J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Symposium - Original Research

Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: Comparison to Oncotype DX

Ajay Basavanhally[1], Michael Feldman[2], Natalie Shih[2], Carolyn Mies[2], John Tomaszewski[2], Shridar Ganesan[1,3], Anant Madabhushi[1,3]

[1]Department of Biomedical Engineering, Rutgers University, Piscataway, NJ 08854, [2]Department of Surgical Pathology, Hospital of the University of Pennsylvania, Philadelphia, PA, 19104, [3]The Cancer Institute of New Jersey, New Brunswick, NJ, 08903, USA.

E-mail: *Anant Madabhushi - anantm@rci.rutgers.edu
*Corresponding author

Received: 08 November 11 Accepted: 08 November 11 Published: 19 January 12

This article may be cited as:
Basavanhally A, Feldman M, Shih N, Mies C, Tomaszewski J, Ganesan S, *et al.* Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: Comparison to Oncotype DX. J Pathol Inform 2011;2:S1.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2011/2/2/1/92027

Abstract

In this paper, we attempt to quantify the prognostic information embedded in multi-parametric histologic biopsy images to predict disease aggressiveness in estrogen receptor-positive (ER+) breast cancers (BCa). The novel methodological contribution is in the use of a multi-field-of-view (multi-FOV) framework for integrating image-based information from differently stained histopathology slides. The multi-FOV approach involves a fixed image resolution while simultaneously integrating image descriptors from many FOVs corresponding to different sizes. For each study, the corresponding risk score (high scores reflecting aggressive disease and vice versa), predicted by a molecular assay (Oncotype DX), is available and serves as the surrogate ground truth for long-term patient outcome. Using the risk scores, a trained classifier is used to identify disease aggressiveness for each FOV size. The predictions for each FOV are then combined to yield the final prediction of disease aggressiveness (good, intermediate, or poor outcome). Independent multi-FOV classifiers are constructed for (1) 50 image features describing the spatial arrangement of cancer nuclei (via Voronoi diagram, Delaunay triangulation, and minimum spanning tree graphs) in H and E stained histopathology and (2) one image feature describing the vascular density in CD34 IHC stained histopathology. In a cohort of 29 patients, the multi-FOV classifiers obtained by combining information from the H and E and CD34 IHC stained channels were able to distinguish low- and high-risk patients with an accuracy of 0.91 ± 0.02 and a positive predictive value of 0.94 ± 0.10, suggesting that a purely image-based assay could potentially replace more expensive molecular assays for making disease prognostic predictions.

Key words: Image-based risk score, breast cancer, estrogen receptor positive, computerized prognosis, outcome prediction, multi-variate histology, H and E, CD34 immunohistochemistry

Access this article online

Website: www.jpathinformatics.org

DOI: 10.4103/2153-3539.92027

Quick Response Code:

INTRODUCTION

Predicting disease aggressiveness and outcome for estrogen receptor-positive (ER+) breast cancers (BCa) patients allows for selective employment of therapeutic options. Specifically, identifying which women will benefit from adjuvant chemotherapy over and above the standard hormonal therapy will help limit the use of chemotherapy to more aggressive BCa.[1] Prognosis and treatment in early stage ER+ BCa are often guided by the Oncotype DX genomic assay (Genomic Health, Inc.), which produces a quantitative recurrence score (RS) correlated with likelihood for recurrence.[1] However, recent work has suggested that molecular assays do not provide additional prognostic power over tumor morphology (e.g. grading) and immunohistochemistry (e.g. receptor status).[2,3] Visual analysis of tumor grade in BCa histopathology has shown significant value in predicting patient outcome;[1,4] yet high inter- and intra-clinician variability[5] has limited its use in clinical practice. Conversely, a quantitative, reproducible, and computerized prognostic tool that uses only digitized BCa histopathology slides would be invaluable for predicting prognosis and guiding therapy. Translational advantages of an image-based predictor over its molecular counterparts include a reduced cost per test, shorter time delay between biopsy and treatment, and easier access to patients in developing countries such as India and China.

In microscopy, pathologists implicitly partition an entire histopathology slide into many fields of views (FOVs) and incorporate image features from each FOV to arrive at a diagnostic decision for the entire slide. Previous computerized approaches to whole-slide classification, however, have primarily involved the extraction of image features (for the training of a classifier) from within empirically selected FOVs.[6,7] The empirical selection of FOVs for computerized analysis of BCa histopathology slides presents two main concerns. First, it is a subjective and time-consuming task that requires manual intervention by an expert, an issue that would impede the development of a truly automated classification system. Second, BCa is known to contain intratumoral heterogeneity,[8] whereby different types of cancer (e.g. ductal carcinoma *in situ* and invasive ductal cancer) and levels of malignancy (e.g. low and intermediate grades) may be present in a single histopathology slide. This phenomenon suggests that multiple FOVs of various sizes will be needed depending on the different types of image features extracted and classification tasks performed.

In this paper, we present a multi-FOV framework[9] to perform robust, reproducible classification of entire ER+ BCa histopathology slides based on low, intermediate, and high disease aggressiveness while addressing limitations arising from both manual FOV selection and BCa heterogeneity. The multi-FOV classifier is fundamentally different from traditional multi-scale (i.e. multi-resolution) approaches.[10,11] In image processing, multi-scale schemes are often used to interpret contextual information over different image resolutions.[10] Most multi-scale frameworks, which operate by exposing a single FOV to classifiers at multiple image resolutions, perform well when quantifying large-scale image patterns. Analyzing local object density (or other localized descriptors), however, is more challenging since object density remains invariant to changes in scale (although our visual perception and ability to detect individual objects within the image will vary).

The multi-FOV framework is used to predict BCa outcome by combining image-based features from CD34 immunohistochemistry (IHC) stained and hematoxylin and eosin (H and E) stained histopathology slides. Our multi-FOV scheme uses a fixed image scale and extracts image features at FOVs of different sizes, a highly desirable attribute in heterogeneous images where it is not clear which FOV sizes will contain class discriminatory information. First, a slide is split into FOVs of a fixed size and relevant image features are extracted. A supervised classifier makes an initial class decision for each FOV and the decisions for all FOVs are aggregated to make a single class prediction for the specific FOV size. This procedure is repeated for a variety of FOV sizes, from which the class predictions at all FOV sizes are aggregated to arrive at a single decision for the entire slide. Hence there is no need to empirically determine the optimal FOV size for classification; rather this approach combines class predictions from image features across all FOV sizes. Class predictions are made by two multi-FOV classifiers applied independently to image features describing (1) nuclear architecture and (2) vascular density from the same patient. These class predictions are subsequently combined to form a multi-parametric prediction for the patient.

The CD34 protein is a popular indicator of angiogenesis and, hence, tumor growth and metastasis.[12] Previously, both qualitative[13] and quantitative[14] assessments of CD34 IHC stained slides have characterized IHC staining via "hotspots", i.e. manually selected FOVs; yet, the pitfalls associated with manual FOV selection suggest that hotspot-based predictions may not accurately represent CD34 expression in an entire slide. In this paper, vascular density is automatically extracted from CD34 IHC stained histopathology and used as the sole image-based feature to quantify angiogenic activity. Specifically, color deconvolution,[15] which takes advantage of light-absorbing properties of histological staining, is used to automatically isolate the brown diaminobenzidine (DAB) compound signifying CD34 expression. The extent of DAB staining in a particular FOV is defined as vascular density and used as a feature in the multi-FOV framework.

In prior work, researchers have demonstrated the ability to model and quantify tumor morphology in H and E stained histology through the construction of various graphs.[7,16,17] A graph (e.g. Voronoi diagram) is constructed by drawing edges between relevant tissue structures (e.g. nuclei) in a particular fashion. Graph-based features describing the spatial arrangement of these structures (e.g. Voronoi cell area) are then defined to quantify tissue architecture. In this paper, nuclear architecture is quantified in H and E stained histopathology by constructing three graphs (Voronoi diagram, Delaunay triangulation, and minimum spanning tree) using individual cancer nuclei as nodes. The nuclei are first identified automatically by isolating the blue hematoxylin stain, which preferentially stains nuclear material, via color deconvolution.[15] A total of 50 graph-based features describing the spatial arrangement of cancer nuclei are extracted from each FOV.[7,18] Note that, while both vascular density and nuclear architecture are widely considered to have prognostic value, their biological foundations are very different. This suggests that these two feature classes are complementary and that their combination may produce an improved predictor of patient outcome.

The main contributions of this work are the following:

1. Development and quantitative evaluation of image-based architectural and vascular features for patient outcome prediction in whole-slide ER+ BCa histopathology and
2. Synergistic combination of image-based features from multi-parametric histological studies to achieve an improved prognostic prediction of patient outcome.

While the ideal ground truth for evaluation of prognostic tools like the one described in this work is long-term patient outcome (i.e. survival data), this type of data is very difficult to obtain. In lieu of patient outcome, we utilize Oncotype DX RS as a relative ground truth. Oncotype DX, which produces a quantitative RS between 0 and 100, is a molecular assay that has been clinically validated to predict the likelihood of 10-year distant recurrence and the expected benefit from adjuvant chemotherapy for early-stage, ER+ BCa patients.[1] Specifically, we evaluate the ability of the multi-FOV framework (in conjunction with vascular and architectural features) to distinguish patients with low, intermediate, and high Oncotype DX RS.

The rest of the paper is organized as follows. In the section "multi-FOV framework", we present the theory and methodology behind the multi-FOV framework. It is followed by the sections "experimental design," "results and discussion", and "conclusion."

Multi-FOV Framework

An image scene $\mathcal{C}=(C,g)$ is defined as a 2D set of pixels $c \in C$ with associated intensity function g and class label $y \in \{0,1\}$. For each C and FOV size $\tau \in T$, a grid containing FOVs $D^\tau=\{d_1^\tau, d_2^\tau,.... d_M^\tau\}$ is constructed, where $d_m^\tau \in C, m \in \{1,2,..., M\}$ is a square FOV with edge length of pixels. We define $\mathbf{f}(d_m^\tau)$ as the function that extracts features from each d_m^τ. Grid construction and feature extraction are repeated likewise for each $\tau \in T$.

Theoretical Motivation for Consensus-based Approach

The theory supporting consensus over multiple FOVs demonstrates that a consensus predictor is inherently superior to classification at individual FOV sizes. A consensus predictor over multiple FOV sizes is defined as $\mathbf{H}(\mathbf{D})= E_\tau [H(D^\tau, \tau)]$, where $\mathbf{D}= \{D^\tau: \forall \tau \in T\}$ is the collective data over all FOV sizes, $H(D^\tau, \tau)$ is a meta-classifier (integrated FOV size prediction via individual FOV classifier) for each τ, and E_τ is the expectation of $H(D^\tau, \tau)$ at FOV size $\tau \in T$. The mean squared error of classification at individual FOV sizes is given by $e_\tau = E_\tau [y - H(D^\tau, \tau)]$ and the error of the consensus predictor is given by $e_A = [y - \mathbf{H}(\mathbf{D})]^2$.

Proposition 1: *Given independent classifiers at FOV sizes* $\tau \in T$, $e_\tau \geq e_A$.

Proof:

$$
\begin{aligned}
& e_\tau = \mathrm{E}_\tau[\mathrm{y} - \mathrm{H}(\mathrm{D}^\tau, \tau)]^2 = \mathrm{y}^2 - 2\mathrm{y}\mathrm{E}_\tau[\mathrm{H}(\mathrm{D}^\tau, \tau)]\mathrm{E}_\tau[\mathrm{H}^2(\mathrm{D}^\tau, \tau)] \\
& \text{Since } \mathrm{E}_\tau[H^2(D^\tau, \tau) \geq [H(\mathrm{D}^\tau, \tau)]]^2, \\
& \geq y^2 - 2\mathrm{y}\mathrm{E}_\tau[H(\mathrm{D}^\tau, \tau)] + [\mathrm{E}_\tau[\mathrm{H}(\mathrm{D}^\tau, \tau)]]^2 \\
& \geq \mathrm{y}^2 - 2\mathrm{y}\mathbf{H}(\mathbf{D}) + \mathbf{H}^2(\mathbf{D}) \\
& \geq [\mathrm{y} - \mathbf{H}(\mathbf{D})]^2 \\
& \geq \mathrm{e}_\mathrm{A}
\end{aligned}
$$

Note that the consensus classifier for multiple FOV sizes is similar to Bagging.[19] However, instead of combining weak learners, independent predictors at different FOV sizes (reasonable assumption since different information is discernible at different FOV sizes in heterogeneous images) are used as the "weak" classifiers used to build the "strong" consensus result. To this end, Proposition 1 ensures that the consensus error e_A will always be less than the mean error e_τ of individual FOV size classifiers.

Integrated FOV Size Prediction via Individual FOV Classifier

A pre-trained classifier $\mathbf{h}(d_M^\tau, \mathbf{f}) \in \{0,1\}$ is first used to assign an initial class decision to each individual FOV d^τ with associated features $\mathbf{f}$. Subsequently, decisions are aggregated for all FOVs D^τ to achieve a combined decision $H(D^\tau, \tau)$ at a single FOV size $\tau \in T$.

Algorithm 1: Individual FOV Classifier
Input: Image $\mathcal{C}$. FOV sizes T. Classified $\mathbf{h}(d_m^\tau, \mathbf{f})$.
Output: Aggregate prediction $H(D^\tau, \tau)$ for each FOV size $\tau \in T$

1. **for All** $\tau \in T$ **do**
2. From $\mathcal{C}$, define M, $\tau \times \tau$ FOVs $D^\tau = \{d_1^\tau, d_2^\tau,.... d_M^\tau\}$.
3. Extract features $\mathbf{f}$ from d_m^τ, $\forall m \in M$.
4. Apply classifier $\mathbf{h}(d_m^\tau, \mathbf{f})$ for initial classification of each d_m^τ.

5. Make aggregate predication $H(D^{\tau}, \tau) = \frac{1}{M}\sum_{m=1}^{M} \mathbf{h}(d_m^{\tau}, \mathrm{f})$ over all FOVs D^{τ}.
6. **end for**

Experimental Design

Dataset

The multi-FOV classifier is leveraged for the task of quantifying BCa disease outcome by distinguishing tumors based on Oncotype DX RS. CD34 immunohistochemistry (IHC) and hematoxylin and eosin (H and E) stained histopathology images from 29 patients (9 low RS, 11 intermediate RS, 9 high RS) were digitized via a whole slide scanner at 1 μm/pixel resolution [Table 1]. Each slide is accompanied by (a) annotations from an expert pathologist denoting extent of invasive cancer, and (b) RS values denoting good ($0 < RS < 18$), intermediate ($18 \leq RS \leq 30$), or poor ($30 < RS < 100$) outcome.

Classification Strategy

In each experiment, classification accuracy is computed by comparing the class predictions made by each classifier (multi-FOV and individual FOV sizes) to the ground truth, i.e. good, intermediate, or poor outcome, delineated by the Oncotype DX RS for each slide. To mitigate the bias associated with the selection of training and testing samples, each classifier is evaluated via a three-fold cross-validation scheme.[20] For each experiment, the dataset is first divided randomly into three subsets of equal size. FOVs from two subsets are used to train the preliminary classifier **h** (via a random forest classifier[21]) and FOVs from the remaining subset is used for evaluation. The training and testing subsets are rotated so that each slide is evaluated once. The entire cross-validation procedure is repeated 10 times to produce mean and standard deviation classification accuracy values.

Table 1: A summary of the data cohort comprising 29 ER+ BCa patients used in this paper. For each class, the number of patients and the range of oncotype DX RS values are given

Patient class	# Patients	RS range
Low	9	0-17
Intermediate	11	18-30
High	9	31-100

Experiment 1: Multi-FOV Classifier for Quantifying Vascular Density in CD34 IHC Stained Histopathology

The density of vascular formation is calculated from CD34 IHC stained histology images [Figures 1a and e] to capture the degree of angiogenesis via the following steps.

Step 1: Color deconvolution[15] splits the image into channels representing DAB (i.e. CD34 expression) and hematoxylin [Figures 1b, c, f, and g].

Step 2: The DAB channel is thresholded to produce a set of brown pixels corresponding to angiogenic vessels [Figures 1d and h].

Step 3: Global vascular density is defined as fraction of brown pixels within region of cancer extent from entire slide.

Step 4: Local vascular density is defined as fraction of

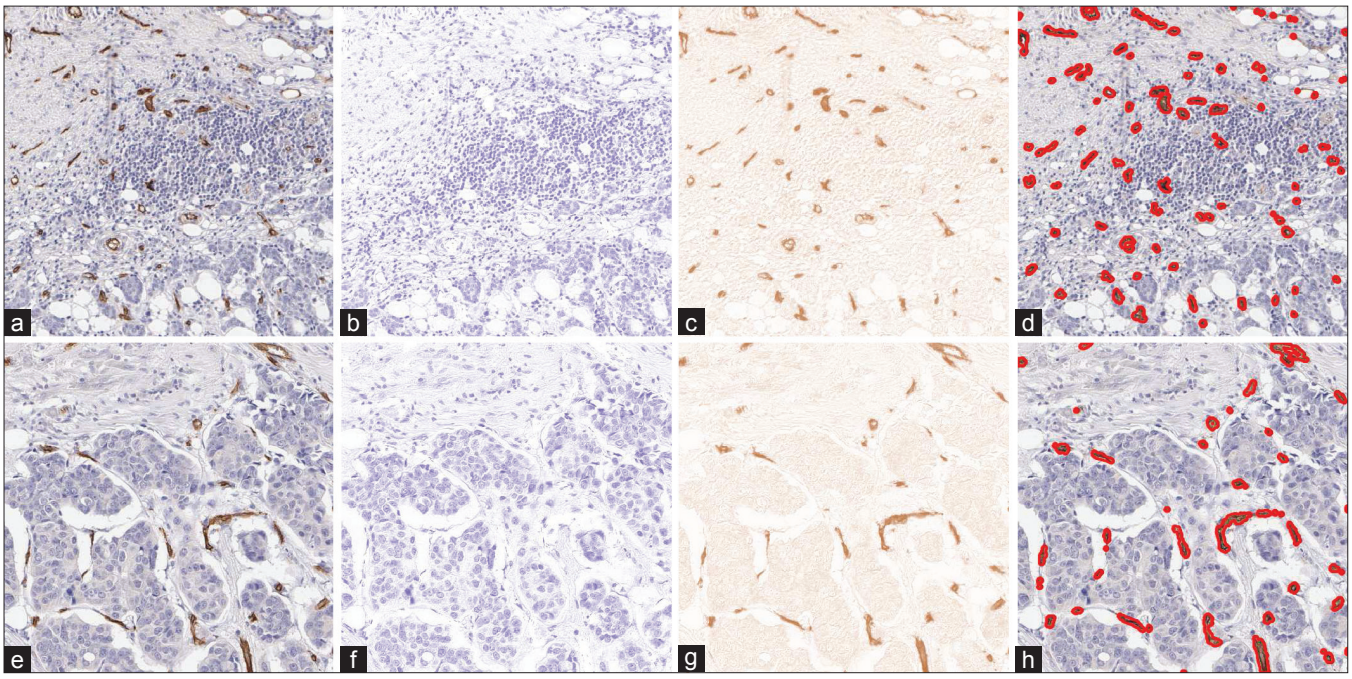

Figure 1: (a), (e) CD34 IHC stained images are separated into (b), (f) hematoxylin and (c), (g) DAB channels via color deconvolution. The DAB channel is thresholded to isolate (d), (h) segmented regions expressing the CD34 protein.

brown pixels from a smaller FOV (of size $\tau \in T$) within region of cancer extent. A range of FOV sizes (T= {250, 500, 1000} pixels) was considered in this work.

Experiment 2: Multi-FOV Classifier for Quantifying Tissue Morphology on H and E Stained Histopathology

The variation in the spatial arrangement of cancer nuclei is quantified to capture the BCa tissue structure in an image via the following steps.

Step 1: Color deconvolution is used to separate image into channels representing hematoxylin and eosin stains [Figures 2b and c].

Step 2: Since hematoxylin stains nuclear material, individual cancer nuclei are detected by thresholding the hematoxylin channel [Figure 2d].

Step 3: Cancer nuclei are used as vertices for construction of Voronoi diagram [Figures 2e, h, and k], Delaunay triangulation [Figures 2f, i and l], and minimum spanning

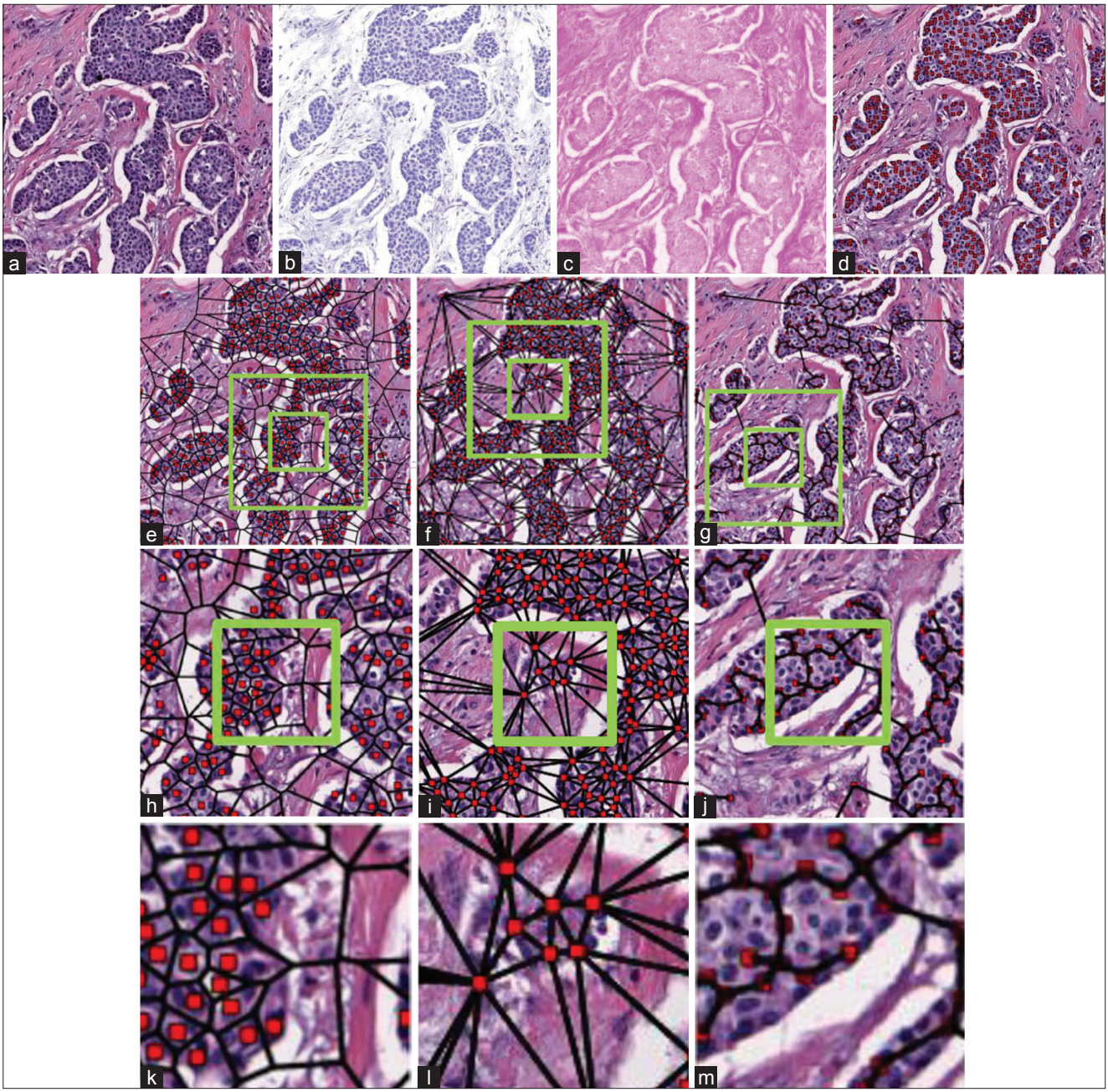

Figure 2: (a) Hematoxylin and eosin stained images are separated into (b) hematoxylin and (c) eosin channels via color deconvolution. The hematoxylin channel is thresholded to detect (d) centroids of individual cancer nuclei, which are used to construct (e), (h), (k) Voronoi diagram, (f), (i), (l) Delaunay triangulation, and (g), (j), (m) minimum spanning tree. The graphs are subsequently used to extract 50 features describing nuclear architecture. Note the variations in nuclear arrangement when exposing (e)-(g) large, (h)-(j) medium, and (k)-(m) small FOVs

tree [Figures 2g, j and m], from which 50 architectural features [Table 2] are extracted for each image.

Step 4: Architectural features are calculated for individual FOVs within regions of cancer extent. A wide range of FOV sizes ($\tau \in \{250, 500, 1000, 2000\}$ pixels) was considered in this paper.

Experiment 3: Multi-Parametric Classifier for Combining Features from H and E and IHC Stained Histopathology

Since vascular density and nuclear architecture utilize distinct biological phenomena to describe disease aggressiveness, we expect a combination of the two data channels to produce improved classification.

Step 1: Perform Experiment 1 and save resulting class decision $\mathbf{H}_{IHC} \in \{0,1\}$ made for each slide.

Step 2: Perform Experiment 2 and save resulting class decision $\mathbf{H}_{HE} \in \{0,1\}$ made for each slide.

Step 3: Generate a decision-level prediction $\hat{\mathbf{H}} = \mathbf{H}_{IHC} \wedge \mathbf{H}_{HE} \in \{0,1\}$ based on the independent class predictions made from the H and E and IHC stained slides. Note that the $\wedge$ operation is defined as "logical AND", whereby $\hat{\mathbf{H}} = 1$ if both $\mathbf{H}_{IHC}=1$ and $\mathbf{H}_{HE}=1$. Conversely, $\hat{\mathbf{H}} = 0$ if either $\mathbf{H}_{IHC}=0$ or $\mathbf{H}_{HE}=0$.

Table 2: A breakdown of the 50 architectural features extracted from the Voronoi diagram, Delaunay triangulation, and minimum spanning tree graphs

Type	#	Description
Voronoi diagram	13	Polygon area, perimeter, chord length
Delaunay triangulation	8	Triangle side length, area
Minimum spanning tree	4	Edge length
Nearest neighbor	25	Density, distance to nearest nuclei

RESULTS AND DISCUSSION

Experiment 1: Vascular Density in CD34 IHC Stained Histopathology

The ability of the multi-FOV classifier to outperform classification at individual FOV sizes is borne out by the local vascular density [Figure 3], which is able to distinguish entire CD34 IHC stained slides with good vs. poor, good vs. intermediate, and intermediate vs. poor Oncotype DX RS values with classification accuracies of 0.82 ± 0.04, 0.75 ± 0.06, 0.86 ± 0.04, respectively, and positive predictive values (PPV) of 0.82 ± 0.06, 0.76 ± 0.06, 0.87 ± 0.06, respectively. The theoretical justification for the multi-FOV framework suggests that a multi-FOV classifier will outperform the majority of classifiers for individual FOV sizes (theoretical motivation for consensus-based approach). This concept is validated empirically in Experiment 1, where multi-FOV classifiers perform as well as (and usually better than) individual FOV sizes in both classification accuracy and PPV [Figure 3].

In addition, global vascular density produces corresponding classification accuracies of 0.60 ± 0.08, 0.40 ± 0.11, 0.46 ± 0.07 and PPV of 0.82 ± 0.09, 0.76 ± 0.07, and 0.72 ± 0.11, respectively [Figure 3]. Experiment 1 demonstrates that the multi-FOV classifier (in conjunction with local vascular density) consistently outperforms whole-slide global vascular density in discriminating ER+ BCa with low, intermediate, and high Oncotype DX RS [Figure 3]. The superior performance of the multi-FOV classifier is likely due to its ability to capture local variations in vascular density and robustness to intra-slide heterogeneity. The multi-FOV framework also has an added benefit in that it readily accommodates the inclusion of complimentary structural information (e.g. nuclear architecture).

Experiment 2: Tissue Morphology on H and E Stained Histopathology

Figure 4 shows that the architectural features (in

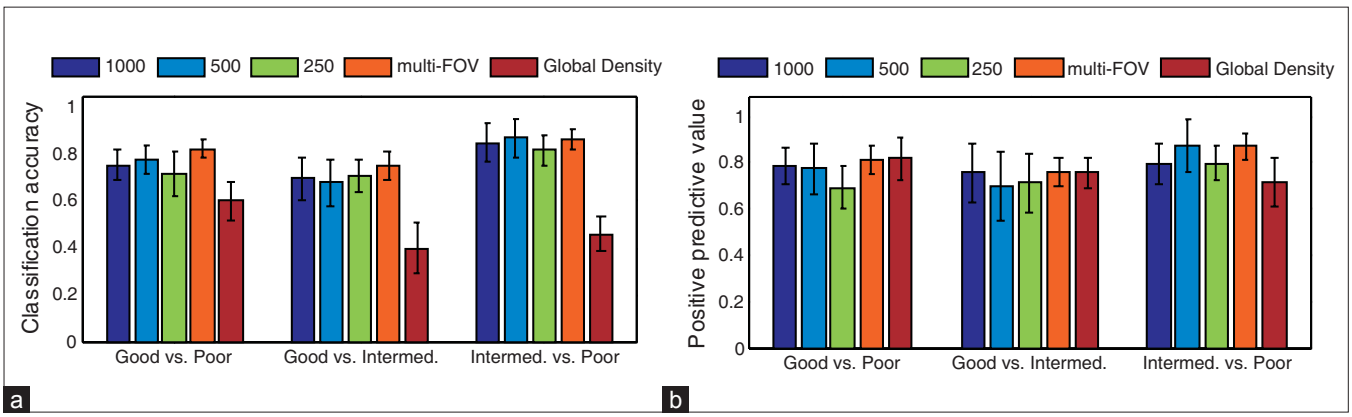

Figure 3: (a) Classification accuracy and (b) positive predictive values for the multi-FOV framework using local vascular density from 29 CD34 IHC stained histopathology slides over 10 trials of three-fold cross-validation. Note that the bar colors represent different FOV sizes as indicated. For comparison, global vascular density was also calculated directly from each slide.

conjunction with the multi-FOV classifier) are able to discriminate H and E stained slides with good vs. poor, good vs. intermediate, and intermediate vs. poor Oncotype DX RS at classification accuracies of 0.91 ± 0.04, 0.72 ± 0.06, 0.71 ± 0.11, respectively, and positive predictive values of 0.92 ± 0.06, 0.74 ± 0.12, 0.68 ± 0.11, respectively. The argument in favor of the multi-FOV classifier is even stronger in Experiment 2 (compared to Experiment 1), where it shows significantly increased performance over individual FOV sizes [Figure 4].

Experiment 3: Combined Features on H and E and IHC Stained Histopathology

Performing a decision-level combination of vascular density and nuclear architecture produces classification accuracies of 0.91 ± 0.02, 0.76 ± 0.05, 0.83 ± 0.08 and PPV of 0.94 ± 0.10, 0.85 ± 0.11, 0.92 ± 0.13, for distinguishing good vs. poor, good vs. intermediate, and intermediate vs. poor RS values, respectively [Table 3]. The fact that vascular density and nuclear architecture exploit such disparate aspects of cancer biology (i.e. angiogenesis and tissue morphology, respectively) suggests that the two feature classes are complimentary and integration will yield improved classification. Experiment 3 shows that a decision-level combination of the two feature sets maintains high levels of classification accuracy while improving positive predictive values [Table 3] over the corresponding multi-FOV classifiers from Experiments 1 and 2 [Figures 3 and 4].

Significance of Multi-FOV Classifier Compared to Individual FOV Sizes

To confirm the significance of our results for the multi-FOV classifier, two-sample *t*-tests were performed with alternative hypotheses asserting that the multi-FOV classifier outperforms individual FOV sizes in terms of classification accuracy [Table 4]. The Bonferroni correction for multiple comparisons[22] has been applied to all *P*-values in Table 4. For good vs. poor outcome, we were able to reject the null hypothesis for all FOV sizes with $P < 0.05$ for vascular density and for 3 of 4 FOV sizes for nuclear architecture. Similarly, the null hypothesis was rejected with $P < 0.05$ for 3 of 4 FOV sizes when comparing good vs. intermediate outcomes and with $P < 0.10$ for 2 of 4 FOV sizes when comparing

Table 3: Classification accuracies and positive predictive values for comparing good, intermediate, and poor Oncotype DX scores via the multi-FOV framework using a combination of vascular density and architectural features over 10 trials of three-fold cross-validation

	Good vs. poor	Good vs. intermed	Intermed. vs. poor
Accuracy	0.91 ± 0.022	0.76 ± 0.051	0.83 ± 0.076
PPV	0.94 ± 0.10	0.85 ± 0.11	0.92 ± 0.13

PPV - Positive predictive values

Table 4: Bonferroni-corrected *P*-values produced by two-sided *t*-tests with a null hypothesis that classification results from the multi-FOV approach are equivalent to results from individual FOV sizes from both IHC stained and H and E stained histopathology slides. The alternative hypothesis asserts that the multi-FOV classifier performs better than individual FOV sizes

FOV size	Good vs. poor	Good vs. intermed.	Intermed. vs. poor
Vascular density in IHC stained histopathology			
1000	0.0288	0.2250	0.9042
500	0.0123	0.1011	1.0000
250	0.0129	0.2313	0.1101
Nuclear architecture in H and E stained histopathology			
2000	0.0570	0.0666	1.0000
1000	0.02657	0.0066	0.1575
500	0.0429	0.0003	0.0657
250	<0.0001	<0.0001	0.0027

IHC: Immunohistochemistry, H and E: Hematoxylin and eosin, FOV: Field-of-view

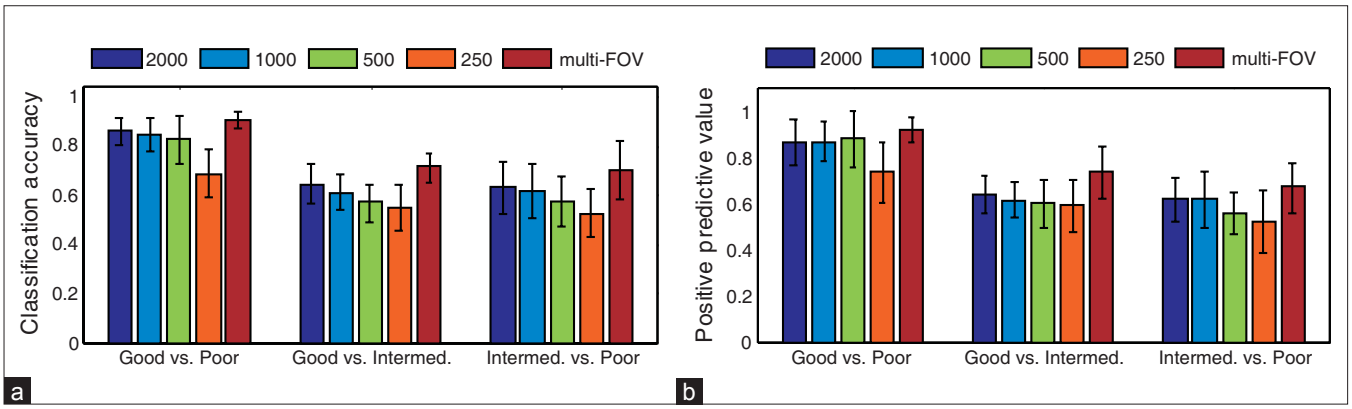

Figure 4: (a) Classification accuracy and (b) positive predictive values for the multi-FOV framework using architectural features from 29 H and E stained histopathology slides over 10 trials of three-fold cross-validation. Note that the bar colors represent different FOV sizes as indicated

intermediate vs. poor outcomes for nuclear architecture. The results also suggest that vascular features in conjunction with the multi-FOV approach do not appear to offer any significant improvement in distinguishing good vs. intermediate and intermediate vs. poor outcomes, suggesting the need to identify higher order features that more accurately quantify vascular formation in IHC stained ER+ BCa histopathology.

Understanding Misclassified Patients in the Context of Oncotype DX as a Relative Ground Truth

It is particularly important to note that the Oncotype DX RS values used as class labels in this work represent a relative ground truth due to their demonstrated correlation with patient outcome.[1] The classification results in this paper do not reflect directly upon the ability of the multi-FOV framework to predict patient outcome; instead, they reveal the level of concordance between the multi-FOV framework and Oncotype DX RS values. Specifically, our results demonstrate the difficulty in using relative ground truth to evaluate BCa prognosis due to the high degree of uncertainty in the "intermediate" class. This problem is illustrated in Figure 5, where only a few of the misclassified slides with intermediate RS fall squarely within the intermediate class (RS 22-29). Meanwhile, the majority of misclassifications lies on the lower end of the intermediate class (RS 18-21) and may actually represent patients with good prognosis.

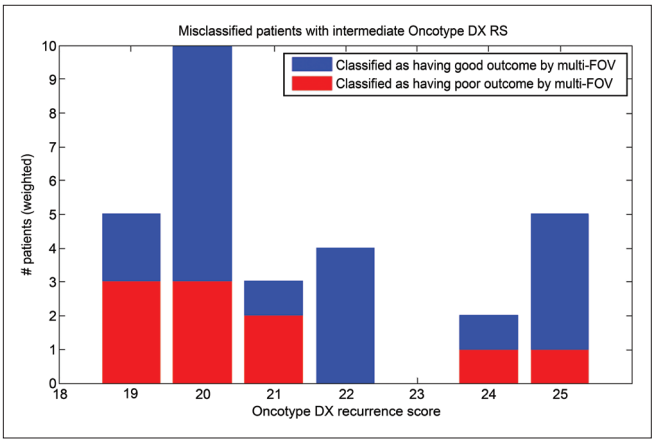

Figure 5: A histogram of patients with intermediate Oncotype DX RS that were found by the multi-FOV classifier to have either good (shown in blue) or poor (shown in red) prognosis

CONCLUSIONS

We have presented a computerized system for predicting disease outcome in ER+ BCa using only image-based features from multi-parametric histopathology images. From a translational perspective, this work illustrates the possibility of a low cost, quantitative, image-based risk predictor that performs on par with expensive gene expression assays (e.g. Oncotype DX) in terms of predicting outcome in ER+ BCa patients. The main contributions of this work are the following:

A multi-FOV framework that integrates vascular and structural information from multiple FOVs at different sizes in ER+ BCa histopathology, and

Quantitative evaluation of vascular density from CD34 IHC stained slides as a prognostic indicator for ER+ BCa via comparison to Oncotype DX RS.

Utilizing image features that quantify angiogenesis and nuclear architecture, we demonstrated the ability of the multi-FOV classifier to discriminate between cancers with good and poor RS, good and intermediate RS, and intermediate and poor RS with accuracies of 0.91, 0.76, and 0.83, respectively. We also establish the importance of using localized FOV-based feature extraction instead of a global approach for classifying heterogeneous histopathology slides. For nuclear architecture, the advantage of the multi-FOV classifier (over classification at individual FOV sizes) is significant in most cases. However, the results for vascular density suggest that additional patients should be included to confirm the superiority of the multi-FOV approach.

A closer look at studies misclassified by the multi-parametric multi-FOV classifier shows that a large number of misclassified patients with intermediate RS are actually distributed closely along the border between low and intermediate RS values. This distinction is particularly important because recent studies comparing Oncotype DX with another molecular assay, PAM50, have suggested that a number of patients assigned intermediate RS may actually have a low risk of recurrence and, hence, do not require adjuvant chemotherapy.[23]

In the current implementation, the entire algorithm (including object detection/segmentation, feature extraction, and classification) requires approximately 2 h per slide on a 2.83 GHz processor. A key advantage of the multi-FOV approach, however, is that computational time can be significantly reduced via parallelization, especially with the rapid proliferation of multi-core CPU and GPU computing. For instance, individual FOV sizes (from all histological channels) can be processed in parallel since the class predictions they make are independent of each other.

Future work will focus on incorporating additional feature classes and a larger variety of histopathology studies. In the long term, we aim to perform a direct comparison against Oncotype DX in terms of predicting patient outcome.

DISCLOSURE/CONFLICT OF INTEREST

AM is a majority stockholder in Ibris Inc. and vascuVis Inc.

ACKNOWLEDGMENTS

This work was made possible by the Wallace H. Coulter Foundation, New Jersey Commission on Cancer Research, National Cancer

Institute (R01CA136535-01, R01CA14077201, R21CA12718601, and R03CA143991-01), and The Cancer Institute of New Jersey.

REFERENCES

1. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-26.
2. Flanagan MB, Dabbs DJ, Brufsky AM, Beriwal S, Bhargava R. Histopathologic variables predict oncotype dx recurrence score. Mod Pathol 2008;21:1255-61.
3. Weigelt B, Reis-Filho JS. Molecular profiling currently offers no more than tumour morphology and basic immunohistochemistry. Breast Cancer Res 2010;12(Suppl 4):S5.
4. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957;11:359-77.
5. Meyer JS, Alvarez C, Milikowski C, Olson N, Russo I, Russo J, *et al.* Breast carcinoma malignancy grading by bloom-richardson system vs proliferation index: Reproducibility of grade and advantages of proliferation index. Mod Pathol 2005;18:1067-78.
6. Sertel O, Kong J, Shimada H, Catalyurek UV, Saltz JH, Gurcan MN. Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. Pattern Recognit 2009;42:1093-103.
7. Basavanhally A, Xu J, Madabhushi A, Ganesan S. Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with oncotype dx assay. In: Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro ISBI '09. 2009. p. 851-4.
8. Torres L, Ribeiro FR, Pandis N, Andersen JA, Heim S, Teixeira MR. Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. Breast Cancer Res Treat 2007;102:143-55.
9. Basavanhally A, Ganesan S, Shih N, Mies C, Feldman M, Tomaszewski J, *et al.* A boosted classifier for integrating multiple fields of view: Breast cancer grading in histopathology. In: Proc. IEEE Int Biomedical Imaging: From Nano to Macro Symp; 2011. p. 125-8.
10. Doyle S, Rodriguez C, Madabhushi A, Tomaszeweski J, Feldman M. Detecting prostatic adenocarcinoma from digitized histology using a multi-scale hierarchical classification approach. In: Proc. 28th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society EMBS '06. 2006. p. 4759-62.
11. Gurcan MN, Kong J, Sertel O, Cambazoglu BB, Saltz J, Catalyurek U. Computerized pathological image analysis for neuroblastoma prognosis. AMIA Annu Symp Proc 2007:304-8.
12. Weidner N, Semple JP, Welch WR, Folkman J. Tumor angiogenesis and metastasis–correlation in invasive breast carcinoma. N Engl J Med 1991;324:1-8.
13. Nassif AE, Filho RT. Immunohistochemistry expression of tumor markers cd34 and p27 as a prognostic factor of clinically localized prostate adenocarcinoma after radical prostatectomy. Rev Col Bras Cir 2010;37:338-44.
14. Erovic BM, Neuchrist C, Berger U, El-Rabadi K, Burian M. Quantitation of microvessel density in squamous cell carcinoma of the head and neck by computer-aided image analysis. Wien Klin Wochenschr 2005;117:53-7.
15. Ruifrok AC, Johnston DA: Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol 2001;23:291-9.
16. Sudbø J, Bankfalvi A, Bryne M, Marcelpoil R, Boysen M, Piffko J, *et al.* Prognostic value of graph theory-based tissue architecture analysis in carcinomas of the tongue. Lab Invest 2000;80:1881-9.
17. Doyle S, Hwang M, Shah K, Madabhushi A, Feldman M, Tomaszeweski J. Automated grading of prostate cancer using architectural and textural image features. In: Proc. 4th IEEE Int. Symp. Biomedical Imaging: From Nano to Macro ISBI 2007. 2007. p. 1284-7.
18. Basavanhally AN, Ganesan S, Agner S, Monaco JP, Feldman MD, Tomaszewski JE, *et al.* Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology. IEEE Trans Biomed Eng 2010;57:642-53.
19. Breiman L. Bagging predictors. Mach Learn 1996;24:123-40.
20. Duda R, Hart P, Stork D. Pattern Classification. Hoboken, New Jersey: John Wiley and Son; 2001.
21. Breiman L. Random forests. Mach Learn 2001;45:5-32.
22. Bonferroni CE. Il calcolo delle assicurazioni su gruppi di teste. In: Studi in Onore del Professore Salvatore Ortu Carboni. Rome; 1935. p. 13-60.
23. Kelly C, Bernard P, Krishnamurthy S, Ebbert M, Bastien R, Boucher K, *et al.* Concordance between the 21 gene assay (oncotypedx) and the pam50 intrinsic classifier for prognosis in early stage er-positive breast cancer. Ann Oncol 2011;22(Suppl 2):ii44-7.