue slides used in our study is approximately $71,623 \times 100,348$ in pixels; with an average disk size of 20 ± 8 GB; hence the average number of image tiles in a whole-slide image is approximately 8900, which still requires significant computation time.

Table 1
Distribution of neuroblasoma cases based on MKI and the grade of differentiation

MKI	UD	PD	D
Low	3	10	6
Intermediate	3	3	1
High	2	3	1

UD, PD, and D correspond to undifferentiated, poorly differentiated and differentiating subtypes, respectively.

One representative whole-slide sample for each subtype (i.e., stroma-rich and stroma-poor) has been used to generate the training image tiles and the remaining 43 were used for whole-slide independent testing. Thirty-two of 43 whole-slide samples are associated with stroma-poor and the rest are associated with stroma-rich subtypes, as determined by an expert pathologist. Five of stroma-rich cases correspond to Ganglioneuroblastoma and six to Ganglioneuroma. The remaining 32 stroma-poor cases correspond to NB. Their morphological characteristics in terms of mitotic-karyorrhectic index (MKI) and the grade of differentiation, according to the International Neuroblastoma Pathology Classification (the Shimada system), are summarized in Table 1.

1.2. Computational infrastructure

The image analysis routines were implemented using MATLAB (version 7.1.0.246, Natick, MA) and the experimental studies have been carried out on a 64-node PC cluster owned by the Department of Biomedical Informatics, The Ohio State University. Each computation node on the cluster is equipped with dual 2.4 GHz Opteron 250 processors and 8 GB of RAM. This parallel system uses a software developed in house, which distributes the processing of image tiles to each node, applies the required MATLAB routines on each image tile locally, collects the classification outputs and stitches them together to create the final classification map. Fig. 3 shows the
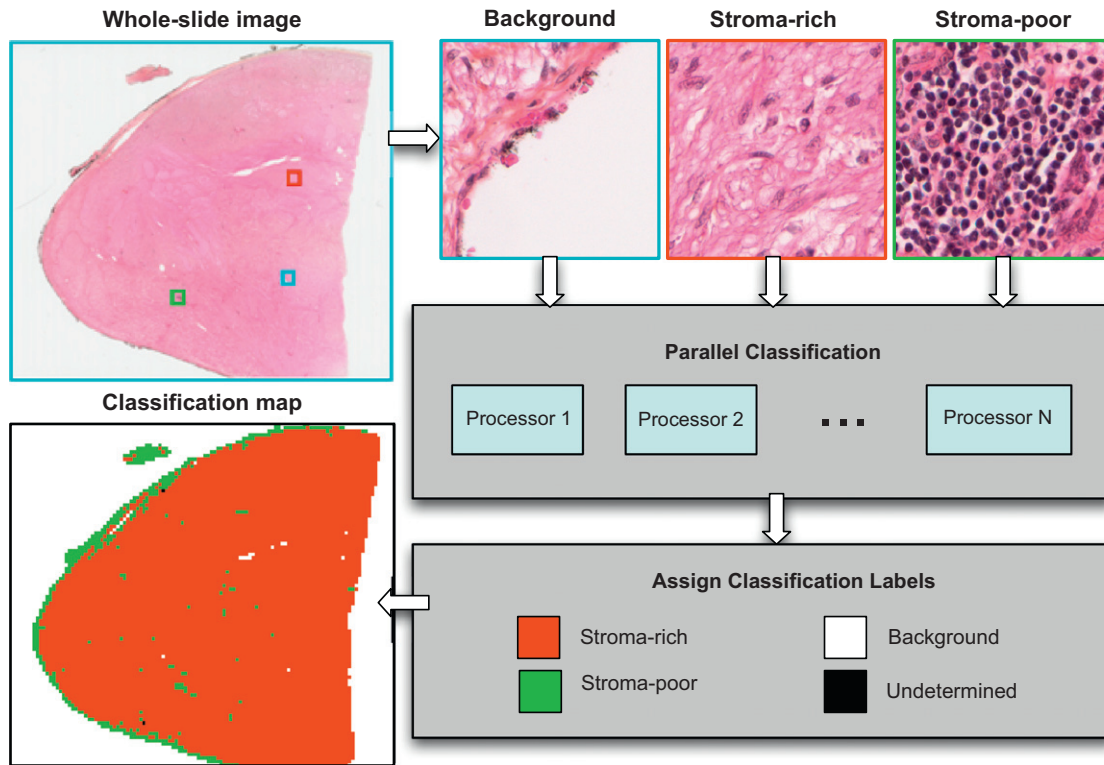
Fig. 3. Computational infrastructure used to process the whole-slide images in parallel.

computational infrastructure used to analyze the whole-slide images in parallel.

An additional extension of this software is a grid-based infrastructure where image analysis modules (i.e., MATLAB or C/C + + files) could be uploaded to the system by multiple developers and pipelined to be applied to the available whole-slide images stored in the common repository using a grid interface [17].

2. Classification of stromal development using texture information

We formulate the classification of digitized microscopy images of NB as a statistical pattern recognition problem [18]. The flowchart of the proposed whole-slide image analysis system is given in Fig. 4. Our system decomposes each whole-slide image into smaller image tiles and each image is processed independently. Given an image tile, we first check whether it contains sufficient amount of tissue component. To test this, we count the number of white pixels. We considered any pixel with intensity level greater than 200 in all the red, green, and blue channels as a white pixel. If the number of these pixels is greater than 80% of the total number of pixels in the particular image tile, then the image tile is assigned a label for background without subsequent processing.

For the classification of stromal development, we used a multi-resolution strategy that provides remarkable computational savings. The proposed multi-resolution approach starts processing the image tiles at the lowest resolution with the least computational burden and switches to higher resolutions only when necessary. At each resolution level, we extracted a set of textural features to discriminate the stroma-rich and stroma-poor tissue subtypes. The classification is performed using a modified k-nearest neighbor classifier, which provides a confidence measure for the classification at each level. Finally, once the classifier is confident enough at a particular resolution level, the system assigns a classification label indicating

stroma-poor or stroma-rich subtype to the image tile. The resulting classification labels from all image tiles of the same slide form the final classification map. As indicated by the International Neuroblastoma Pathology Classification (the Shimada system), a whole-slide image is considered stroma-rich if image tiles classified as stroma-rich is more than 50%. The following subsections explain each step of the proposed approach in more detail.

2.1. Gaussian pyramids

For implementation of the multi-resolution approach, we used the Gaussian pyramid approach, which provides the low-resolution representations of a given image [19]. In the Gaussian pyramid representation, the original image, I_0, becomes the bottom or (in other words) the zero level of the pyramid. Other representations through the upper levels in the pyramid consist of low-pass filtered and down-sampled versions of the preceding levels, respectively. Eq. (1) gives the computation of each level in the pyramid:

$$I_l(i,j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(n,m)I_{l-1}(2i+m, 2j+n), \tag{1}$$

where $w(n,m)$ the weighting function can be written as follows:

$$w(n,m) = w(n)w(m), \tag{2}$$

$$w = [0.25 - 0.5\alpha, \ 0.25, \ \alpha, \ 0.25, \ 0.25 - 0.5\alpha]. \tag{3}$$

The pixel values are computed as a weighted average of values within a 5-by-5 window in the preceding level. It should be noted that due to the down-sampling process, the resolution of the image is being reduced by half in two dimensions each time we compute the new level. In our study, we computed a four-level representation (i.e., I_0, I_1, I_2 and I_3) of each image tile, and the parameter α is chosen to be $\alpha = 0.4$ to create a Gaussian-like smoothing kernel.
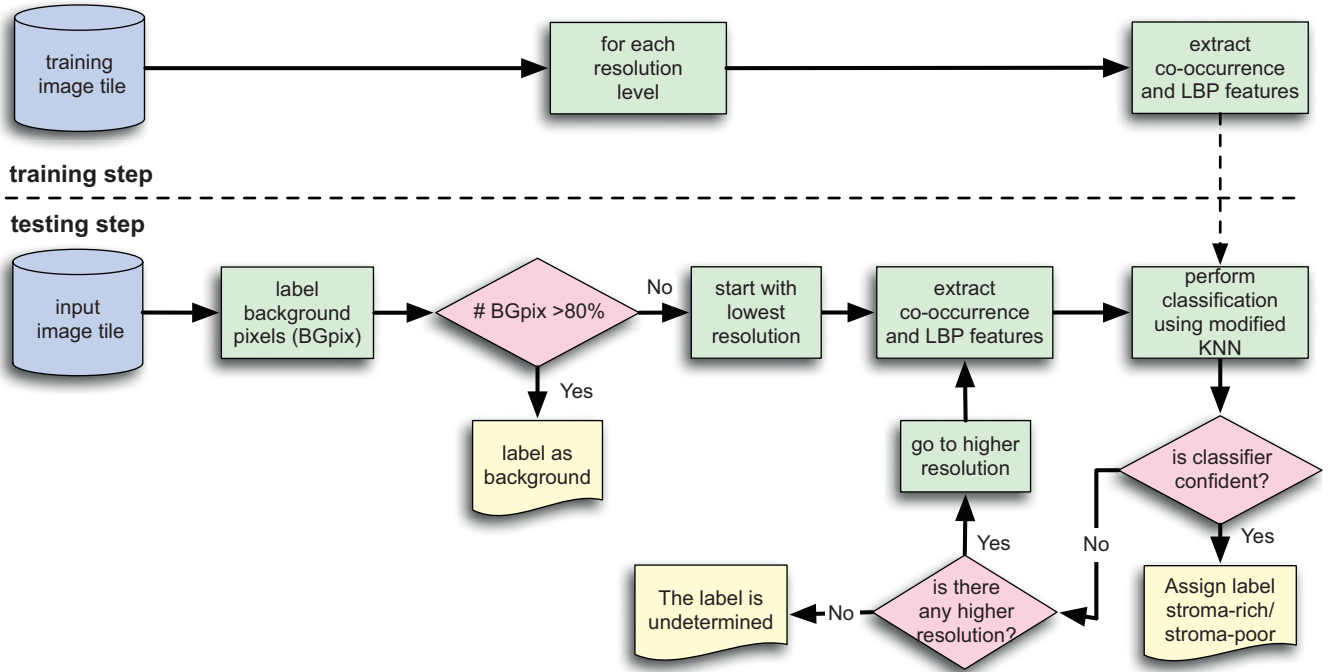
Fig. 4. Flowchart of the whole-slide image analysis system for NB prognosis. KNN and LBP correspond to k-nearest neighbor and local binary patterns, respectively.

2.2. La*b* color conversion

To better represent the texture and color information independently, in addition to the red-green-blue (RGB) color space, we used the La*b* color space developed by the International Commission on Illumination (CIE). The La*b* is a perceptually uniform color space meaning a change of the same amount in a color value should produce the same amount of perceptual difference of visual importance [20]. L channel corresponds to illumination and, a* and b* channels correspond to the color opponent dimensions. It is derived from the CIE XYZ color space, which is based on direct measurements of human visual perception.

Our goal using the La*b* color space was not to compensate staining dissimilarities. We aimed to separate color and illumination information and obtain a perceptually uniform color space so that comparing two color vectors using the Euclidean distance could be more appropriate. Furthermore, the variability in staining is minimal in our case since the tissue samples are acquired according to the commonly accepted Children's Oncology Group protocols, and the extracted texture features can compensate these differences by convention. The LBP features are invariant to any local or global illumination change, which is also valid for color since we apply the LBP operator on each color channel.

2.3. Texture features

The texture of the stroma septa given in Figs. 2(a) and (b) is quite different than the neurophil meshwork seen in Figs. 2(c) and (d). The hair-like fibrin structures exhibit patterns that are locally organized in particular directions, where the neurophil meshwork randomly distributed between neuroblastic cells (typical rosetta pattern associated with differentiating subtype in Fig. 2(c) and differentiated cells in Fig. 2(d)) do not exhibit any directional organization. To capture texture differences, we constructed a set of features for each image tile using second order statistics [21] and LBP [22]. These features are extracted from each channel in the La*b* and the RGB color spaces.

Table 2
Features used for automated classification of image tiles as stroma-rich and stroma-poor

Features	Description
Contrast	The variation between intensity levels
Correlation	The measure of correlation between a pixel and its neighborhood
Energy	The measure of uniformity
Entropy	The measure of randomness
Homogeneity	The degree of homogeneity
Dissimilarity	The measure of dissimilarity of the intensity values
LBP	Captures the proportion of micro-features

A summary of the set of features used in our system is given in Table 2. The features given in the first six rows of Table 2, also known as Haralick features, were extracted using co-occurrence histograms [23].

A co-occurrence histogram is simply the joint frequency of the intensity levels of a pair of pixels with a given spatial relationship. This relationship is often indicated using a displacement vector $d(dx, dy)$, where dx and dy are the offsets in horizontal and vertical dimensions of the image, respectively. Using this spatial relationship, the computation of a co-occurrence histogram, $P_d(i,j)$ of an image tile, I is as follows:

$$P_d(i,j) = |I(x,y) = i, I(x + dx, y + dy) = j|, \tag{4}$$

where, $x, y \in N \times M$, $N \times M$ is the image resolution and $|\cdot|$ denotes the cardinality of a set. For the construction of the co-occurrence histogram, we used a distance of one pixel in eight directions and computed their mean, P_{mean} to obtain a rotation-invariant representation as follows:

$$P_{mean} = \frac{1}{8} \sum_{dx=-1}^{1} \sum_{dy=-1}^{1} P_d(dx, dy) \quad \text{and} \quad |dx| + |dy| \neq 0. \tag{5}$$

LBP is a feature extraction method of capturing the proportion of the micro-patterns in the texture of the image. The conventional
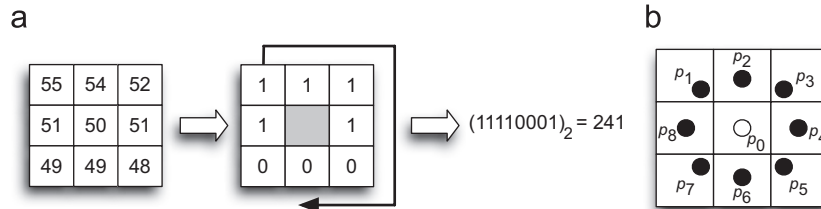
Fig. 5. (a) The conventional LBP operator; (b) circular pattern used to compute rotation invariant uniform patterns.

LBP operator is applied to every pixel p_c by examining the eight neighbors to see if their intensity is greater than that of p_c as shown in the following equation:

$$LBP(p_c) = \sum_{i=0}^{N-1} 2^i \tau(p_c - p_i), \tag{6}$$

where N is the number of neighbors with consistent order for each pixel, p_i and $\tau(x)$ is

$$\tau(x) = \begin{cases} 1 & \text{if } x \geqslant 0, \\ 0 & x < 0. \end{cases} \tag{7}$$

The results from the eight neighbors are used to construct an eight-digit binary number, $b_1 b_2 \ldots b_8$, where $b_i = 0$, if the intensity of the ith neighbor is less than, or equal to, that of p_c, otherwise $b_i = 1$. The histogram of these values is used to represent the texture of the image. The computation of the LBP feature for a given pixel is illustrated in Fig. 5(a).

Ojala et al. introduced rotation-invariant LBP features by using circularly sampled neighborhood pixels [22]. Instead of using the rectangular kernel, they propose the use of a circular kernel to increase the degree of freedom for rotational invariance, as shown in Fig. 5(b). The circular kernel around each pixel is computed using bilinear interpolation. The authors observed that when the image is rotated, the binary patterns will only be circularly shifted through the center pixel. Based on this observation, they introduced the uniform patterns. An LBP is called uniform if it contains, at most, two bitwise transitions from 0 to 1 or vice versa in its circular string. These uniform patterns are referred to as bright spot, flat area, dark spot, and edges of varying positive and negative curvature. The authors experimentally showed that the uniform patterns account for almost 90% of all patterns in texture images. Following this powerful representation that is invariant to rotation and any local or global intensity change, we computed the LBP features using a radius of one pixel to construct the circular pattern with eight samples around each pixel. The texture features are constructed for each image tile at each resolution level to perform classification.

2.4. Feature selection

As in many other pattern recognition applications, some of the features extracted from each image tile might be redundant; without contributing much to the discrimination power. In fact, if too many features are used in a classification system, it can degrade the overall performance, i.e., the trained classifier may provide poor generalization on the independent testing set. This is also known as the curse of the dimensionality [18,24]. Furthermore, the construction of features requires certain computation. Thus, we employed an automated feature selection step not only to improve accuracy, but also to decrease the computational complexity of the proposed system.

Let Y be the original set of features with n dimensions and d representing the desired number of selected features X, where $X \subseteq Y$. Let the feature selection criterion function over the set X be $J(X)$ to

Table 3
Features selected at each resolution level by the SFFS algorithm that minimizes the leave-one-out classification error on the training set

Resolution level	Selected features
3 (112×112)	LBP (L), Energy (R), Entropy (a*)
2 (224×224)	Entropy (a*), Entropy (b*), LBP (L)
1 (448×448)	Entropy (b*), LBP (L), Homogeneity (L)
0 (896×896)	Homogeneity (L), LBP (L), Entropy (b*), Energy (a*), Energy (b*)

be maximized. The problem of feature selection is to find a subset $X \subseteq Y$, where $|X| = d$ using

$$J(X) = \max_{Z \subseteq Y, |Z| = d} J(Z). \tag{8}$$

The number of possible subsets increases exponentially in the case of an exhaustive search that requires examining all possible d-subsets of the feature set Y, making it impractical for moderate values of n. Therefore, several techniques have been introduced to find the optimal subset without searching the entire solution space. Among several feature selection algorithms, we employed the sequential floating forward selection (SFFS) method, which has been shown to be superior to other sequential procedures [25,26]. As the feature selection criterion, we used leave-one-out cross-validation over the training set to compute the classification accuracy. This automated feature selection step was applied in an offline manner at the training step for each resolution level and only the selected feature sets were extracted at the testing stage.

In Table 3 we report the selected features at each resolution level. The letters in the parenthesis represent the color channel from which the corresponding features have been constructed (e.g., entropy (b*) corresponds to entropy feature extracted from the b* channel of the image represented in the La*b* color space). As shown in Table 3, the LBP feature was selected at each resolution level, thus, it can be referred to as the most discriminating feature in our feature set. It should also be noted that there is a smooth transition between the subset of features selected at each resolution level. When we go to higher resolutions that contain more detailed information, only some of the features were replaced with more discriminating features.

Fig. 6 shows the scatter plot of the training samples in the 2D feature spaces associated with different resolution levels. Subfigures 3–0 correspond to the scatter plots of the training data from the lowest to the highest image resolutions, respectively. The 2D feature space plots given in Fig. 6 have been obtained by applying the principal component analysis (PCA) method [24], where the first two principal components were plotted. As can be seen from these plots, when we switch to higher resolutions, the between class discrimination increases. The discrimination is still satisfactory at lower resolutions, where a large number of cases can be classified correctly with less computational cost. To quantify this observation, we computed the classification accuracies at each resolution level using the KNN
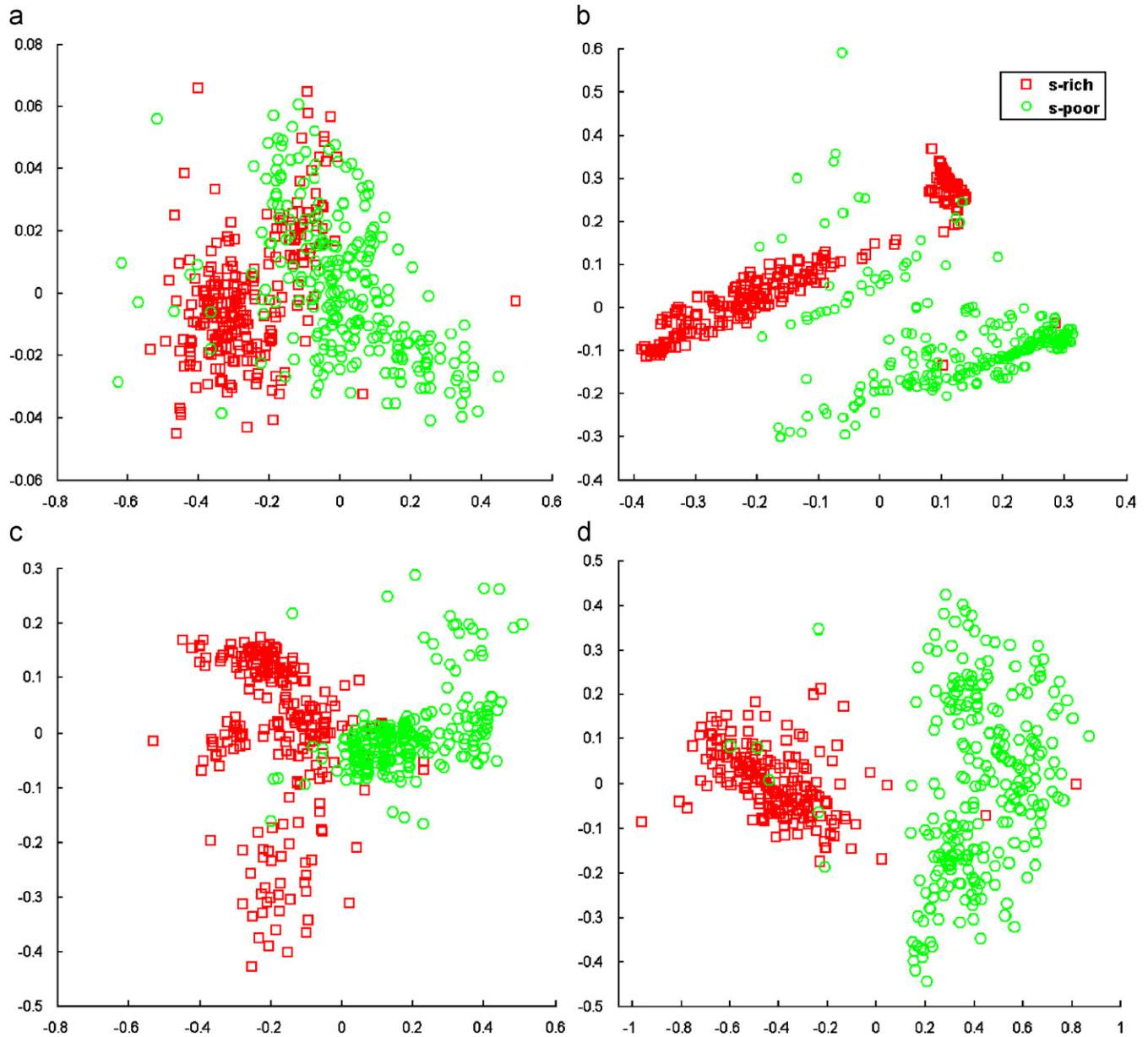
Fig. 6. The 2D scatter plot of feature spaces associated with different resolution levels. The discrimination of the class samples increases proportionally with the increasing resolutions 3–0, respectively. (a) Resolution level 3, (b) resolution level 2, (c) resolution level 1 and (d) resolution level 0.

Table 4
Classification accuracies among the training set using leave-one-out and 10-fold cross-validation at different resolutions

Resolution level	Leave-one-out (%)	10-fold (%)
3 (112×112)	92.0	86.0
2 (224×224)	95.4	94.4
1 (448×448)	96.2	96.6
0 (896×896)	98.4	98.0

classifier. Table 4 shows leave-one-out and 10-fold cross-validations results among the training set at each resolution level. The improvement in classification accuracy indicates that the separability between the two classes increases with the increasing resolution level.

2.5. Classification

Using the proposed multi-resolution approach, NB images are processed starting at the lowest resolution level. The system switches to higher resolution levels only if the classification results are not sufficient at the particular resolution level. A modified KNN classifier was employed to determine the confidence level at the particular resolution level. Typically, the KNN classifier first determines the k closest training samples of the testing sample in the feature space using the Euclidean distance. Then, the label of the class that has the majority of the samples among these nearest neighbors is assigned as the label of the test sample. We modified this rule by introducing a confidence level for each resolution level. Thus, for the lowest resolution level, instead of seeking the majority, the classifier requires at least 90% of the samples belong to a certain class. Being so restrictive allows only the most representative test samples to be classified

at lower resolutions where others needed to be examined at higher resolutions. The majority requirement is decreased gradually each time the resolution level is increased. Hence, the multi-resolution approach is quite strict on making the decision at the low resolution levels as opposed to higher resolutions. This helps to improve computational performance remarkably without losing any discrimination power.

Below, we provide the pseudo-code for the proposed classification scheme based on KNN, where x corresponds to a sample feature vector to be classified, k is the number of neighbor samples, ω_i is the ith class and R_j is the jth resolution level. $CL(R_j)$ corresponds to the confidence level at resolution level j. In our experiments, we set $\alpha = 0.1$, $CL(R_3) = 0.9$.

Algorithm 1. Modified KNN

$j = 3$
for a testing sample x
while $j \geqslant 0$ **do**
    find k closest samples using Euclidean distance
    compute the probability, $P_{\omega_i}(\omega_i|x) = \frac{k_i}{k}$
    if $(P_{\omega_i}(\omega_i|x)) \geqslant CL(R_j)$ **then**
        $\omega_x = \omega_i$
    else
        $j = j - 1$
        $CL(R_j) = CL(R_{j+1}) - \alpha$
    end if
end while

3. Experimental results

Our training dataset consists of 500 image tiles cropped from two whole-slide samples associated with a stroma-rich and a stroma-poor subtype, respectively. During the training step, we first applied the multi-resolution decomposition to the training images using the Gaussian pyramid approach. This is followed by color space conversion and texture feature construction at each resolution level, as shown in Fig. 7. Finally, we applied feature selection and stored the corresponding feature space information associated with each resolution level.

We tested the proposed computerized classification system on 43 different whole-slide NB images obtained from Nationwide Children's Hospital. Thirty-two of these slides correspond to stroma-poor subtype and the remaining 11 correspond to stroma-rich subtype. Using the aforementioned computational infrastructure, we performed the automated classification and obtained the corresponding classification maps. Fig. 8 shows classification results obtained from two sample whole-slide NB images. Figs. 8(a) and (d) are the original slides with resolutions 88,438 × 109,065 and 40,604 × 67,900 in pixels associated with stroma-rich and stroma-poor classes, respectively. Figs. 8(b) and (e) are the corresponding classification maps determined by the proposed computerized system. Each pixel in the classification maps corresponds to the class label of an image tile of size 896 × 896 in pixels. In these classification maps red, green, and white colors correspond to the stroma-rich tissue, the stroma-poor tissue and background regions, respectively. In cases where the computerized system is not confident at assigning a label to a given image tile, it is assigned black color indicating that it will not be used in final decision mechanism. If a relatively large portion (i.e., 5%) of the slide is unclassified, no conclusive decision can be made by the computerized system. However, this has never occurred in our testing database. Figs. 8(c) and (f) show the number of image tiles classified at a particular resolution level in a log-scale graph. The first bar on the left correspond to the background tiles, and the rest from left to right correspond to resolution levels from lowest to highest, respectively.

For the first case given in Fig. 8(a), 94.28% of the image tiles have been assigned to the stroma-rich subtype; therefore, the whole-slide image has been classified as stroma-poor since the threshold defined by the International Neuroblastoma Pathology Classification (the Shimada system) is 50%. For the second case given in Fig. 8(d), the stroma-rich tiles only correspond to 7.81% of the total number of image tiles. Therefore, the whole-slide image has been classified as stroma-poor. Theautomated classification of the slides is quite accurate, with most of the misclassified tiles lying on the boundaries, as shown in Figs. 8(b) and (e).

Figs. 8(c) and (f) show that most of the image tiles have been classified at the lower resolution levels, leading to remarkable computational savings. The computation time associated with different resolution levels is given in Table 5. Based on these, the
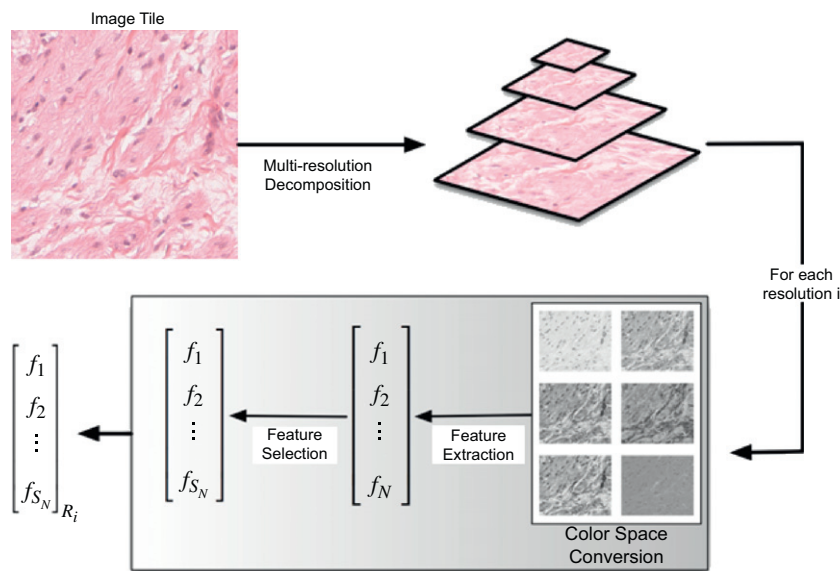

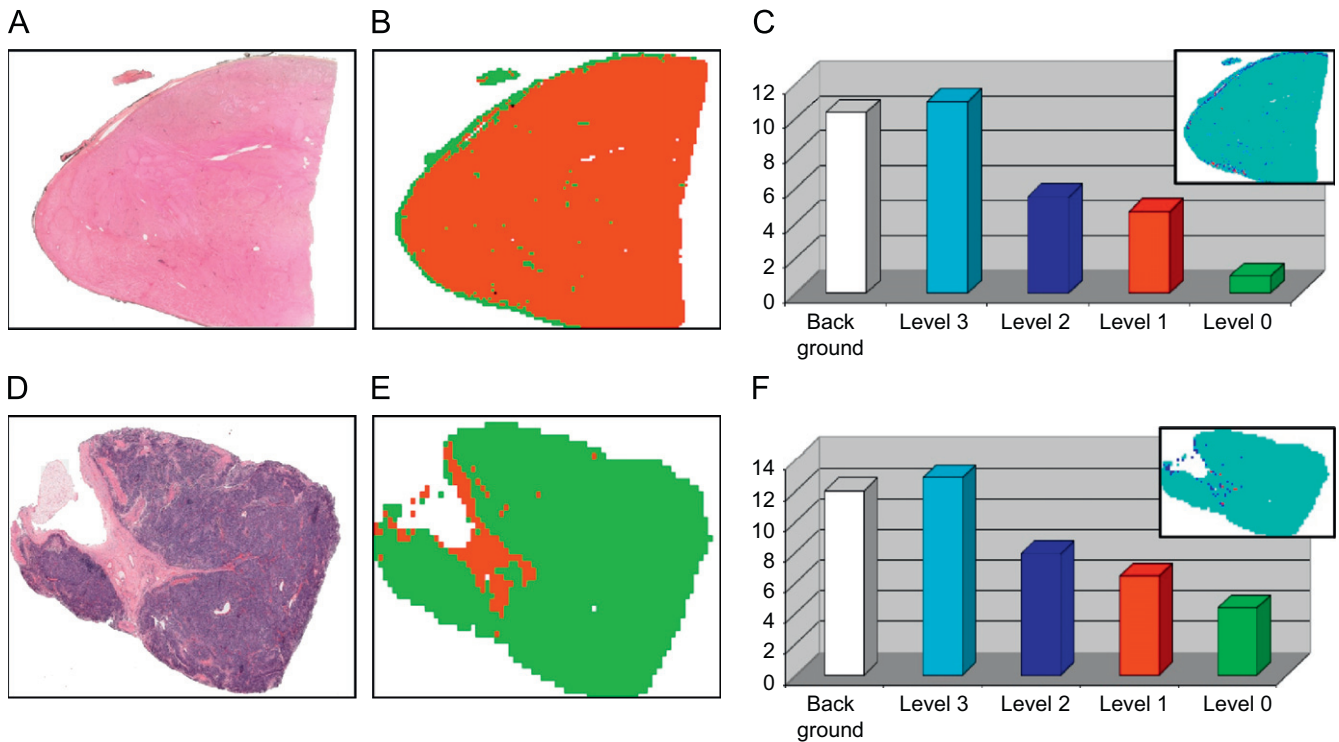**Fig. 7.** Flowchart of the training process.

Fig. 8. Sample classification results after processing a whole-slide NB image. (a) and (d) are the H&E stained NB slides associated with stroma-rich and stroma-poor by an expert pathologist. (b) and (e) are the classification maps identified by the computerized system where the red color corresponds to stroma-rich regions and the green color corresponds to stroma-poor regions. (c) and (f) are the corresponding decision level statistics that show in log-scale the number of image tiles classified at a certain resolution level. In the resolution level map on the upper right, cyan color represents the lowest resolution and green color represents the highest resolution, respectively.

Table 5
Required times for the classification of an image tile at different resolution levels

Resolution level	Time required (in ms)
3 (112×112)	4.48
2 (224×224)	9.20
1 (448×448)	15.06
0 (896×896)	32.77

proposed multi-resolution approach reduced the execution time by 85% on average when compared to the single resolution approach.

The classification results of 43 different whole-slide images are given in Figs. 9 and 10. Since each whole-slide image has a different resolution, they have a different number of image tiles indicated by the y-axis. The number of stroma-rich, stroma-poor, background and undetermined image tiles is represented by different colors in each bar. As shown in Fig. 9, the majority of the image tiles were assigned to the stroma-rich subtype; hence, all of the 11 stroma-rich cases were classified correctly with a 100% perfect accuracy. Out of 32 stroma-poor cases, 27 of them were classified correctly leading to a classification accuracy of 84.38%, as given in Fig. 10. The first five cases reported in Fig. 10 were misclassified as stroma-rich. Four of these misclassified stroma-poor cases correspond to NB, differentiating subtype with low MKI indices. This indicates that the proposed approach does not perform very well on images associated with NB, differentiating subtype. However, the rest of the samples were correctly classified and the computerized system achieved the best overall classification accuracy of 88.38% which is quite promising for the clinical applications.

4. Conclusions

The evaluation of the entire tissue sample is critically important for NB prognosis where the stromal development is determined as the ratio of stroma-poor and stroma-rich regions. For practical reasons, pathologists typically examine representative regions in each slide before they come up with a prognostic decision (i.e., sampling). However, sampling may lead to prognostic errors, particularly in tumors that show heterogeneity. Since the computer analyzes the whole slide, it can potentially warn the pathologist in case of an inconsistency in recognizing stroma-poor or stroma-rich regions from pathological slides. This could have an important effect on the prognostic decision and hence could help reducing the intra- and inter-reader variability.

Independent tests show that the proposed system can determine the stroma-rich regions with 100% accuracy and stroma-poor regions with 84.38% accuracy, on a 43 slide dataset. It should be noted that the automated computer-aided system should never be considered a replacement for pathologists. Instead, it should only be used as assistance to the decision mechanism. In case of a disagreement between the computerized system and the human reader, the final decision is that of the human reader.

Digitized whole-slide histopathological images have very large resolution; hence, processing them in a reasonable amount of time requires novel computational methods. In this work, we investigated a multi-resolution framework, which provides computational savings up to 85% compared to a single-resolution approach developed earlier. This computational saving is useful toward developing a computerized system that can be used in clinical practice. Currently, the classification of an average sized whole-slide image takes around 7–8 min on eight computer nodes. However, we are
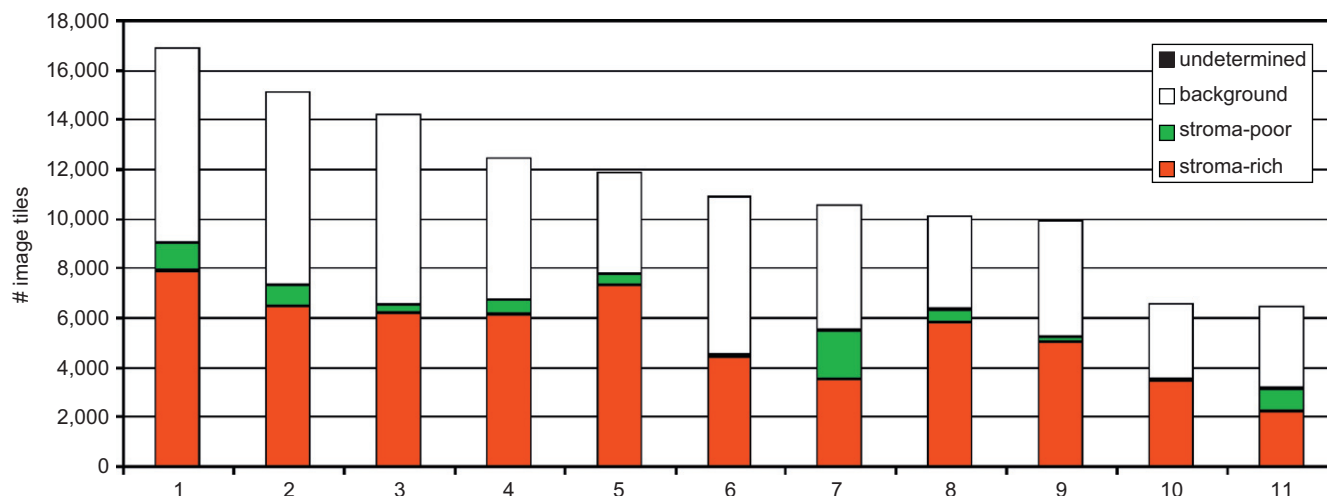
Fig. 9. The distribution of the image tiles over the 11 stroma-rich, whole-slide images. The red, green, and white colors represent the number of image tiles classified as stroma-rich, stroma-poor, and background tiles, respectively. The black color represents the number of image tiles that were not able to be classified by the computerized system.
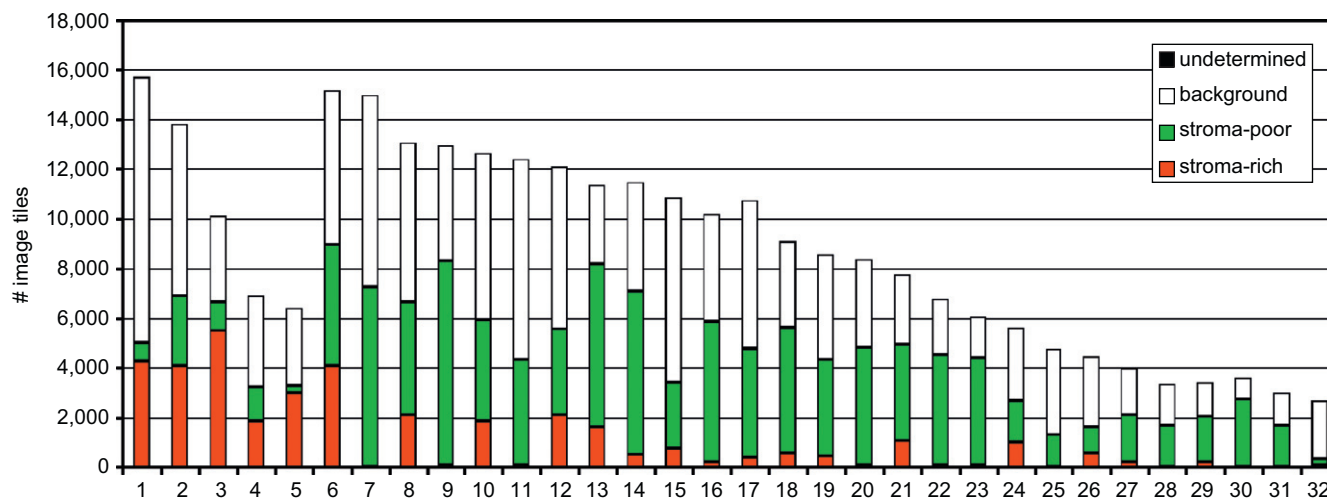

Fig. 10. The distribution of the image tiles over the 32 stroma-poor, whole-slide images. The red, green, and white colors represent the number of image tiles classified as stroma-rich, stroma-poor, and background tiles, respectively. The black color represents the number of image tiles that were not able to be classified by the computerized system.

conducting further research on computational aspects of histopathological image analysis such as the use of graphics processing units (GPUs) and IBM™ cell processors. A developed system with GPUs reduces the processing time for an analysis of an average sized whole-slide image to less than 30 s [27,28].

In our future work, we will concentrate on improving the accuracy of the system to enhance the performance on images associated with NB, differentiating subtype. Therefore, we will incorporate additional feature construction methods such as the use of morphological and topological characteristics of different types of cells in a graph theoretical framework. We will also study linear and non-linear dimensionality reduction methods to improve classification performance.

Acknowledgments

References

[1] American Cancer Society ⟨http://www.cancer.org⟩ (2007).

[2] H. Shimada, I.M. Ambros, L.P. Dehner, J. Hata, V.V. Joshi, B. Roald, D.O. Stram, R.B. Gerbing, J.N. Lukens, K.K. Matthay, R.P. Gastlebery, The international neuroblastoma pathology classification (the Shimada system), Cancer 86 (2) (1999) 364–372.

[3] H. Shimada, I.M. Ambros, L.P. Dehner, J. Hata, V.V. Joshi, B. Roald, Terminology and morphologic criteria of neuroblastic tumors: recommendation by the international neuroblastoma pathology committee, Cancer 86 (2) (1999) 349–363.

[4] L.A. Teao, R.S.A. Khayat, S. Qualman, G. Reaman, D. Parham, The problem and promise of central pathology review: development of a standardized procedure for the children's oncology group, Pediatr. Dev. Pathol. 10 (2007) 199–207.

[5] L. Burhenne, S. Wood, C. D'Orsi, S. Feig, D. Kopans, K. O'Shaughnessy, E. Sickles, L. Tabar, C. Vyborny, R. Castellino, Potential contribution of computer-aided detection to the sensitivity of screening mammography, Radiology 215 (2000) 554–562.

[6] M.N. Gurcan, B. Sahiner, N. Petrick, H. Chan, E. Kazerooni, P. Cascade, L. Hadjiiski, Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system, Med. Phys. 29 (11) (2002) 2552–2558.

[7] D. Comaniciu, P. Meer, D.J. Foran, Image-guided decision support system for pathology, Mach. Vision Appl. 11 (1999) 213–224.

[8] W. Chen, P. Meer, B. Georgescu, W. He, L.A. Goodell, D.J. Foran, Image mining for investigative pathology using optimized feature extraction and data fusion, Comput. Methods Programs Biomedicine 79 (2005) 59–72.

[9] S. Cross, J. Bury, T. Stephenson, R. Harrison, Image analysis of low magnification images of fine needle aspirates of the breast produces useful discrimination between benign and malignant cases, Cytopathology 8 (1997) 265–273.

[10] S. Petushi, F.U. Garcia, M. Habe, C. Katsinis, A. Tozeren, Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer, BMC Med. Imaging 6 (14) (2006).

[11] K. Jafari-Khouzani, H. Soltanian-Zadeh, Multiwavelet grading of pathological images of prostate, IEEE Trans. Biomedical Eng. 50 (2003) 697–704.

[12] S. Doyle, M. Hwang, K. Shah, A. Madabhushu, M. Feldman, J. Tomaszeweski, Automated grading of prostate cancer using architectural and textural image features, in: Proceedings of IEEE International Symposium on Biomedical Imaging, 2007, pp. 1284–1287.

[13] A. Tabesh, M. Teverovskiy, H. Pang, V.P. Kumar, D. Verbel, A. Kotsianti, O. Saidi, Multifeature prostate cancer diagnosis and Gleason grading of histological images, IEEE Trans. Med. Imaging 26 (2007) 1366–1378.

[14] M.N. Gurcan, T. Pan, H. Shimada, J. Saltz, Image analysis for neuroblastoma classification: segmentation of cell nuclei, in: Proceedings of IEEE International Conference on Engineering in Medicine and Biology Society, 2006, pp. 4844–4847.

[15] J. Kong, H. Shimada, K. Boyer, J. Saltz, M. Gurcan, Image analysis for automated assessment of grade of neuroblastic differentiation, in: Proceedings of IEEE International Symposium on Biomedical Imaging, 2007, pp. 61–64.

[16] J. Kong, O. Sertel, H. Shimada, K. Boyer, J. Saltz, M. Gurcan, Computer-aided grading of neuroblastic differentiation: multi-resolution and multi-classifier approach, in: Proceedings of IEEE International Conference on Image Processing, 2007, pp. 525–528.

[17] B.B. Cambazoglu, O. Sertel, J. Kong, J. Saltz, M.N. Gurcan, U.V. Catalyurek, Efficient processing of pathological images using the grid: computer-aided prognosis of neuroblastoma, in: Proceedings of IEEE Workshop on Challenges of Large Applications in Distributed Environments, 2007, pp. 35–41.

[18] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 22 (2000) 4–37.

[19] E. Adelson, C. Anderson, J. Bergen, P. Burt, J. Ogden, Pyramid methods in image processing, RCA Engineer 29 (1984) 33–41.

[20] G. Paschos, Perceptually uniform color spaces for color texture analysis: an empirical evaluation, IEEE Trans. Image Process. 10 (2001) 932–937.

[21] M. Tuceryan, A.K. Jain, Texture analysis, in: The Handbook of Pattern Recognition and Computer Vision, second ed., World Scientific Publishing Co., Singapore, 1998.

[22] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 24 (2002) 971–987.

[23] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. SMC 3 (1973) 610–621.

[24] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1990.

[25] P. Pudil, F. Ferri, J. Novovicova, J. Kittler, Floating search methods in feature selection, Pattern Recognition Lett. 15 (11) (1994) 1119–1125.

[26] A. Jain, D. Zongker, Feature selection: evaluation, application, and small sample performance, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 19 (1997) 153–158.

[27] A. Ruiz, O. Sertel, M. Ujaldon, U. Catalyurek, J. Saltz, M. Gurcan, Pathological image analysis using the gpu: stroma classification for neuroblastoma, in: Proceedings of IEEE International Conference on Bioinformatics and Biomedicine, 2007, pp. 78–85.

[28] A. Ruiz, O. Sertel, M. Ujaldon, U. Catalyurek, J. Saltz, M.N. Gurcan, Stroma classification for neuroblastoma on graphics processor, Int. J. Data Mining Bioinformatics 3 (4) (2008).

About the Author—OLCAY SERTEL received his BS from Yildiz Technical University, Istanbul, Turkey, MSc degree from Yeditepe University, Istanbul, Turkey, in 2004 and 2006, both in Computer Engineering. He is currently a PhD student at the Department of Electrical and Computer Engineering and working as a Graduate Research Associate at the Department of Biomedical Informatics at The Ohio State University, Columbus, OH. His research interests include computer vision, image processing and pattern recognition with applications in medicine.

About the Author—JUN KONG received the BS in Information and Control and MS in Electrical Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and 2004. He is currently a PhD student in Department of Electrical and Computer Engineering at The Ohio State University, Columbus, OH. His research interests include computer vision, machine learning, and pathological image analysis.

About the Author—HIROYUKI SHIMADA is currently an Associate Professor of clinical in Children's Hospital Los Angeles and Department of Pathology and Laboratory Medicine at The University of Southern California. He is the Pathologist-of-Record for the Children's Cancer Group (CCG) Neuroblastoma Studies. He is also a core member of International Neuroblastoma Pathology Committee and chairing workshops and other activities. In 1999 this Committee established the International Neuroblastoma Pathology Classification (the Shimada System) by adopting my original classification published in 1984. His research interests include investigating morphological characteristics of pediatric tumors.

About the Author—UMIT V. CATALYUREK is an Associate Professor in the Department of Biomedical Informatics at The Ohio State University. His research interests include graph and hypergraph partitioning algorithms, grid computing, and runtime systems and algorithms for high-performance and data-intensive computing. He received his PhD, MS and BS in computer engineering and information science from Bilkent University, Turkey, in 2000, 1994 and 1992, respectively.

About the Author—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 MD-PhD 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.

About the Author—METIN N. GURCAN is an Assistant Professor at the Department of Biomedical Informatics at The Ohio State University (OSU). He received his MSc degree in Digital Systems Engineering from UMIST, England and his BSc and PhD 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 OSU 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 recipient of the young investigator award from the Children's Neuroblastoma Cancer Foundation, author of over 50 peer-reviewed publications and has a patent in computer-aided diagnosis in volumetric imagery.