COMPUTER-AIDED PROGNOSIS OF ER+ BREAST CANCER HISTOPATHOLOGY AND CORRELATING SURVIVAL OUTCOME WITH ONCOTYPE DX ASSAY

Ajay Basavanhally, Jun Xu,
Anant Madabhushi

Rutgers, The State University of New Jersey
Department of Biomedical Engineering
Piscataway, New Jersey, 08854, USA

Shridar Ganesan

The Cancer Institute of New Jersey
New Brunswick, New Jersey, 08903, USA

ABSTRACT

The current gold standard for predicting disease survival and outcome for lymph node-negative, estrogen receptor-positive breast cancer (LN-, ER+ BC) patients is via the gene-expression based assay, Onco*type* DX. In this paper, we present a novel computer-aided prognosis (CAP) scheme that employs quantitatively derived image information to predict patient outcome analogous to the Onco*type* DX Recurrence Score (RS), with high RS implying poor outcome and vice versa. While digital pathology has made tissue specimens amenable to computer-aided diagnosis (CAD) for disease detection, our CAP scheme is the first of its kind for predicting disease outcome and patient survival. Since cancer grade is known to be correlated to disease outcome, low grade implying good outcome and vice versa, our CAP scheme captures quantitative image features that are reflective of BC grade. Our scheme involves first semi-automatically detecting BC nuclei via an Expectation Maximization driven algorithm. Using the nuclear centroids, two graphs (Delaunay Triangulation and Minimum Spanning Tree) are constructed and a total of 12 features are extracted from each image. A non-linear dimensionality reduction scheme, Graph Embedding, projects the image-derived features into a low-dimensional space, and a Support Vector Machine classifies the BC images in the reduced dimensional space. On a cohort of 37 samples, and for 100 trials of 3-fold randomized cross-validation, the SVM yielded a mean accuracy of 84.15% in distinguishing samples with low and high RS and 84.12% in distinguishing low and high grade BC. The projection of the high-dimensional image feature data to a 1D line for all BC samples via GE shows a clear separation between, low, intermediate, and high BC grades, which in turn shows high correlation with low, medium, and high RS. The results suggest that our image-based CAP scheme might provide a cheaper alternative to Onco*type* DX in predicting BC outcome.

Index Terms— Breast cancer, Image analysis, Histopathology, Cancer grade, Oncotype DX, prognosis

1. INTRODUCTION

Breast cancer (BC) is one of the leading causes of cancer-related deaths in women, with an expected annual incidence greater than 182,000 in the United States in 2008 (source: *American Cancer Society*). One subset of BC comprises cancer cells that have not spread to the lymph nodes and with overexpression of the estrogen receptor (LN-, ER+ BC). Although cases of LN-, ER+ BC are treated with a combination of chemotherapy and adjuvant hormone therapy, the specific prognosis and treatment is often determined by the Onco*type* DX gene expression assay [1]. This assay produces a Recurrence Score (RS) between 0–100 that is positively correlated to the likelihood for distant recurrence and the expected benefit from chemotherapy [1].

In this paper, we present an image-based computer-aided prognosis (CAP) scheme that seeks to replicate the prognostic power of molecular assays in BC histopathology. Using only the tissue slide samples, a mechanism for digital slide scanning, and a computer, our image-based CAP scheme aims to overcome many of the drawbacks associated with Onco*type* DX, including the

- high cost associated with the assay,
- limited laboratory facilities with specialized equipment, and
- length of time between biopsy and prognostic prediction.

Our research includes key methodological contributions and the use of several state of the art machine learning schemes, including

- a robust, efficient method to automatically detect BC nuclei,
- image features to describe the architectural arrangement of BC nuclei and hence, quantitatively describe cancer grade, and
- the use of non-linear dimensionality reduction to classify and visualize the underlying biological data structure in a low-dimensional representation.

The manual detection of BC nuclei in histopathology is a tedious and time-consuming process that is unfeasible in the clinical setting. Previous approaches to cell segmentation – thresholding [2], clustering [3], and active contour models [4] – are not very robust to the highly variable shapes and sizes of BC nuclei, as well as artifacts in the histological fixing, staining, and digitization processes. We present an semi-automated nuclear detection scheme based on the Expectation Maximization (EM) algorithm [5].

Previous work [1] has shown that the Onco*type* DX RS is correlated with BC grade. Cancer grade reflects the architectural arrangement of the tissue and is correlated with survival (high grade implies poor outcome). However, pathologists often disagree on the grade of a BC study. Hence an image analysis system that can reproducibly and quantitatively characterize tissue architecture can be used to predict patient outcome. In fact, with the recent advent of digital pathology, researchers have begun to explore automated image analysis of

Thanks to the Coulter Foundation, New Jersey Commission on Cancer Research, the National Cancer Institute, the Society for Imaging Informatics in Medicine, and the Life Sciences Commercialization Award for supporting this work. The authors would like to thank Dr. David Foran for his assistance in digitizing whole slide histology.

978-1-4244-3932-4/09/$25.00 ©2009 IEEE

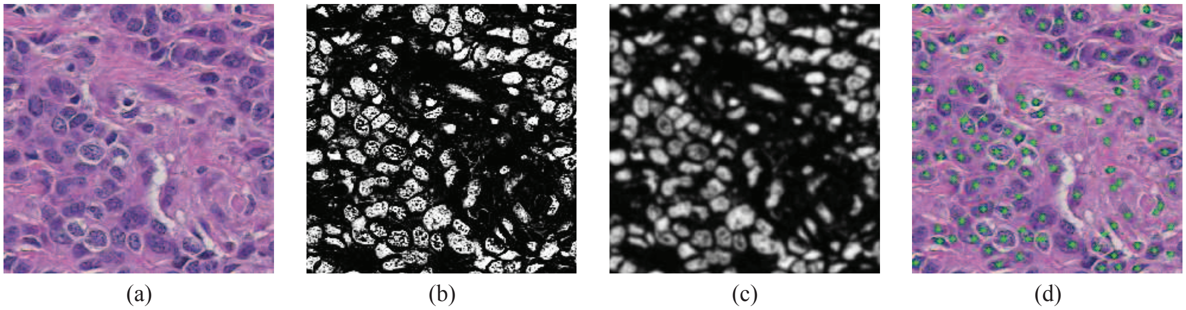


Fig. 1. A (a) LN-, ER+ BC histopathology image shown along with its corresponding (b) EM-based segmentation of BC nuclei. The segmentation in (b) is (c) smoothed to help detect individual nuclei and the (d) final detected nuclear centroids are used for feature extraction.

BC histopathology. Wolberg et al. [6] used nuclear features from manually segmented BC nuclei to distinguish benign and malignant images. Bilgin et al. [7] explored the use of hierarchical graphs to model the architecture of BC histopathology. Textural features were used by Hall et al. [8] to examine variations in immunohistochemical staining. In this paper we derive architectural features to characterize the arrangement of BC nuclei and hence capture BC grade. Our scheme is similar to the method presented by Doyle et al. [9], where different graphs were constructed using BC nuclei as vertices and the quantitative features derived from these graphs were used to successfully stratify BC grade.

In this paper we also employ Graph Embedding (GE), a non-parametric type of non-linear dimensionality reduction [9], to project the image-derived features from each BC tissue specimen onto a reduced 3D space, and subsequently, down to a 1D line. A Support Vector Machine (SVM) classifier [10] is employed to evaluate the discriminability of the architectural features with respect to BC grade in the reduced 3D space. The further projection of the image data to a 1D line allows us to define image-based risk scores, analogous to the Oncotype DX RS, for poor, intermediate, and good outcome. This image-based risk score predictor could potentially supplant Onco*type* DX to predict BC outcome and survival.

2. AUTOMATED DETECTION OF NUCLEI USING EM-BASED GAUSSIAN MIXTURE

2.1. Dataset

A total of 37 H & E stained breast histopathology images were collected from a cohort of 17 patients and scanned into a computer using a high resolution whole slide scanner at 20x optical magnification. For all methods, we define the data set $\mathbf{Z} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_\mathcal{M}\}$ of $\mathcal{M}$ images, where an image $\mathcal{C} = (C, g)$ is a 2D set of pixels $c \in C$ and g is the associated intensity function. Each $\mathcal{C}$ is associated with a architectural feature set $\mathbf{F}(\mathcal{C})$, an Onco*type* DX RS $L^{\mathrm{RS}}(\mathcal{C}) \in \{0, 1, \ldots, 100\}$, and BC grade $L^{\mathrm{GR}}(\mathcal{C}) \in \{\mathrm{LG}, \mathrm{MG}, \mathrm{HG}\}$, where LG, MG, and HG represent low-, medium-, and high- grade cancer, respectively. The 37 samples were also classified based on their RS, where $L^{\mathrm{RS}}(\mathcal{C})$ is binned into good (RS<22), intermediate (23<RS<30), and poor (31<RS<100) prognosis categories.

2.2. EM-Based Segmentation of Cancer Nuclei

To segment BC nuclei, each image $\mathcal{C}$ is modeled as a Gaussian mixture of $K = 5$ components, where $\kappa \in \{1, 2, \ldots, K\}$. We optimize the model parameter set $\gamma^i = \{\mu^i_\kappa, \sigma^i_\kappa, \mathbf{p}^i_\kappa : \forall \kappa\}$, comprising the mean μ_κ, covariance σ_κ, and *a priori* probability $\mathbf{p}_\kappa$ at iteration i. The mixture is initialized to γ^0 via K-means clustering of RGB values over all $c \in C$. The Expectation step calculates the posterior probability

$$p^i(\kappa|g(c)) = \frac{\mathbf{p}^i_\kappa N(g(c)|\mu^i_\kappa, \sigma^i_\kappa)}{\sum_{j=1}^{K} \mathbf{p}^i_j N(g(c)|\mu^i_j, \sigma^i_j)},$$

where $N(g(c)|\mu_\kappa, \sigma_\kappa)$ represents the value of Gaussian component κ at intensity $g(c)$. The Maximization step uses p^i to calculate $\gamma^{i+1} = \{\mu^{i+1}_\kappa, \sigma^{i+1}_\kappa, \mathbf{p}^{i+1}_\kappa\}$ [5]. The EM algorithm converges when $\|(\mathcal{L}^{i+1} - \mathcal{L}^i)/\mathcal{L}^i\|_2 < \epsilon$, where $\mathcal{L}^i$ is the log likelihood of the Gaussian mixture model with parameters γ^i and $\epsilon = 10^{-5}$ is determined empirically. Based on posterior probability, a grayscale "scene" over all $c \in C$ (Figure 1(b)) is saved for each $\kappa \in \{1, 2, \ldots, K\}$.

The scene that best represents BC nuclei (Figure 1(b)) is selected manually and smoothed (Figure 1(c)) to reduce intra-nuclear intensity variations. Morphological and connected component operations are then applied to identify individual objects corresponding to BC nuclei and the corresponding set of n nuclear centroids $\mathbf{V} \in \{v_1, v_2, \ldots, v_n\}$ is found for each $\mathcal{C} \in \mathbf{Z}$ (Figure 1(d)).

3. FEATURE EXTRACTION

A complete, undirected graph $\mathcal{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W})$ is defined by a vertex-set of nuclear centroids $\mathbf{V}$, an edge-set $\mathbf{E} = \{E_1, E_2, \ldots, E_m\}$ connecting the nuclear centroids such that $(v_1, v_2) \in \mathbf{E}$ with $v_1, v_2 \in \mathbf{V}$, and a set of weights $\mathbf{W} = \{W_1, W_2, \ldots, W_m\}$ proportional to the length of each $E \in \mathbf{E}$. A total of 12 architectural features are extracted for each image based on the Delaunay Triangulation and a Minimum Spanning Tree [9] (Figure 2). Below we describe the construction of these 2 graphs.

3.1. Delaunay Triangulation

A Delaunay graph $\mathcal{G}_\mathcal{D} = (\mathbf{V}, \mathbf{E}_\mathcal{D}, \mathbf{W}_\mathcal{D})$ (Figure 2(b)) is a spanning subgraph of $\mathcal{G}$ that is easily calculated from the Voronoi Diagram $\mathcal{R}$. Each $\mathcal{R}$ is defined by a set of polygons $\mathbf{P} = \{P(v_1), P(v_2), \ldots, P(v_n)\}$ surrounding all nuclear centroids $\mathbf{V}$. Each pixel $c \in C$ is linked to the nearest $v \in \mathbf{V}$ (via Euclidean distance) and added to the associated polygon $P(v) \in \mathbf{P}$. The Delaunay graph $\mathcal{G}_\mathcal{D}$ is simply the dual graph of $\mathcal{R}$ and is constructed such that if $P(v_a), P(v_b) \in \mathbf{P}$ share a side, their nuclear centroids $v_a, v_b \in \mathbf{V}$ are connected by an edge $(v_a, v_b) \in \mathbf{E}_\mathcal{D}$. The mean,

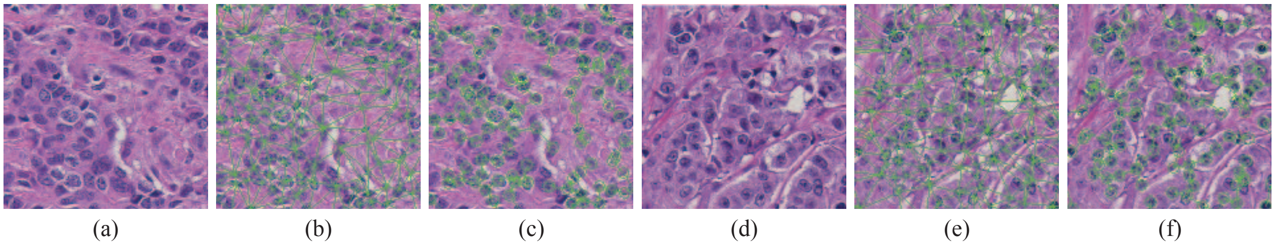


Fig. 2. (a), (d) Low and high grade LN-, ER+ BC samples are shown with (b), (e) Delaunay Triangulation and (c), (f) Minimum Spanning Tree graphs overlaid.

standard deviation, minimum/maximum (min/max) ratio, and disorder are calculated for the side length and area of all triangles in $\mathcal{G}_{\mathcal{D}}$, providing a set of 8 features $\mathbf{f}_{\mathcal{D}}$ for each $\mathcal{C} \in \mathbf{Z}$.

3.2. Minimum Spanning Tree

A spanning tree $\mathcal{G}_{\mathcal{S}} = (\mathbf{V}, \mathbf{E}_{\mathcal{S}}, \mathbf{W}_{\mathcal{S}})$ refers to any spanning subgraph of $\mathcal{G}$. The total weight $\widehat{\mathbf{W}}_{\mathcal{S}}$ for each subgraph is calculated by summing all individual weights $W \in \mathbf{W}_{\mathcal{S}}$. The Minimum Spanning Tree (Figure 2(c)) $\mathcal{G}_{\mathrm{MST}} = \arg\min_{\mathcal{G}_{\mathcal{S}} \in \mathcal{G}} \widehat{\mathbf{W}}_{\mathcal{S}}$ is the subgraph with the lowest total weight. The mean, standard deviation, min/max ratio, and disorder of the branch lengths in $\mathcal{G}_{\mathrm{MST}}$ provide a set of 4 features $\mathbf{f}_{\mathrm{MST}}$ for each $\mathcal{C} \in \mathbf{Z}$.

4. DIMENSIONALITY REDUCTION USING GRAPH EMBEDDING AND SUPPORT VECTOR MACHINE BASED CLASSIFICATION

4.1. Projecting Data to a 3D Space via Graph Embedding

We use Graph Embedding (GE) to transform the architectural feature set into a low-dimensional embedding [9]. For each $\mathcal{C} \in \mathbf{Z}$, a 12-dimensional architectural feature set is defined as the superset $\mathbf{F}(\mathcal{C}) = \{\mathbf{f}_{\mathcal{D}}, \mathbf{f}_{\mathrm{MST}}\}$ containing all features derived from Delaunay Triangulation and Minimum Spanning Tree. Given histopathology images $\mathcal{C}_a, \mathcal{C}_b \in \mathbf{Z}$, a confusion matrix $\mathcal{W}(a,b) = \exp(-\|\mathbf{F}(\mathcal{C}_a) - \mathbf{F}(\mathcal{C}_b)\|_2) \in \mathbb{R}^{\mathcal{M} \times \mathcal{M}}$ is first constructed $\forall a, b$. The optimal embedding vector $\mathbf{F}'$ is obtained from the maximization of the function,

$$\mathcal{E}(\mathbf{F}') = 2(\mathcal{M} - 1) \cdot \mathrm{trace}\left[\frac{\mathbf{F}'^{\mathrm{T}}(\mathcal{A} - \mathcal{W})\mathbf{F}'}{\mathbf{F}'^{\mathrm{T}}\mathcal{A}\mathbf{F}'}\right],$$

where $\mathcal{A}(a,a) = \sum_b \mathcal{W}(a,b)$. The low-dimensional embedding space is defined by the Eigen vectors corresponding to the β smallest Eigen values of $(\mathcal{A} - \mathcal{W})\mathbf{F}' = \lambda \mathcal{A} \mathbf{F}'$.

Reducing the high-dimensional feature space to a 3D Eigen subspace allows us to evaluate the discriminability of the image-derived features in distinguishing samples with different cancer grade patterns and hence different prognoses.

4.2. SVM-based Classification via Cross-Validation

A support vector machine (SVM) classifier [10] is trained using image-derived features to distinguish images with different grades using a k-fold cross-validation scheme. The data set $\mathbf{Z}$ is divided into training $\mathbf{Z}_{\mathrm{tra}}$ and testing $\mathbf{Z}_{\mathrm{tes}}$ subsets, where $\mathbf{Z}_{\mathrm{tra}} \cap \mathbf{Z}_{\mathrm{tes}} = \emptyset$. The SVM classifier projects $\mathbf{F}(\mathbf{Z}_{\mathrm{tra}})$ onto a higher dimensional space using a linear kernel and the hyperplane that most accurately separates the two classes is determined. The classifier is evaluated by projecting $\mathbf{F}(\mathbf{Z}_{\mathrm{tes}})$ and comparing all $\mathcal{C} \in \mathbf{Z}_{\mathrm{tes}}$ to the hyperplane. Image $\mathcal{C}$ is said to be correctly classified if its SVM-derived class matches the clinician's ground truth label.

SVM training is performed via stratified, randomized k-fold cross-validation algorithm, whereby $\mathbf{Z}$ is divided randomly into k subsets. The samples from $k-1$ subsets are pooled into $\mathbf{Z}_{\mathrm{tra}}$ and the remaining subset is used as $\mathbf{Z}_{\mathrm{tes}}$. For each of the k iterations, a different subset is used as $\mathbf{Z}_{\mathrm{tes}}$ while the remaining subsets are used for $\mathbf{Z}_{\mathrm{tra}}$. Using a value of $k = 3$, the entire cross-validation algorithm is repeated for 100 trials and the resulting mean μ_{ACC} and standard deviation σ_{ACC} of the classification accuracy are determined.

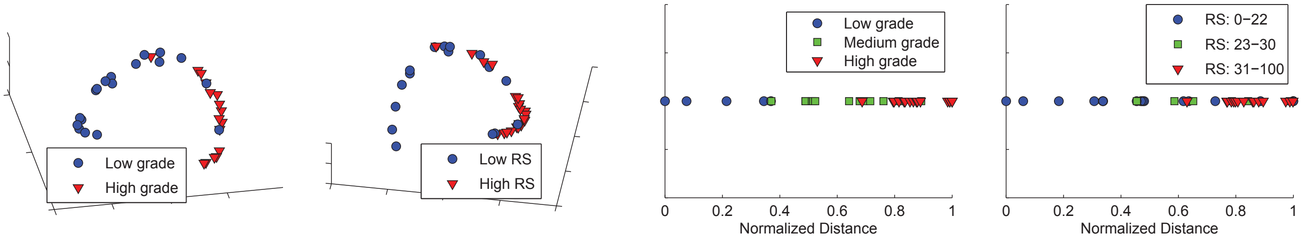
4.3. Geodesic Distance-based Projection from 3D to 1D

The 3D GE manifold (obtained as described in Section 4.1) can be "unwrapped" into a 1D (linear) space simply by selecting the image $\mathcal{C}_1$ at one end of the manifold as an anchor point and using the Euclidean distance metric to find the next image nearest to it on the 3D manifold. By using $\mathcal{C}_a$ as the new anchor point, this process is repeated until all $\mathcal{C} \in \mathbf{Z}$ are included. Thus the geodesic distances between all scenes $\mathcal{C}$ embedded on the mainfold are determined and GE is again employed to project the data down to a 1D line. By uncovering the grade (outcome) labels of the samples on this 1D projection and their relative locations, an image-based recurrence score can be determined to distinguish between low, intermediate, and high BC grades (and hence outcomes). For any new image $\mathcal{C}_b$ projected onto this line, the relative distance of $\mathcal{C}_b$ from poor, intermediate, and good outcome samples on the trained manifold will enable prediction of prognosis for $\mathcal{C}_b$.

5. RESULTS AND DISCUSSION

5.1. Image-Based Discrimination of Grade

A SVM trained via 3-fold cross-validation on the unreduced $\mathbf{F}$ and reduced (3D) $\mathbf{F}'$ architectural feature sets was able to distinguish high and low grade BC histopathology images with classification accuracies of 74.82% ± 5.74% and 84.12% ± 5.42%, respectively, over 100 runs (Table 1). These results appear to confirm that GE has embedded the architectural feature set without any significant loss of information. The success of the architectural features is confirmed qualitatively by the clear separation between high and low BC grade on the 3D manifold (Figure 3(a)).



(a) (b) (c) (d)

Fig. 3. Graph Embedding plots of architectural features show clear separation of different (a) BC grades and (b) RS labels. The embeddings are projected into a 1D line, where (c) BC grade and (d) RS are defined by a single score.

5.2. Correlation Between Image-Based Signatures in Grade Discrimination and Oncotype DX Recurrence Scores

Replacing the grade labels with the RS labels, the SVM trained via 3-fold cross-validation on $\mathbf{F}'$ and $\mathbf{F}$ yielded classification accuracies of 84.15% ± 3.10% and 84.56% ± 2.69%, respectively (Table 1). This shows that a correlation exists between molecular prognostic assays such as Onco*type* DX and the spatial arrangement of nuclei and histological structures in BC histopathology. The 3D manifolds in Figures 3(a), (b) reveal a similar underlying biological stratification that exists in BC grade and Onco*type* DX RS, in turn suggesting that the quantitative image information employed to characterize BC grade could recapitulate the prognostic capabilities of Onco*type* DX. The curvilinear 3D manifold on which the different BC grades (low to high) reside in a smooth continuum may potentially offer insight into BC biology as well.

	Automated Detection	Manual Detection
RS ($\mathbf{F}'$)	84.15% ± 3.10%	71.92% ± 4.66%
Grade ($\mathbf{F}'$)	84.12% ± 5.42%	85.71% ± 4.89%
RS ($\mathbf{F}$)	84.56% ± 2.69%	71.65% ± 5.88%
Grade ($\mathbf{F}$)	74.82% ± 5.74%	85.00% ± 4.51%

Table 1. μ_{ACC} and σ_{ACC} over 100 trials of 3-fold cross-validation for both automatically and manually detected BC nuclei. Results are reported for the original $\mathbf{F}$ and low-dimensional $\mathbf{F}'$ feature sets using both the RS (L^{RS}) and cancer grade (L^{GR}) labels.

5.3. Creating an Image-Based Assay Using 1D Projection

Figures 3(c), (d) represent the 1D projections of the 3D manifolds shown in Figures 3(a), (b), respectively. The manifolds reveal a smooth, continuous progression from low to medium to high levels in terms of both RS and histological (grade) for all LN-, ER+ BC samples considered. The similarity between the 1D manifolds (Figures 3(c), (d)) suggest that our image-based CAP system can be used to generate a prognostic assay to predict survival scores in much the same way as Oncotype DX.

6. CONCLUDING REMARKS

We have presented an image analysis framework for prognosticating disease outcome and survival of LN-, ER+ BC samples that appears to replicate the prognostic capabilities of Onco*type* DX. Furthermore, we have shown that the relationship between image-based signatures and established prognostic indicators (RS and cancer grade) can be represented by a continuous, low-dimensional manifold. The ability to unwrap this manifold into a one-dimensional scale could be modeled into an image-based assay that will yield a prognostic score from quantitative analysis of the biopsy specimens alone. Our findings in this paper suggest that more expensive molecular assays could be replaced with a cheap image-based test for predicting patient outcome. In future work, we will validate our results on a much larger data cohort.

7. REFERENCES

[1] M.B. Flanagan, D.J. Dabbs, et al., "Histopathologic variables predict oncotype dx recurrence score.," *Mod Pathol*, vol. 21, no. 10, pp. 1255–1261, Oct 2008.

[2] V.R. Korde, H. Bartels, et al., "Automatic segmentation of cell nuclei in bladder and skin tissue for karyometric analysis," in *Biophotonics. Proceedings of the SPIE,*, 2007, vol. 6633.

[3] L. Latson et al., "Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy.," *Anal Quant Cytol Histol*, vol. 25, no. 6, pp. 321–331, Dec 2003.

[4] X. Xie et al., "Mac: Magnetostatic active contour model," *IEEE Trans on PAMI*, vol. 30, no. 4, pp. 632–646, 2008.

[5] A. Ramme, N. Devries, et al., "Semi-automated phalanx bone segmentation using the expectation maximization algorithm.," *J Digit Imaging*, Sep 2008.

[6] W.H. Wolberg, W.N. Street, et al., "Computer-derived nuclear features distinguish malignant from benign breast cytology.," *Hum Pathol*, vol. 26, no. 7, pp. 792–796, Jul 1995.

[7] C. Bilgin et al., "Cell-graph mining for breast tissue modeling and classification.," in *IEEE EMBS*, 2007, pp. 5311–5314.

[8] B Hall, W Chen, et al., "A clinically motivated 2-fold framework for quantifying and classifying immunohistochemically stained specimens," in *MICCAI*, 2007, pp. 287–294.

[9] S. Doyle, S. Agner, et al., "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *ISBI*, 2008, pp. 496–499.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.