AUTOMATED GRADING OF BREAST CANCER HISTOPATHOLOGY USING SPECTRAL CLUSTERING WITH TEXTURAL AND ARCHITECTURAL IMAGE FEATURES

Scott Doyle, Shannon Agner, Anant Madabhushi *

Rutgers University
Department of Biomedical Engineering
Piscataway, New Jersey, 08854

Michael Feldman, John Tomaszewski

University of Pennsylvania,
Department of Surgical Pathology,
Philadelphia, Pennsylvania, USA 19104

ABSTRACT

In this paper we present a novel image analysis methodology for automatically distinguishing low and high grades of breast cancer from digitized histopathology. A set of over 3,400 image features, including textural and nuclear architecture based features, are extracted from a database of 48 breast biopsy tissue studies (30 cancerous and 18 benign images). Spectral clustering is used to reduce the dimensionality of the feature set. A support vector machine (SVM) classifier is used (1) to distinguish between cancerous and non-cancerous images, and (2) to distinguish between images containing low and high grades of cancer. Classification is repeated using different subsets of features to compare their performance. The system achieves a 95.8% accuracy in distinguishing cancer from non-cancer using texture-based characteristics (Gabor filter features), and 93.3% accuracy in distinguishing high from low grades of cancer using architectural features. In addition, we investigate the underlying manifold structure on which the different grades of breast cancer lie as revealed through spectral clustering. The manifold shows a smooth spatial transition from low to high grade breast cancer.

Index Terms— Histopathology, Breast cancer, Image analysis, Automated grading

1. INTRODUCTION

Approximately 178,000 new cases of invasive breast cancer are diagnosed and approximately 41,000 women are lost to breast cancer each year in the U.S. (source: *American Cancer Society*). Fortunately, proper screening and diagnostic techniques dramatically increase the survival rate of diagnosed women. The current screening protocol consists of a mammography to identify suspicious regions of the breast, followed by a tissue biopsy and analysis by a pathologist to determine presence and malignancy of cancer. The malignancy of the disease is determined by examining the degree of tubule formation, mitotic index, and nuclear pleomorphism in the tissue and assigning a numeric grade from 1 (mostly normal) to 3 (mostly abnormal). The Bloom-Richardson (BR) grade is calculated as the sum of these three numbers and correlates well with disease prognosis [1]. However, it has been shown that there is variability among pathologists when using the BR grading system, which can lead to suboptimal treatment [2]. Meyer, et al. [3], using the kappa agreement score between seven pathologists, found that the BR grading scheme is only moderately reproducible. Due to the importance of histological grading in the treatment of breast cancer [4], an automated and quantifiable method of grading breast cancer is desired. Use of computer-aided diagnosis (CAD) for breast mammography has been shown to increase the sensitivity of lesion detection by radiologists by as much as 21% [5].

In contrast to CAD for radiology, relatively little work has been done in quantitative image analysis of breast histopathology using hematoxylin and eosin stained tissue images. Weyn, et al. [6] used wavelets and densiometric features to classify breast tissue nuclei as belonging to high or low grade cancer with an accuracy of 61.52%. Petushi, et al. [7] also found that quantitative histopathological image features contain information that can be used to differentiate between grades of breast cancers. Axelrod, et al. [8] investigated the use of texture and morphology of nuclei to distinguish between low, intermediate, and high nuclear grades of cancer with an overall accuracy of 65%. Tözeren, et al. [9] showed that texture features can distinguish regions of interest on breast tissue histology slides.

We have previously developed a CAD methodology for automated Gleason grading of prostate cancer on digitized histopathology [10]. In this work, we present a quantitative CAD system for detection and grading of breast cancer from digital images of breast tissue. The aim of this work is threefold: (1) to distinguish between "cancer" and "non-cancer" images of breast cancer; (2) to distinguish between low and high BR grades of breast cancer, and (3) to analyze the underlying structure of the nonlinear data extracted from images of breast histopathology and to correlate it with breast cancer biology.

An overview of our system is presented in Figure 1. Our system calculates over 3,400 textural and architectural features from breast tissue images, including Haralick textures, Gabor filtered images, gray level statistical features, and graph-based features derived from spatial arrangement of nuclei. This feature space is then reduced using a spectral clustering (SC) algorithm called graph embedding [11]. The motivations for using spectral clustering are twofold: (1) to visualize the high-dimensional manifold of the data in a linear low-dimensional space, and (2) to avoid the *curse of dimensionality*, a phenomenon that decreases classification accuracy when the number of dimensions greatly exceeds the number of classification objects. By reducing the feature data to three dimensions, we can plot the images as points in $\mathbb{R}^3$ to visualize the high-dimensional data in a low-dimensional space where object adjacencies are preserved. We can then determine if there is a relationship between

*This work was made possible due to grants from the Coulter foundation, New Jersey Commission on Cancer Research, the National Cancer Institute (R21CA127186-01, R03CA128081-01), Society for Imaging Informatics in Medicine, and the United States Department of Defense.

those breast cancer cases which are geodesically adjacent in high-dimensional space and the corresponding tumor biology. Following spectral clustering, we use a support vector machine (SVM) classifier to first distinguish between "cancer" or "non-cancer" images, and then to classify cancerous images as high or low grade cancer. We repeat spectral clustering and classification using individual subgroups of image features to determine the most discriminating feature subgroups for each classification task. Our paper is organized as follows. In Section 2 we describe the methodology of the current work. Section 3 contains the results of our analysis, and in Section 4 we present our concluding remarks.

2. METHODS

2.1. System Overview

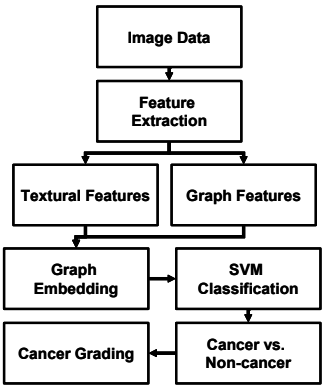


Fig. 1. System overview.

An overview of our system is presented in Figure 1. Glass slides of hematoxylin and eosin stained breast biopsy tissue are scanned into a computer at 40x optical magnification. An expert pathologist labeled regions of tissue within each image according to the Bloom-Richardson (BR) grading scheme [1]. A total of 48 images of tissues were used in this study, comprising 18 benign images and 30 cancerous images. Of the cancer images, 21 were "low grade" cancer (BR grades 5 or 6), and 9 were "high grade" cancer (grades 7 or 8). We denote an image R as $R = (C, f)$, where C is a 2D grid of pixels $c \in C$ and $f(c)$ is a function which assigns a value to c. Each histological image contains m nuclei, and the pixels corresponding to the centroids of the nuclei are manually labeled. We denote these centroid pixels as $c_n^1, c_n^2, \cdots, c_n^m$.

2.2. Feature Extraction

From each image R, we extract a set of graph- and texture-based features to capture the discriminating characteristics of the tissue patterns in each image. A feature vector $\mathbf{f}$ is created for R where each element of $\mathbf{f}$ is a distinct feature value. These values are calculated as described below.

2.2.1. Textural Features

The proliferation of nuclei in cancerous tissue suggests that textural characteristics can help discriminate between different grades of breast cancers [6]. The following features are extracted from each of the three channels of an image (hue, saturation, intensity), using three different window sizes (3x3, 5x5, and 7x7 pixels).

A. Grey Level Features - We calculate 15 gray level features from R as described in [10]. The average, standard deviation, minimum-to-maximum ratio, and mode over all $c \in C$ are then calculated for the values in the feature image to yield a total of 540 gray level feature values for R.

B. Haralick Features - Second-order co-occurrence texture features are described by 16 Haralick features presented in [12]. We calculate a co-occurrence matrix $Z \in \mathbb{R}^{\mathcal{M}\times\mathcal{M}}$ for image R, which is used to generate 16 Haralick feature images. The average, standard deviation, minimum-to-maximum ratio, and mode of the values in the feature images are calculated to yield 576 Haralick texture feature values for R.

C. Gabor Filter Features - Steerable Gabor filters respond to a variety of textural differences in an image. A unique filter kernel G is defined by an orientation parameter $\theta \in \{0, \frac{\pi}{8}, \cdots, \frac{7\pi}{8}\}$ and a scale parameter $s \in \{0, 1, \cdots, 7\}$. We construct 64 Gabor filtered images by varying θ and s. The average, standard deviation, minimum-to-maximum ratio, and mode over all $c \in C$ are calculated for each feature image to yield a total of 2,304 Gabor feature values for R.

2.2.2. Graph Features

The shape and arrangement of nuclei within a histological image region is also related to the cancer progression, and this architecture may be quantified using graph-based techniques [13].

A. Voronoi Diagram - The Voronoi diagram $\mathbf{V}$ [13] comprises a set of polygons $\mathbf{P} = \{P_1, P_2, \cdots, P_m\}$. Any pixel $c \in R$ is included in polygon P_a if $\mathbf{d}(c, c_n^a) = \min_j\{||c - c_n^j||\}$ where $a, j \in \{1, 2, \cdots, m\}$ and $\mathbf{d}(c, d)$ is the Euclidean distance between any two pixels $c, d \in R$. We calculate the area, perimeter length, and chord length of all $\mathbf{P} \in \mathbf{V}$, and the average, standard deviation, minimum-to-maximum ratio, and disorder [13] are calculated over all $\mathbf{P}$, giving 12 feature values for R.

B. Delaunay Triangulation - The Delaunay graph $\mathbf{D}$ is constructed such that any two unique nuclear centroids c_n^a and c_n^b, where $a, b \in \{1, 2, \cdots, m\}$, are connected by an edge $E^{a,b}$ if P_a and P_b share a side in $\mathbf{V}$. We calculate the side lengths and areas for all triangles in $\mathbf{D}$, and take the average, standard deviation, minimum-to-maximum ratio, and disorder of these to obtain 8 additional feature values for R.

C. Minimum Spanning Tree - Given a connected, undirected graph, a spanning tree $\mathbf{S}$ of that graph is a subgraph that connects all vertices. A single graph can have many different $\mathbf{S}$. Weights $\omega_{\mathbf{S}}^E$ are assigned to each edge E in each $\mathbf{S}$ based on the length of E in $\mathbf{S}$. The sum of all weights $\omega_{\mathbf{S}}^E$ in each $\mathbf{S}$ is determined to give the weight $\widehat{\omega}_{\mathbf{S}}$ assigned to each $\mathbf{S}$. The minimum spanning tree (MST) denoted by $\mathbf{S}'$ has a weight $\widehat{\omega}'_{\mathbf{S}}$ less than or equal to $\widehat{\omega}_{\mathbf{S}}$ for every other spanning tree $\mathbf{S}$. We calculate the average, standard deviation, minimum-to-maximum ratio, and disorder of the branch lengths in $\mathbf{S}'$ to obtain 4 additional features for R.

D. Nuclear Features - Nuclear density $\Pi^{\mathcal{D}}$ is computed as $\Pi^{\mathcal{D}} = \frac{m}{|R|}$, where $|R|$ is the cardinality of R. For each nuclear centroid c_n^a, $N(\mu, c_n^a)$ is the set of pixels $c \in R$ contained within a circle with its center at c_n^a and radius μ. We compute the number of c_n^j, $j \neq a$, $j, a \in \{1, 2, \cdots, m\}$ which are in set $N(\mu, c_n^a)$ for $\mu \in \{10, 20 \cdots, 50\}$. We also compute the μ required to obtain $N(\mu, c_n^a) \in \{3, 5, 7\}$. The average, standard deviation, and disorder of these values for all c_n^j in R is calculated to yield 24 additional features for R.

Figure 2 illustrates these graphs for normal benign tissue (Figure 2 (a)-(d)), low-grade cancer (Figure 2 (e)-(h)), and high-grade cancer (Figure 2 (i)-(l)).

2.3. Spectral Clustering

Spectral clustering (SC) algorithms reduce the dimensionality of a data set from M to M' dimensions, where $M' << M$. We used

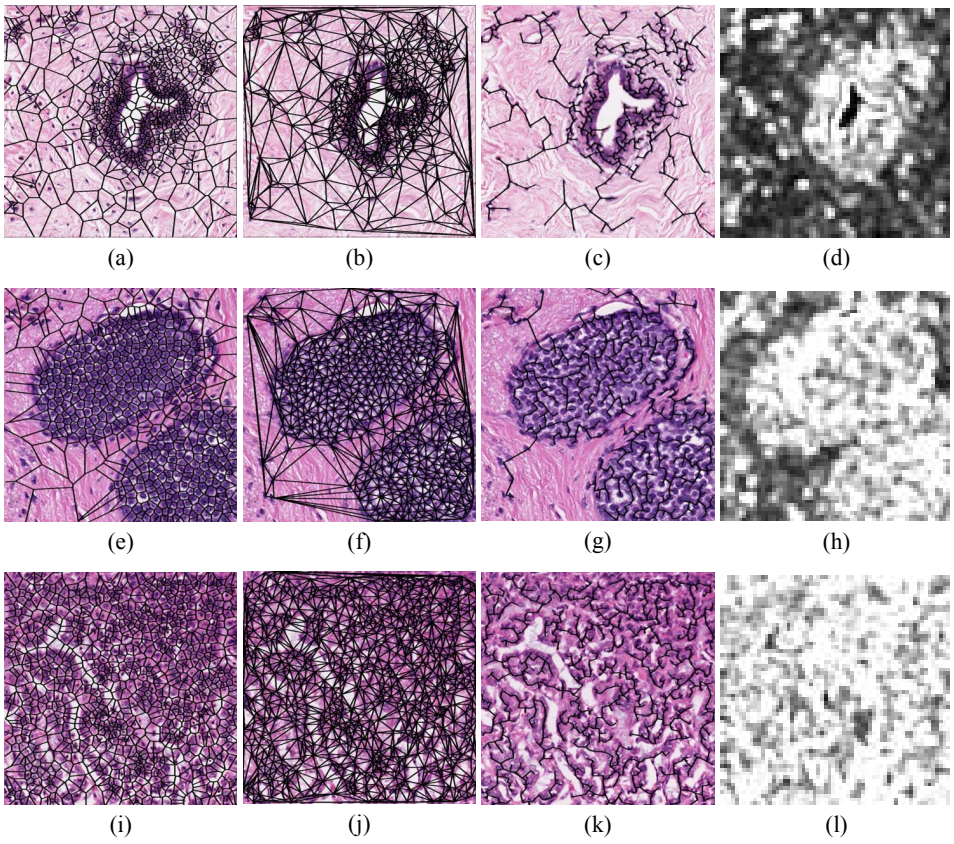


Fig. 2. Example of the ((a), (e), (i)) Voronoi, ((b), (f), (j)) Delaunay, and ((c), (g), (k)) Minimum Spanning Tree graphs, as well an example of a ((d), (h), (l)) Haralick texture image, calculated for ((a)-(d)) benign tissue, ((e)-(h)) low-grade cancer, and ((i)-(l)) high-grade cancer.

graph embedding [11], a nonlinear technique which seeks to find an optimal projection of the data. We construct a confusion matrix $\mathcal{Y}$ to describe the pairwise similarity between the database of images:

$$\mathcal{Y}(p,q) = e^{-||\mathbf{f}_p - \mathbf{f}_q||} \in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}, \quad (1)$$

where $\mathbf{f}_p$ and $\mathbf{f}_q$ are the feature vectors computed for any two images R_p and R_q, respectively, where $p, q \in \{1, 2, \cdots, \mathbf{N}\}$ and where $\mathbf{N}$ is the total number of images in the data set. The embedding vector $\mathcal{X}$ is obtained from the maximization of the function:

$$\mathcal{E}_{\mathcal{Y}}(\mathcal{X}) = 2\eta \frac{\mathcal{X}^{\mathbf{T}}(\mathcal{B}-\mathcal{Y})\mathcal{X}}{\mathcal{X}^{\mathbf{T}}\mathcal{B}\mathcal{X}}, \quad (2)$$

where $\mathcal{B}(p,p) = \sum_q \mathcal{Y}(p,q)$ and $\eta = |\mathbf{N}| - 1$. The M'-dimensional embedding space is defined by the eigenvectors corresponding to the smallest M' eigenvalues of $(\mathcal{B}-\mathcal{Y})\mathcal{X} = \lambda\mathcal{B}\mathcal{X}$. For image R defined by feature vector $\mathbf{f}$, the embedding $\mathcal{X}(R)$ contains the coordinates of R in the embedding space and is given as $\mathcal{X}(R) = [w_z(R)|z \in \{1, 2, \cdots, M'\}]$, where $w_z(R)$ are the z eigenvalues associated with $\mathcal{X}(R)$.

3. RESULTS AND DISCUSSION

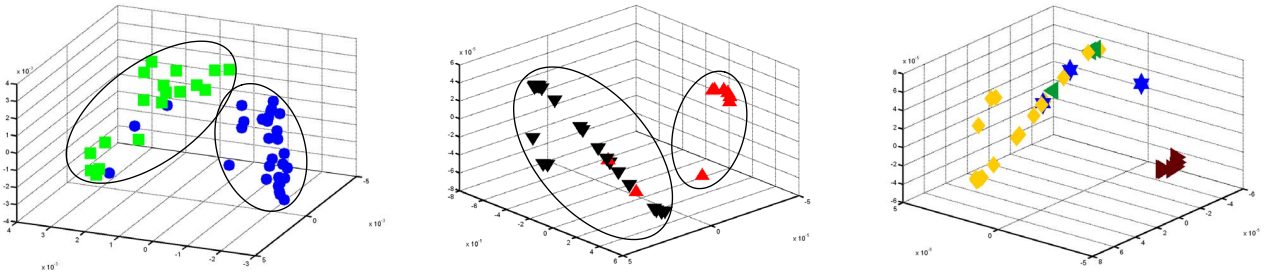
In this work we are concerned with three aims: (1) distinguishing between "cancer" and "non-cancer" images of breast tissue; (2) distinguishing between low and high BR grades of breast cancer, and (3) analyzing the underlying structure of the nonlinear data extracted from images of breast histopathology. For each classification task, we determine which feature subgroup described in Section 2.2 is most discriminating for each classification task. We divide the feature set into 10 groups as shown in the first column in Table 1. Features are not exclusive to one subgroup, since "All Textural" features include Gabor, Grey Level, and Haralick features, and "All Architectural" features includes Voronoi, Delaunay, MST, and Nuclear features. These subgroups are reduced with SC and used for classification. For each classification task, a third of the dataset is randomly selected for training, while all remaining images are used for testing.

3.1. Classifier Accuracy

SVM classification accuracies are given in Table 1. The highest accuracy in distinguishing cancerous from non-cancerous images is 95.8%, obtained when using only Gabor filter features. The highest classification accuracy obtained when distinguishing high from low grade cancer images is 93.3%, obtained using all of the architectural features together: Voronoi, Delaunay, MST, and nuclear features.

3.2. Spectral Clustering Analysis

The scatter plots in Figure 3 show the data in the low-dimensional space obtained through spectral clustering. The axes of the plots correspond to the three dominant eigenvectors found by SC. Fig-



(a) (b) (c)

Fig. 3. Graph Embedding results for (a) cancerous (blue circles) vs. non-cancerous images (green squares), (b) high-grade (red up-triangles) vs. low-grade images (black down-triangles), and (c) individual grades of images: Grade 5 (orange diamonds), Grade 6 (green left-triangles), Grade 7 (blue stars), and Grade 8 (maroon right-arrows). Note that the manifold in (b) and (c) is the same; only the view and the labels on the data have been changed. The manifold structure in (c) reveals a smooth transition in BR grade from low-grade (BR grade 5) to high-grade (BR grade 8) cancer.

Feature Subtype	**Classification Task**	
	Cancer vs. Benign	**Low vs. High Grade**
All Features (3,468)	0.667	0.700
All Textural (3,420)	0.667	0.733
All Architectural (51)	0.771	**0.933**
Gabor (2,304)	**0.958**	0.700
Grey Level (540)	0.938	0.700
Haralick (576)	0.625	0.767
Voronoi (12)	0.792	0.900
Delaunay (8)	0.854	0.900
MST (4)	0.938	0.900
Nuclear (27)	0.729	0.900

Table 1. Classification accuracy using different feature subsets for cancer vs. non-cancer images and high vs. low grade images.

ure 3 (a) shows cancerous images (blue circles) versus non-cancer images (green squares). Figure 3 (b) shows high-grade cancer images (red up-triangles) versus low-grade cancer images (black down-triangles). Black ellipses denote class clusters in both figures. Figure 3 (c) shows a rotated view of (b), with labels altered to show BR grades of cancer: the orange diamonds are grade 5, green left-triangles are grade 6, blue stars are grade 7, and maroon right-arrows are grade 8. There is a transition from the low grade images on the left of the plot through intermediate grades (6 and 7) in the middle, ending with high grade images on the right. This suggests that the structure of the data reflects the biology of the images, and that by analyzing this structure in a low-dimensional subspace, we can appreciate the relationship between the underlying biology and the image features that are calculated from the image.

4. CONCLUDING REMARKS

In this paper, we present an automated system for quantitative histopathological analysis of breast tissue images. The main contributions of this work are:

- A set of image features that can distinguish different grades of breast cancer on digitized histopathology,
- A system capable of distinguishing between cancer and benign images and between images of high and low grades of cancer with a high degree of accuracy, and
- Use of spectral clustering to visualize the underlying biological relationship between different breast cancer studies.

The results obtained from SC suggest that we can appreciate the underlying manifold structure on which different grades of breast cancer lie. The manifold shows a smooth transition from low to intermediate to high grade breast cancer, and with a densely populated manifold, we can determine if there is a connection between an image's location on the high-dimensional manifold and the biological grade of the cancer growth.

5. REFERENCES

[1] H. J. Bloom and W. W. Richardson, "Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years.," *Br J Cancer*, vol. 11, no. 3, pp. 359–377, Sep 1957.

[2] L. W. Dalton et al., "Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement.," *Mod Pathol*, vol. 13, no. 7, pp. 730–735, Jul 2000.

[3] J. S. Meyer et al., "Breast carcinoma malignancy grading by bloom-richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index.," *Mod Pathol*, vol. 18, no. 8, pp. 1067–1078, Aug 2005.

[4] G. Contesso et al., "The importance of histologic grade in long-term prognosis of breast cancer: a study of 1,010 patients, uniformly treated at the institut gustave-roussy.," *J Clin Oncol*, vol. 5, no. 9, pp. 1378–1386, Sep 1987.

[5] R. F. Brem et al., "Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial.," *AJR Am J Roentgenol*, vol. 181, no. 3, pp. 687–693, Sep 2003.

[6] B. Weyn et al., "Automated breast tumor diagnosis and grading based on wavelet chromatin texture description.," *Cytometry*, vol. 33, no. 1, pp. 32–40, Sep 1998.

[7] S. Petushi et al., "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer.," *BMC Med Img*, vol. 6, pp. 14, 2006.

[8] D.E. Axelrod et al., "Effect of quantitative nuclear image features on recurrence of ductal carcinoma in situ (dcis) of the breast," *Cancer Informatics*, vol. 4, pp. 99–109, 2008.

[9] A. Tözeren and B. Kara¸cali, "Automated detection of regions of interest for tissue microarray experiments: An image texture analysis," *BMC Medical Imaging*, vol. 7, 2007.

[10] S. Doyle A. Madabhushi et al., "Automated grading of prostate cancer using architectural and textural image features," in *ISBI 2007*, 2007, pp. 1284–1287.

[11] A. Madabhushi et al., "Graph embedding to improve supervised classification and novel class detection: application to prostate cancer.," *MICCAI*, vol. 8, no. Pt 1, pp. 729–737, 2005.

[12] R M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[13] J. Sudbø, R. Marcelpoil, and A. Reith, "New algorithms based on the voronoi diagram applied in a pilot study on normal mucosa and carcinomas.," *Anal Cell Pathol*, vol. 21, no. 2, pp. 71–86, 2000.