

## Research Article

# Novel Image Analysis Approach Quantifies Morphological Characteristics of 3D Breast Culture Acini with Varying Metastatic Potentials

Lindsey McKeen Polizzotti,<sup>1</sup> Basak Oztan,<sup>2</sup> Chris S. Bjornsson,<sup>1</sup>  
Katherine R. Shubert,<sup>1</sup> Bülent Yener,<sup>2</sup> and George E. Plopper<sup>1</sup>

<sup>1</sup>Department of Biology, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

<sup>2</sup>Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Correspondence should be addressed to Basak Oztan, oztanb@cs.rpi.edu

Received 1 November 2011; Revised 21 February 2012; Accepted 22 February 2012

Academic Editor: James Sherley

Copyright © 2012 Lindsey McKeen Polizzotti et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Prognosis of breast cancer is primarily predicted by the histological grading of the tumor, where pathologists manually evaluate microscopic characteristics of the tissue. This labor intensive process suffers from intra- and inter-observer variations; thus, computer-aided systems that accomplish this assessment automatically are in high demand. We address this by developing an image analysis framework for the automated grading of breast cancer in *in vitro* three-dimensional breast epithelial acini through the characterization of acinar structure morphology. A set of statistically significant features for the characterization of acini morphology are exploited for the automated grading of six (MCF10 series) cell line cultures mimicking three grades of breast cancer along the metastatic cascade. In addition to capturing both expected and visually differentiable changes, we quantify subtle differences that pose a challenge to assess through microscopic inspection. Our method achieves 89.0% accuracy in grading the acinar structures as nonmalignant, noninvasive carcinoma, and invasive carcinoma grades. We further demonstrate that the proposed methodology can be successfully applied for the grading of *in vivo* tissue samples albeit with additional constraints. These results indicate that the proposed features can be used to describe the relationship between the acini morphology and cellular function along the metastatic cascade.

## 1. Introduction

Breast cancer is the second most common cancer in women and is also the second leading cause of cancer-related death in women [1]. In its most common form, the tumor arises from the epithelial cells in the breast tissue. Histological grading systems are commonly used to predict the prognosis of tumors. The most frequently used tumor grading system for breast cancer is the modified Scarrf-Bloom-Richardson method [2] where pathologists analyze the rate of cell division, percentage of tumor forming ducts, and the uniformity of cell nuclei to determine the cancer grade in H&E stained biopsies. While precancerous (or lower-grade) tumors tend to grow slowly and are less likely to spread, invasive (or higher-grade) tumors typically gain the

ability to proliferate and spread rapidly. Subjectivity and variability of the results affect the accuracy of prognosis and subsequent patient treatment. A recent study indicates that the rate of misdiagnosis of breast cancer varies widely between clinicians and is nearly 40% in some cases [3]. Thus, there is an unmet need for robust methods that reduce the variability and subjectivity in the grading of breast tumors and lesions. Development of quantitative tools for image analysis and classification is rapidly expanding fields that constitute a great potential for improving diagnostic accuracy [4, 5].

In this paper, a method for automated grading of breast cancer in three-dimensional (3D) epithelial cell cultures is presented. *In vitro* epithelial breast cells cultured in laminin rich extracellular matrix form *acinar*-like structures that

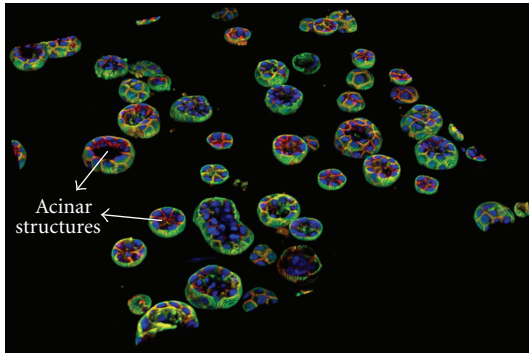


FIGURE 1: Three-dimensional view of a typical nonmalignant cell line culture. Acinar structures are surrounded by the ECM proteins.

both morphologically and structurally resemble *in vivo* acini of breast glands and lobules [6]. Therefore, these culture systems constitute suitable and controllable environments for breast cancer research [7–9].

Figure 1 shows a 3D view of a typical nonmalignant breast culture that comprises several acini that are surrounded by extracellular matrix (ECM) proteins. Figure 2 shows an enlarged view of a cross section from a typical nonmalignant acinus. Acinar structures in healthy/nonmalignant cultures include polarized luminal epithelial cells. (A polarized cell has specialized proteins localized to specific cell membranes such as the basal, lateral, or apical side) and a hollow lumen. The lateral membranes of neighboring cells are in close proximity due to cellular junctions and adhesion proteins. The basal side of cells contacts the surrounding ECM proteins, while the apical sides of cells face the hollow lumen. Malignant cancers result in loss of cell polarity that induces changes in the morphology of acinar structures. In this paper, we investigate morphological characteristics of mammary acinar structures in nonmalignant, noninvasive carcinoma, and invasive carcinoma cancer grades in six MCF10 series of cell lines grown in 3D cultures. We propose novel features to characterize these changes along the metastatic cascade and exploit them in a supervised machine learning setting for the automated grading of breast cancer. Proposed features capture not only similar factors as the Scarff-Bloom-Richardson grading system but also additional subvisual changes observed in breast cancer progression in a quantitative manner to reduce variability. As shown by the grading accuracies, the proposed features efficiently capture the differences caused by the metastatic progression of the cancer.

Previous work on this problem includes examining the change in the morphological characteristics of nontumorigenic MCF10A epithelial acini over time and exploiting them to model the growth of culture over time. Chang et al. examined the elongation of the MCF10A acini at 6, 12, and 96 hours after a particular treatment [10]. In a more predictive setting, Rejniak et al. used the number of cells per acini, proliferation, and apoptosis rates to computationally model the MCF10A epithelial acini growth using fluid-dynamics-based elasticity models [11]. In addition to these

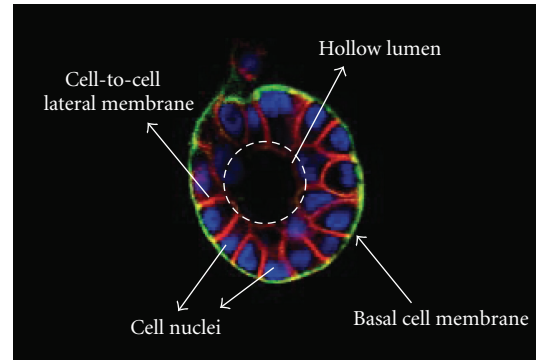


FIGURE 2: Example of a typical nonmalignant acinar structure cross section. The acinar structure includes polarized epithelial cells and hollow lumen. Basal sides of the cells are surrounded by the ECM proteins and apical sides face the hollow lumen.

features, Tang et al. utilized features like acinus volume, density, sphericity, and epithelial thickness to investigate the relationship between acinus morphology and apoptosis, proliferation, and polarization [12]. Specifically, they built a computational model that can predict the growth of acini over a 12-day period. In addition, graph theoretical tools [13–15] were exploited to highlight the structural organization of the cells within the malignant tissues. Our method is different than these characterization efforts in that the grading of cancer is achieved over a richer and more discriminative set of small-scale (local) morphological features that are statistically significant. In addition, the features proposed here closely mimic the features that current pathological grading systems utilize. The presented work builds upon and extends our prior work in this area that introduced the underlying framework [16]. In this work, we provide extended details of our methodology and also present analysis that tests the performance of different supervised machine learning methods and investigates the discriminative influence of the proposed features. Furthermore, the overall grading accuracy is significantly improved by eliminating the acini that are in the preliminary stages of their formations from our analysis. Finally, we perform a preliminary study on the grading of *in vivo* tissue section using our framework and demonstrate that the proposed features can also be used on *in vivo* tissue slides albeit with additional constraints on the preparation of the tissue for our analysis.

## 2. Materials and Methods

**2.1. Cell Culture.** Cells were grown on tissue culture-treated plastic T75 flasks and incubated at 37°C in a humidified atmosphere containing 95% air and 5% carbon dioxide in the manufacturer suggested media. Once 80% confluent, cells were split using 0.25% trypsin with 0.2 g/L EDTA and seeded into a new flask or under experimental conditions. Six MCF10 series cell lines that represent three grades of breast cancer along the metastatic cascade were grown. The cell lines used in this experimentation were

MCF10A (10A), MCF10AT1 (AT), MCF10AT1K.c12 (KCL), MCF10DCIS.com (DCIS), MCF10CA1h.c12 (CA1H), and MCF10CA1a.c11 (CA1A). The DCIS, CA1H, and CA1A cell lines were grown in DMEM/F12 with 5% horse serum and 1% fungizone/penicillin/streptomycin. 10A, AT, and KCL cells were grown in the same base media with additional factors including 20 ng/mL epidermal growth factor, 0.5 mg/mL hydrocortisone, 100 ng/mL cholera toxin, and 10  $\mu$ g/mL of insulin. The 10A cell line was obtained from the American Type Culture Collection (ATCC), AT, KCL, CA1A, and CA1H cell lines were obtained from the Barbara Ann Karmanos Cancer Institute at Wayne State University, and the DCIS cell line was purchased from Asterand Inc.

The advantages of the MCF10 series of cell lines include their derivation from a single biopsy and subsequent mutations forming cell lines of ranging metastatic ability. These cell lines were acquired from a tissue sample diagnosed as noncancerous fibrocystic disease [17, 18]. The biopsy cells, MCF-10M, were cultured and spontaneously gave origin to two immortal sublines. One of these cell lines was named 10A for its adherent ability. 10A cells were nontumorigenic in mice xenografts; however, the first derivative cell line was transfected with oncogene T-24 H-Ras to promote expression of a constitutively active form of Ras resulting in precancerous lesion formations. These cells were then serially passaged for six months and the derived cell line was named MCF10AneoT [17]. AT cells were derived from a 100-day-old MCF10AneoT lesion that formed a squamous carcinoma but failed to produce carcinomas when injected back into mice [19]. KCL cell line was obtained from a 367 day tumor xenograft of the AT cell line. KCL cells were injected into a mouse, where a tumor was formed and isolated after 292 days. Isolated cells derived from this tumor were cultured and yielded the DCIS cell line, which formed ductal carcinoma in situ tumors when injected into mice [18]. Additional cells from the tumors which were derived from the KCL cell line were implanted within a mouse and yielded the MCF10CA cell lines; two of these were included in this study: the CA1H and CA1A [18]. CA1H cells were found to develop invasive carcinoma in mice while the CA1A cells were of higher-grade malignancy and able to metastasize to the lungs. After evaluating the metastatic potentials of these six cell lines, we considered the 10A and AT cell lines as nonmalignant, KCL and DCIS cell lines as noninvasive carcinoma, and CA1H and CA1A cell lines as invasive carcinoma that constituted the three grades of breast cancer we considered in this study.

**2.1.1. 3D Culture System.** Cells were suspended at a concentration of one million cells per milliliter in Matrigel (laminin rich extra cellular matrix) on ice [20]. The gel-cell solution was seeded at 30  $\mu$ L per glass bottom 96 well for 14 days in their respective media. The Matrigel-cell solution was allowed to solidify for 30 minutes at 37°C before the media was added and was changed every 2-3 days thereafter.

**2.2. Immunocytochemistry Staining for Imaging.** Following 14 days in culture, samples were washed once with phosphate buffer solution (PBS) and then fixed with 3% paraformaldehyde at room temperature for 30 minutes. Next, samples

were rinsed with PBS and treated with cell blocking solution (PBS with 1% bovine serum albumin (BSA) and 0.25% Tween20) for one hour at 4°C. Samples were then permeabilized with 0.25% TritonX100-PBS solution for ten minutes.

After the permabilization, samples were washed with PBS and then treated with the primary antibodies at their determined working dilution in PBS at 4°C overnight. Next, the samples were washed with PBS three times for 30 minutes at room temperature. The secondary antibodies were added to the samples in PBS at 4°C overnight. Finally, samples were treated with DAPI for thirty minutes followed by three 30-minute PBS washes and stored at 4°C in PBS prior to imaging.

**2.2.1. Antibodies and Dyes.** Integrin  $\alpha$ 3 antibody (mouse monoclonal) (Abcam: ab11767) was used at a dilution of 1:40 for 3D confocal fluorescent imaging. Secondary antibody, Alexa Fluor 568 goat anti-mouse IgG highly cross absorbed (Invitrogen A-11031) at a dilution of 1:200, was used to visualize the localization of integrin  $\alpha$ 3. Integrin  $\alpha$ 6 FITC-conjugated antibody (Rat monoclonal) (Abcam: ab21259) was used at a dilution of 1:25 for 3D confocal fluorescent imaging during the secondary antibody step previously described. Nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI; Molecular Probes) for 30 minutes at room temperature. 3D volumes were obtained using a Zeiss LSM510 META confocal microscope with a 40x water immersion objective as z-stacks. Proper color filters were used to capture the red, green, and blue fluorescent signals. Images were captured using a multitrack setting where fluorescent channels were acquired sequentially to avoid fluorescence crosstalk. Thickness of the z-stacks ranged from 10 to 40  $\mu$ m (1  $\mu$ m slices) with an initial depth of at least 10  $\mu$ m. Slices had 512  $\times$  512 pixels (320  $\mu$ m  $\times$  320  $\mu$ m) cross-section area. In order to include larger volumes of the cultures in our analysis, we captured four tiles (in 2  $\times$  2 formation) with approximately 20% overlap and stitched these tiles using the 3D image registration technique proposed by Preibisch et al. [21] implemented in ImageJ [22]. A total of five stitched images were captured for each of the six cell lines.

**2.3. Segmentation of Acinar Structures.** Segmentation of the acinar structures was accomplished by the well-known watershed segmentation [23]. Watershed segmentation exploits the morphological characteristics of the regions of interest and is particularly useful for the segmentation of adjoined objects. The process starts with the identification of the acinar structures. In this study, cell nuclei were marked with blue fluorescent marker DAPI, cell-to-cell borders were identified by the red fluorescent marker that shows the localization of integrin  $\alpha$ 3, and basal sides of the cell membranes were identified by the green fluorescent marker that shows the localization of integrin  $\alpha$ 6. As these components are observed in the individual color channels of the captured images and, we used the combination of the three color channels to identify the acinar structures.

First, the color channels of the image were individually binarized. Image values were separated into foreground and

background classes, where the foreground class represented the stained components and the background class included the combination of the gel medium and extracellular proteins. We employed a local adaptation to Otsu's well-known global thresholding algorithm [24] to binarize the color channels. In each slice along the depth direction, we divided the image into rectangular blocks along the horizontal and vertical directions and binarized them separately. This approach handles the spatial variations in the foreground-background contrast better than global thresholding. The noise produced due to the binarization of local regions that contain hardly any information was eliminated by using the edge-based noise elimination that cleaned the regions that did not contain any edge from the resulting binary image [25]. Next, the resulting binary color channels were superposed to obtain a single monochrome binary image by logical OR operation. Enclosing acinar structures were identified by applying morphological *close* operation followed by a morphological *fill-hole* operation [26] to the resulting binary image. Finally, 3D watershed segmentation was applied to label the individual acinar structures. For this purpose, we first obtained a topographic interpretation of the resulting binary image by taking its Euclidean distance transform where the shortest distance to the nearest background pixel for each pixel was measured [27]. The resulting transformed image was then inverted, while forcing the values for the background class pixels to  $-\infty$ , to construct the catchment basins and the watershed segmentation method was finally applied to construct watershed lines to divide these basins and identify the unique acinar structures. Acinar structures with less than four nuclei were considered to exhibit reduced ability to form polarized acinar structures and excluded from our analysis. Note that this elimination was not carried out in our preliminary work on this problem [16] that resulted in a poorer overall grading accuracy (79.0%) than what we achieve with the elimination (89.0%) as we will present in Section 3. From the 30 images we analyzed, 99 10A, 49 10AT, 81 KCL, 80 DCIS, 29 CA1H, and 62 CA1A acinar structures were identified using this segmentation method yielding a total 400 acinar structures.

**2.4. Characterization of Acinar Structure Morphology.** Visual investigation of the acinar structures shown in Figure 1 reveals differences in acini morphology and localization of integrin subunits that are highlighted in Figure 2. Acinar structures in nonmalignant cell lines are comprised of polarized cells that are layered around the hollow lumen that closely approximate the acini formations in mammary glands and lobules. On the other hand, the acinar structures in the tumorigenic cell lines consist of nonpolarized cells that form clusters of cells rather than explicit acini. As shown in Figures 3(a) and 3(b), nontumorigenic cell line 10A and precancerous cell line AT, respectively, exhibit polarized acinar structures that are characterized by the integrin  $\alpha 6$  localization at the basal membrane of the cells, integrin  $\alpha 3$  localization along the lateral cell membranes, and clear hollow lumen formations. Acinar structures in KCL and DCIS cell lines shown in Figures 3(c) and 3(d), respectively, exhibit significant changes in the integrin subunit densities and

their colocalizations. While the basal and lateral membrane protein densities decrease, relative colocalizations of these proteins increase within the acinus. The acinar structures from the noninvasive carcinoma cell lines and acinar-like structures from the invasive carcinoma cell lines CA1H and CA1A shown in Figures 3(e) and 3(f), respectively, are more elongated and exhibit smaller hollow lumens than the acinar structures from the nonmalignant cell lines.

The observed variations in the morphology of acinar structures motivated the development of the features that characterize the level of cell polarity within the acinar structures. The features of interest primarily capture (i) the morphology of the acinar structure and hollow lumen, (ii) the basal, and (iii) the lateral protein densities within the acinar structure. These features were computed in each acinar structure for the slice that the acinar structure had the largest cross-section area along the depth direction (z-stack).

**2.4.1. Features Capturing the Morphology of Acinus and Hollow Lumen.** The first subset of features we propose captured the shape of the acinar structures, number of nuclei that constitute the acinus, and the relative size of the hollow lumen. The segmentation of the cell nuclei was also accomplished by the watershed segmentation technique described previously using the blue (nuclei) channel image only. Figure 4(a) shows the number of nuclei per acinar structure across the six cell lines. As expected the invasive carcinoma cell lines exhibit the largest number of nuclei per cell cluster compared to the other cell lines acini due to the unregulated division of cells.

Acinar structures in the nonmalignant grades of cancer typically have symmetrical round shapes. Malignant cancer grades result in deformations in the acinar structures that cause the shapes to be more elongated. We measured the roundness of the acinar structures by taking the ratio between the minor and major axes of the ellipse fitted as illustrated in Figure 4(d). This fitting was achieved by finding the ellipse that has the same normalized second central moments as the region. Roundness values close to 1 comes from rounder/more symmetrical acinar structures whereas values close to 0 indicate an elongated shape. As expected, noninvasive and invasive cell lines exhibit statistically significant decrease in acinar structure roundness compared to the nonmalignant cell lines as shown in Figure 4(b).

The final feature in this subset captured the differences in the relative size of hollow lumen by computing the ratio between the areas of the hollow lumen and the acinar structure. Values closer to 1 indicate larger hollow lumens, and those to 0 indicate smaller hollow lumens. In order to compute the hollow lumen area, we assumed that the hollow lumen was circular and computed its radius as the Euclidean distance between the centroid of the acinar structure and the nearest nucleus to the centroid as illustrated in Figure 4(e). The hollow lumen for the precancerous cell line AT is the largest among the six cell lines. Noninvasive and invasive cell line acinar structures exhibit statistically significant shrinkage in hollow lumen compared against the nonmalignant cell line acinar structures as shown in Figure 4(c).

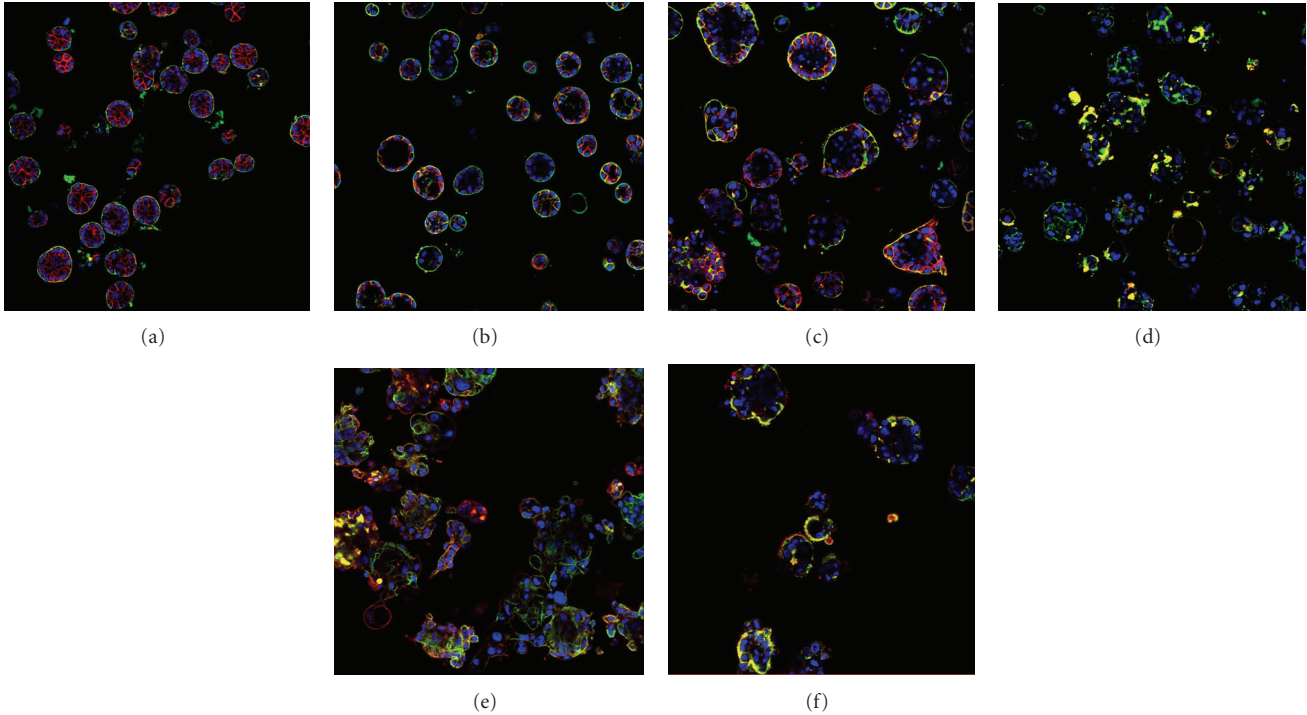


FIGURE 3: MCF10 series cell lines exhibit variation in acinar structure morphology along the metastatic cascade after 14 days in 3D laminin rich environment culture systems. Cells were cultured in 3D Matrigel suspensions and stained with integrin  $\alpha 3$  (red), integrin  $\alpha 6$  (green), and DAPI (blue). Example image slices from the six cell lines are typical of those captured during the course of the experiments and are displayed from (a) to (f) as follows: 10A, AT, KCL, DCIS, CA1H, and CA1A.

**2.4.2. Features Capturing the Basal Protein Integrin  $\alpha 6$  Density.** The next subset of features analyzed and quantified the localization and structural relationships of the basal membrane-localized protein integrin  $\alpha 6$ . The first feature measured the ratio of integrin  $\alpha 6$  expression within the acinar structure (excluding the expression along the basal membrane) to the acinar structure area as illustrated in Figure 5(e). As nonmalignant cancerous cells exhibit high cellular polarity, integrin  $\alpha 6$  is localized to the basal membrane of these cells. Hence, a low density of integrin  $\alpha 6$  is expected within the acinar structure. Reduced cellular polarity in the malignant grades of cancer enables greater internal expressions of integrin  $\alpha 6$  within the acinar structure. DCIS, CA1H, and CA1A cell lines exhibited statistically significant increase in the internal expression of integrin  $\alpha 6$  compared with the less malignant cell lines as shown in Figure 5(a).

The next feature characterized the spatial distribution of integrin  $\alpha 6$  within the acinar structure. It was computed as the ratio between the amount of the green fluorophores located in concentric circles that were centered at the centroid of the acinar structure and the total amount of green fluorophores in the acinar structure. The radii of the concentric circles were increased by one-tenth of the radius of the circle that circumscribes the acinar structure at each step as illustrated in Figure 5(g). Cell lines that have expressions of integrin  $\alpha 6$  near the centroid of the acinar structure show relatively early rise in the cumulative densities than the others. It is observed that integrin  $\alpha 6$  is localized near the

basal membrane in the three lower-grade cell lines, and near the centroids of the acini in the three advanced grade cell lines as shown in Figure 5(b). Since the cumulative densities near the acinus center and the basal membrane are similar for all the cell lines, we decide to include the cumulative densities corresponding to 30 to 70% of the radius of the circle that circumscribes the acinus in our analysis.

Another feature captured the amount of integrin  $\alpha 6$  expressed along the basal cell membrane as illustrated in Figure 5(f). First, we took the difference between the acinar structure and its morphologically eroded version to obtain a binary contour mask. This mask was then overlaid with the corresponding slice of the binary green channel image to obtain the integrin  $\alpha 6$  localized along the basal membrane of the acinar structure. Next, we superposed the resulting image with rays that initiate at the centroid of the acinus oriented at angles that varied from  $1^\circ$  to  $360^\circ$  with  $1^\circ$  steps scanning a complete circle. Finally, the total number of times that these rays intersected with the green markers was counted and normalized between 0 and 1 to obtain the continuity of integrin  $\alpha 6$  along the acinar basal side. Values close to 1 correspond to intact basal sides and, thus, indicate high cellular polarity. As shown in Figure 5(c), 10A and AT cell lines have the significantly more continuous expression of integrin  $\alpha 6$  along the basal side of the cells than the four malignant cell lines. We note that KCL cell line exhibits a statistically significant higher localization of integrin  $\alpha 6$  at the basal side than the three more tumorigenic cell lines.

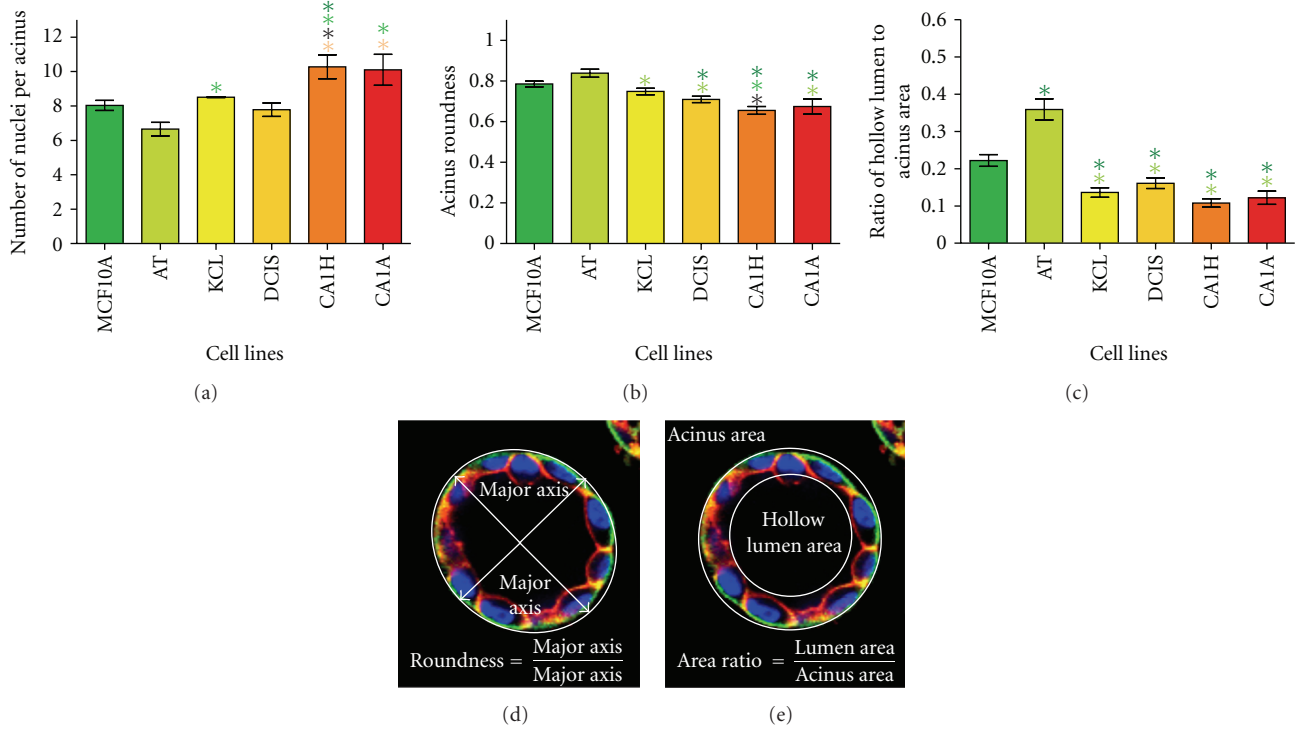


FIGURE 4: Features quantifying acinus and hollow lumens morphology are statistically significant between cell lines and states of health. (a) shows the number of nuclei per acinar structure, (b) shows the acinar structure elongation (roundness) that is measured as illustrated in (d) where the ratio between the minor and major axis of the ellipse fitted to the acinar structure is taken, and (c) shows the ratio between the hollow lumen and acinar structure areas that is computed as illustrated in (e). Error bars correspond to the standard error of the means. Two-sided  $t$ -tests were performed and cell lines that exhibit statistical significance ( $P < 0.05$ ) are marked with asterisks (\*).

The final feature in this category measured the ratio of the amount of integrin  $\alpha 6$  colocalized with integrin  $\alpha 3$  to the total expression of integrin  $\alpha 6$  to determine the amount basal membrane protein overlapped with the lateral membrane protein. This feature gets higher values when there is higher internal expression of integrin  $\alpha 6$  or higher basal membrane localization of integrin  $\alpha 3$ . As plotted in Figure 5(d), 10A and DCIS exhibit statistically significant less colocalization of integrin  $\alpha 6$  with  $\alpha 3$  than the other four cell lines.

**2.4.3. Features Capturing the Lateral Protein Integrin  $\alpha 3$  Density.** The final subset of features captured the density of lateral membrane protein integrin  $\alpha 3$  within the acinar structure. The following features measured the expression and localization of lateral membrane protein and adhesion molecule integrin  $\alpha 3$ . The first feature determined the amount integrin  $\alpha 3$  expressed along the basal cell membrane using the same method described previously. We anticipate observing less amount of integrin  $\alpha 3$  expressed along the basal cell membrane in the malignant grades of cancer than the nonmalignant grade.

Next, we captured the overall expression of integrin  $\alpha 3$  within the acinar structure as illustrated in Figure 6(e) where the total amount of integrin  $\alpha 3$  expressed in the acinar structure was divided by the area of the acinar structure. It is seen in Figure 6(b) that this feature monotonically decreases from the nontumorigenic 10A to noninvasive DCIS acinar

structures. The invasive carcinomas CA1H and CA1A show significantly higher expressions compared to the DCIS cells.

We then measured the ratio between the amount of integrin  $\alpha 3$  colocalized with the integrin  $\alpha 6$  and the total amount of integrin  $\alpha 3$  within the acinar structure. As expected nonmalignant cell lines exhibit significantly less colocalization of integrin  $\alpha 3$  with integrin  $\alpha 6$  than the malignant cell lines as shown in Figure 6(c).

Final feature in this subset measured the ratio between the amount of integrin  $\alpha 3$  localized between the hollow center and basal membrane side of the acinar structure to the total integrin  $\alpha 3$  expression. This feature, thus, quantified the density of  $\alpha 3$  integrin along the lateral membrane of the cells. Since integrin  $\alpha 3$  localization is not confined to cell-to-cell lateral membranes in malignant tumors, expressions of this protein are expected within the hollow lumen as well as the cell-to-extracellular matrix border.

**2.5. Automated Grading of Cancer Using Supervised Machine Learning.** We used supervised machine learning [28] to grade the acinar structures into nonmalignant, noninvasive carcinoma, and invasive carcinoma forms of breast cancer using the features defined previously. In supervised learning, the data are first divided into training and test sets. The classifier is trained with the labeled training data and the classes of the test data are then predicted using the resulting classifier. We tested the grading performance of five

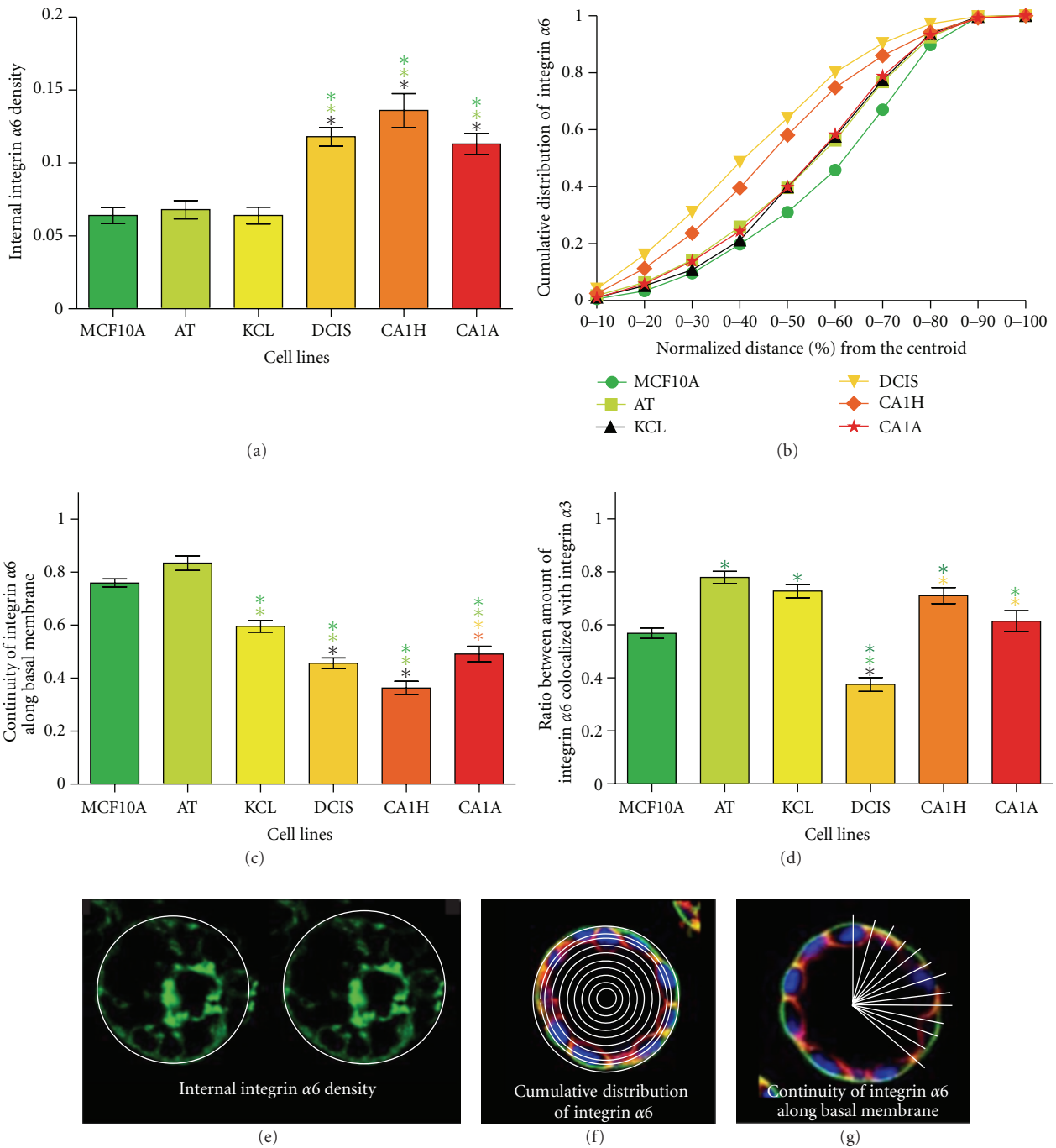


FIGURE 5: Features capturing the level of cell polarity based on the basal membrane protein integrin  $\alpha 6$  density. (a) shows the internal integrin  $\alpha 6$  density that is computed as illustrated in (e) where the amount of integrin  $\alpha 6$  localized inside the acini is divided to the total amount of integrin  $\alpha 6$ , (b) shows the cumulative spatial distributions of integrin  $\alpha 6$  that is computed as illustrated in (g) where the amount of integrin  $\alpha 6$  localized within the concentric circles are divided by the total amount of integrin  $\alpha 6$  within the acini, (c) shows the continuity of the integrin  $\alpha 6$  along the basal membrane that is measured as illustrated in (f) where the number of times that rays initiate from the centroid oriented at from 1 to 360 degrees intersected with the green fluorophores along the basal membrane is counted and normalized between 0 and 1, and (d) shows the ratio between the amount of integrin  $\alpha 6$  colocalized with integrin  $\alpha 3$  and the total integrin  $\alpha 6$  expression. Error bars correspond to the standard error of the means. Two-sided  $t$ -tests were performed and cell lines that exhibit statistical significance ( $P < 0.05$ ) are marked with asterisks (\*).

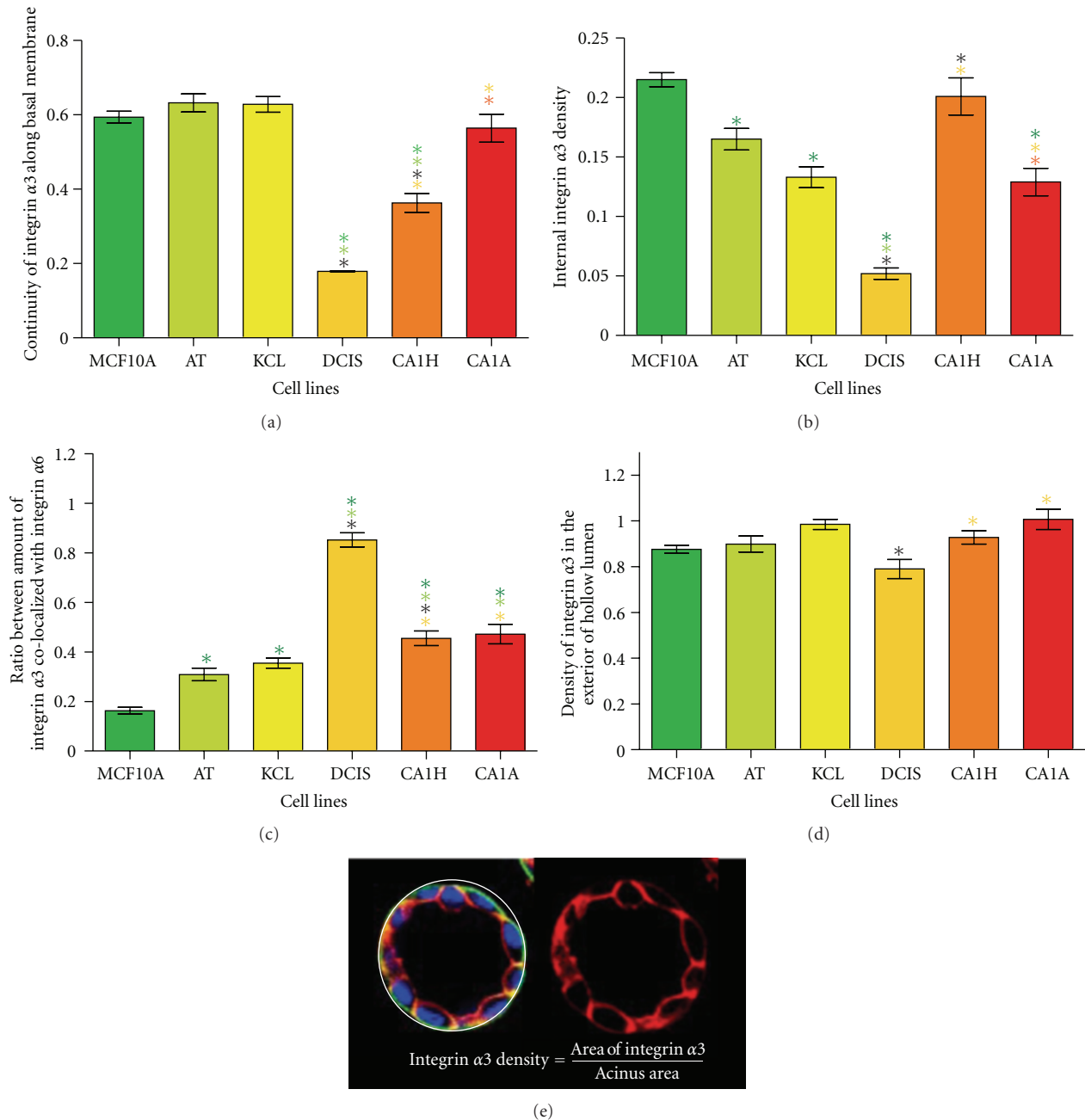


FIGURE 6: Features capturing the level of cell polarity based on the lateral membrane protein integrin  $\alpha 3$  density. (a) Shows the continuity of the integrin  $\alpha 6$  along the basal membrane, (b) shows the total integrin  $\alpha 3$  density that is computed as illustrated in (e) where the total amount of integrin  $\alpha 3$  within the acinar structure is divided by area of the acinar structure, (c) shows the ratio between the amount of integrin  $\alpha 3$  colocalized with integrin  $\alpha 6$  and the total integrin  $\alpha 3$  expression, and d shows the density of integrin  $\alpha 3$  in the exterior of the hollow lumen. Error bars correspond to the standard error of the means. Two-sided  $t$ -tests were performed and cell lines that exhibit statistical significance ( $P < 0.05$ ) are marked with asterisks (\*).

supervised learning algorithms. Linear discriminant analysis fits a multivariate normal density to each of the training class assuming equal covariance matrices for each class [29]. It then separates the classes with a hyperplane that is established by seeking the projection that maximizes the sum of intraclass distances and minimize the sum of interclass distances. Quadratic discriminant analysis is similar to

the linear discriminant analysis with the distinction that covariance matrix for each class is estimated separately [29]. Naïve Bayes classifier assumes that the classes are arising from independent distributions; hence, a diagonal covariance matrix is assumed.  $K$ -nearest neighbor classifier finds the  $K$  closest training data to the test data based on the Euclidean distance and classifies the sample using



the majority voting of the nearest points. Support vector machines classifier maps the training data into a higher dimensional space with a kernel function where the data become linearly separable [30]. A separating maximum-margin hyperplane is established to separate the different class data with minimal misclassification via quadratic optimization [30]. The test data are classified by determining the side of the hyperplane they lie on in the kernel-mapped space. In order to extend SVM for three-class classification, we employed the one-against-one approach [31] where three two-class SVM classifiers were established for each pair of classes in the training data set. Each sample in the test data was assigned to a class by these classifiers and the class with the majority vote was chosen as the final result. If there is equal voting for the three classes, we chose the class that has the largest margin from the separating hyperplane.

In order to obtain unbiased performance estimates, 10-fold cross-validation was performed. The feature set was first randomly divided into 10 disjoint partitions of equal size. For each partition, a classifier was trained with the remaining nine partitions and then tested on the retained partition. The results for each partition were then combined to find the overall grading accuracy. In order to reduce the scale differences within the features, the data were normalized so that the features had zero mean and unit variance across the samples. We performed a parametric search to determine the number of neighbors in the nearest neighbor identification. We tested between 8 and 15 nearest neighbors and determined that identifying 12 nearest neighbors to a test point achieved the highest grading accuracy. For the SVM classifiers, we used radial basis function, also referred to as Gaussian kernel, in the form of  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$  that mapped the data into an infinite dimensional Hilbert space [30]. We performed a parameter search to identify  $\sigma$  that achieves the highest grading accuracy. We sought  $\sigma$  in the set of candidate values that varied from 1.0 to 2.5 with 0.1 steps and determined that  $\sigma$  being equals to 2.0 achieved the best performance in the grading of the acinar structures. Table 1 shows the performance of the five supervised learning algorithms. It is clear that SVM-based classifier achieves the highest overall accuracy in the grading of the acinar structures. This is not unexpected as SVM classifiers are known to be highly successful in biological applications [32].

### 2.5.1. Discriminative Influence of the Proposed Features.

After identifying SVM-based classifier as the most accurate grading method for our data set, we performed feature selection to determine the discriminative capabilities of the features to characterize the data. A good feature selection algorithm identifies the features that are consistent within the class and exhibit differences between the classes. Fisher score (F-score) is documented to be a powerful feature selection tool [33]. For a dataset  $x_i$  with two classes, denote the instances in the first class with  $x_i^{(+)}$  and instances in the other class with  $x_i^{(-)}$ . The Fisher score of the  $i$ th feature is then given by  $F(i) = ((\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2) / (\sigma_{x_i^{(+)}}^2 + \sigma_{x_i^{(-)}}^2)$ , where  $\bar{x}_i$  is the mean value of the  $i$ th features,  $\bar{x}_i^{(+)}$  and  $\bar{x}_i^{(-)}$  are the mean

TABLE 1: Overall grading accuracies for the supervised machine learning techniques used in our analysis. SVM classifiers clearly achieve significantly higher grading accuracy than the other methods. The data set includes 400 acinar structures.

Learning method	Overall grading accuracy (%)
Linear discriminant analysis	80.75
Quadratic discriminant analysis	80.00
Naïve Bayes	69.75
$K$ -nearest neighbors	79.50
Support vector machines	<b>89.00</b>

values over the positive and negative instances, respectively, and  $\sigma_{x_i^{(+)}}^2$  and  $\sigma_{x_i^{(-)}}^2$  are the variances over the positive and negative instances, respectively, for the  $i$ th feature. In order to extend this method for a feature set with three classes, we computed the F-score for each feature in each pair of classes and then took the average of the three possible combinations. Larger values of F-score indicate stronger discriminative influence; therefore, after obtaining the F-scores for all the features, we ranked them in descending order.

Table 2 shows the average F-scores of the features and their corresponding discriminative rank. It is seen that continuity of integrin  $\alpha 6$  along the basal membrane is the most discriminative feature to describe the data. Discriminative influence of the ratio between the hollow lumen and acinar structure area is also high. These two features are important as they are strong indicators of the level of cell polarity within the acinar structure and can be considered as measures of duct forming tumors as used in the commonly used Scarff-Bloom-Richardson system. Some of the other most discriminative features quantify the differences that are difficult to assess by visual inspection such as the colocalization between the integrin  $\alpha 3$  and integrin  $\alpha 3$  and internal densities of the integrin subunits. We note that these features constitute a subset of the novel features introduced in this paper and, therefore, this analysis particularly highlights the importance of the proposed features and the methodology.

We performed cancer grading using subsets of the resulting most discriminative features, where at each stage we increased the number of most discriminative features in the training feature set. Figure 7 shows the overall grading accuracy with respect to the number of most discriminative features selected in the grading. It is seen that the highest grading accuracy 89.0% is achieved when all of the features are used to train the classifiers. Considering that the most grading systems are typically based on the assessment of a limited number of features, we also investigate the grading performance of our methodology in a similar setting. When only the first five most discriminative features are used for grading, we achieve 82.0% overall grading accuracy. Though relatively worse than the highest grading accuracy, this constitutes a highly promising setting considering the limited number of features.

2.6. Tissue Sections. Breast tissue from patients with invasive carcinoma, ductal carcinoma *in situ*, and from healthy

TABLE 2: Discriminative influence of the proposed features based on the F-score values.

Feature label	F-Score	Rank
Number of nuclei within the acinar structure	0.0547	14
Elongation of the acinar structure	0.1122	6
Ratio between the hollow lumen and acinar structure areas	0.1602	4
Internal integrin $\alpha 6$ Density	0.1205	5
Density of integrin $\alpha 6$ inside the 30% circle	0.0620	11
Density of integrin $\alpha 6$ inside the 40% circle	0.0611	12
Density of integrin $\alpha 6$ inside the 50% circle	0.0885	9
Density of integrin $\alpha 6$ inside the 60% circle	0.0953	8
Density of integrin $\alpha 6$ inside the 70% circle	0.0598	13
Continuity of integrin $\alpha 6$ along the basal membrane	0.5725	1
Amount of integrin $\alpha 6$ colocalized with integrin $\alpha 3$	0.0337	15
Continuity of integrin $\alpha 3$ along the basal membrane	0.1046	7
Total integrin $\alpha 3$ density	0.2651	3
Amount of integrin $\alpha 3$ colocalized with integrin $\alpha 6$	0.3526	2
Density of integrin $\alpha 3$ in the exterior of the hollow lumen	0.0716	10

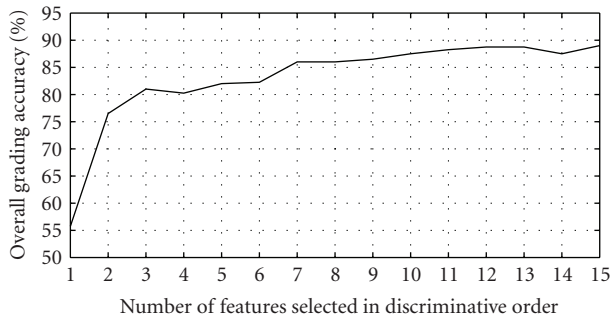


FIGURE 7: Overall grading accuracy with respect to the number of most discriminative features selected in the grading. Highest overall grading accuracy is achieved when the whole feature set is considered.

individuals was obtained from ProteoGenix (Culver City, CA) and stored at  $-80^{\circ}\text{C}$ . Portions of each tissue block were embedded in O.C.T. Compound (Sakura Finetek USA, Torrance, CA) and sections  $20\mu\text{m}$  thick were cut using a Microm HM505E cryostat. Tissue sections were adhered to Superfrost Plus Gold microscope slides (Fisher Scientific, Morris Plains, NJ) and vapor fixed using 2-3 Kimwipes (Kimberly-Clark Worldwide, Roswell, GA) soaked in 4% paraformaldehyde in a small chamber at  $-20^{\circ}\text{C}$  for 30 minutes prior to immunohistochemistry. Tissue sections were encircled with an ImmEdge Pen (Vector Labs, Burlingame, CA) and immunohistochemistry and confocal microscopy were performed as described for *in vitro* samples in Section 2.2. An Alexa488-conjugated goat anti-fluorescein secondary antibody (Invitrogen, Carlsbad, CA) was included at a 1 : 300 dilution in the overnight secondary incubation.

### 3. Results and Discussion

A workflow diagram involved in the automated grading of breast cancer is shown in Figure 8. First, 3D fluorescent confocal images of acinar structures were collected after 14 days within an *in vitro* culture system as described in Sections 2.1 and 2.2. A total of 30 images were captured (5 images for each cell line) and a total of 400 acinar structures were identified using the image segmentation method described in Section 2.3 in these images. For each acinar structure, we extracted the morphological features described in Section 2.4. These features exhibit high-level of statistical significance across the nonmalignant, noninvasive carcinoma, and invasive carcinoma grades of breast cancer. Using the resulting unique feature profiles, we then trained the SVM-based classifiers as described in Section 2.5 for the automated grading of acini and test on the data.

Overall grading accuracy of 89.0% is achieved when the whole feature set is used in the grading. As shown in the first half of Table 3, the nonmalignant 10A and AT cell lines are graded with 97.0% and 73.5% accuracy, respectively. Due to the strong resemblance between the AT and KCL cell lines, some portions of the AT acinar structures are graded as noninvasive carcinoma. Acinar structures in KCL cell line are correctly graded as noninvasive carcinoma 72.1% of the time, and as nonmalignant 11.1% of the time which should be considered as high success as KCL falls between the advanced-early stages of cancer and noninvasive carcinoma. We observe significantly high grading accuracies in the remaining three cell lines. Noninvasive DCIS acinar structures are graded with 95.0% accuracy, and the acinar structures in the invasive cell lines CA1H and CA1A are graded with 93.1% and 93.5% accuracy, respectively. We note that neither the DCIS nor invasive cell line acinar structures are graded as nonmalignant. When only the five

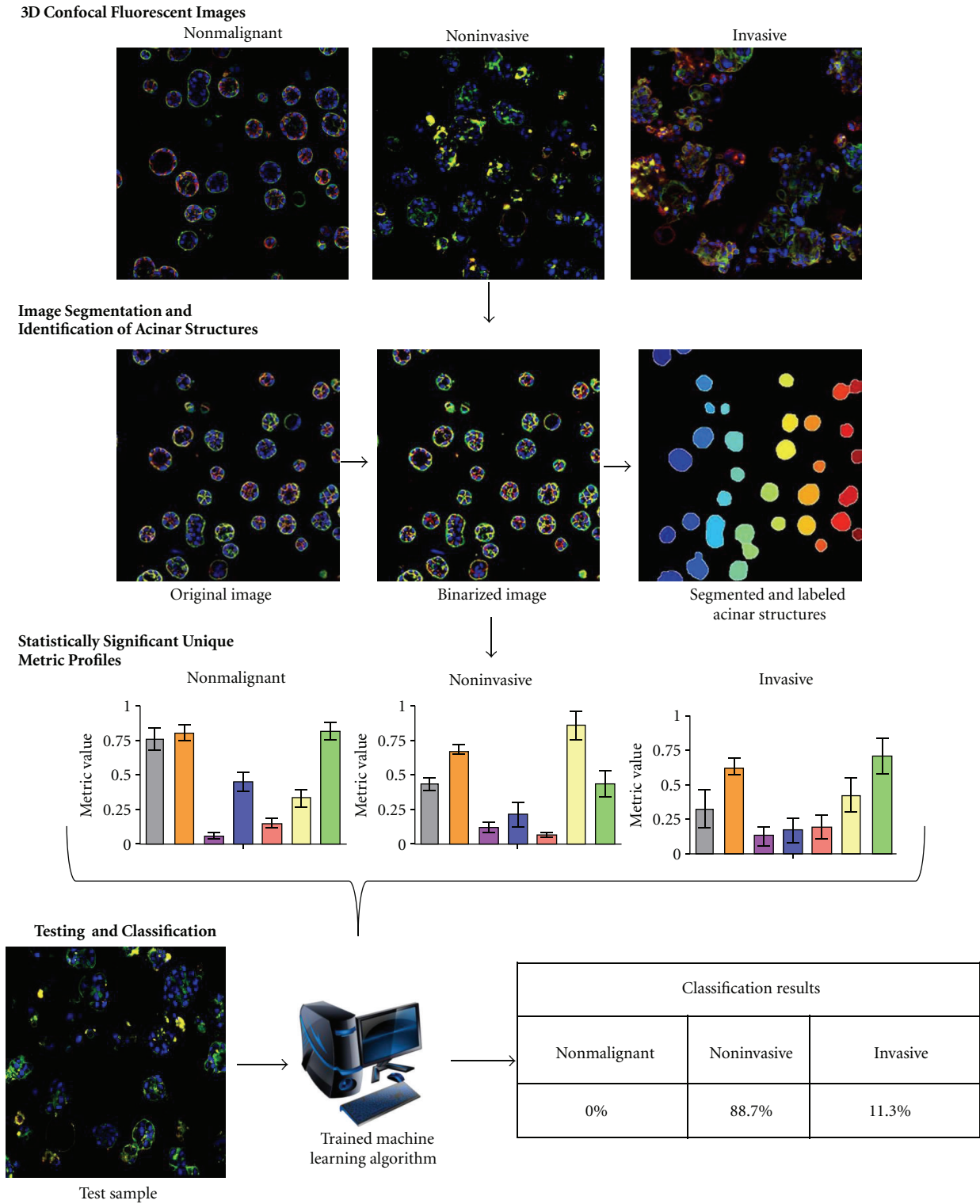


FIGURE 8: A flow chart of the proposed quantitative analysis approach in measuring morphological-based features of acini. First, fluorescently labeled images of the cultures are captured by 3D confocal microscopy. Next, individual acinar structures in these images are segmented and labeled. We then extract the proposed features based on the acini morphology that lead to statistically significant unique feature profiles between functionally different stages of cancer. Finally, automated grading of cancer is achieved by supervised machine learning.

TABLE 3: Performance of the SVM-based classifier with the proposed quantitative features. Acinar structures are graded as nonmalignant as trained from the 10A and AT samples, noninvasive carcinoma as trained from the KCL and DCIS samples, and invasive carcinoma as trained from the CA1H and CA1A samples using all of the proposed features as well as the five most discriminative features: continuity of integrin  $\alpha 6$  along the basal membrane, colocalization of integrin  $\alpha 3$  with integrin  $\alpha 6$ , integrin  $\alpha 3$  density, ratio between hollow lumen and acinar structure areas, and integrin  $\alpha 6$  density. The data set includes 400 acinar structures: 99 10A, 49 10AT, 81 KCL, 80 DCIS, 29 CA1H, and 62 CA1A.

Grading results utilizing overall feature set		Grading categories (%)		
		Nonmalignant	Noninvasive Carcinoma	Invasive Carcinoma
Nonmalignant	10A	<b>97.0</b>	2.0	1.0
	AT	<b>73.5</b>	24.5	2.0
Noninvasive Carcinoma	KCL	11.1	<b>72.1</b>	14.8
	DCIS	0.0	<b>95.0</b>	5.0
Invasive Carcinoma	CA1H	0.0	6.9	<b>93.1</b>
	CA1A	0.0	6.5	<b>93.5</b>
Grading results utilizing the five most discriminative features		Grading categories (%)		
		Nonmalignant	Noninvasive Carcinoma	Invasive Carcinoma
Nonmalignant	10A	<b>93.0</b>	3.0	4.0
	AT	<b>85.7</b>	10.2	4.1
Noninvasive Carcinoma	KCL	18.5	<b>53.1</b>	28.4
	DCIS	0.0	<b>93.8</b>	6.2
Invasive Carcinoma	CA1H	0.0	27.6	<b>72.4</b>
	CA1A	0.0	8.1	<b>91.9</b>

most discriminative features are used, we achieve 82.0% overall accuracy in the grading. As shown in the second half of Table 3, in this case AT acinar structures are graded with higher success. However, the grading accuracy of noninvasive KCL and invasive CA1H acinar structures significantly decreases. It is not unexpected that some of the features not included after the feature selection capture significant characteristics about the acini morphology. The following discussion on the proposed features helps us understand how they relate to the underlying biological implications.

Changes in the hollow lumen size and acini shape are visually observed and quantified across multiple cancer stages as described in Section 2.4.1. While the average number of cells per acini is a concrete measurement and could be determined manually, it is useful and more practical to utilize automated techniques when analyzing large volumes of image data. With this feature we clearly determine that the two invasive cell lines CA1H and CA1A have more nuclei in the acinar structures than the other less malignant cell lines. This could arise from several biological factors. A simple reason is that large acinar structures are comprised of more cells. Alternatively, this could arise from higher density of cells within the acinus due to the loss of the hollow lumen. Both of these explanations likely contribute to the resulting feature making it challenging to determine the exact cause. Our next feature, the ratio between the areas of the hollow lumen and acinar structure can also help us explain the general trend in the number of nuclei per acinus with the progression along the metastatic cascade. Larger hollow lumens in the lower cancer grades possibly limit the space

that the cells can proliferate; thus smaller number of nuclei per acinus is observed at lower-grade cell lines.

The hollow lumen to acinar structure area ratio captures a key change in acinar structure which is challenging to assess by visual inspection. While the presence or absence of a hollow lumen is observable by eye, evaluating the relative size of the lumen compared to the overall acinar structure is difficult and poses subjectivity. This feature enabled us to quantitatively characterize this relationship and helps identify the significant changes between cancer grades. The loss of hollow lumens is associated with increasing cell division and cell survival within the acinar structure. In native breast tissue a decrease in the size of the hollow lumen arises from increasing metastatic capability. This is consistent in our *in vitro* system and the resulting quantitative features. However, despite generally observing a decrease in the hollow lumen to acinar structure area ratio with the progression of cancer, precancerous AT cell line exhibits an increase compared to the 10A samples. While it is unclear why this is the case, it could potentially be due to the cells becoming flatter rather than columnar in shape creating a larger hollow lumen.

In addition to the changes in the hollow lumen morphology, the acinar structure elongation quantifies the subtle changes in the roundness of acini that are also difficult assess by eye. As shown by our results, we capture statistically significant differences in the acini elongation caused by the increasing metastatic capability, reflecting the progressive loss of structural integrity in acini as breast cancer develops. Replacing these qualitative observations on the acini and

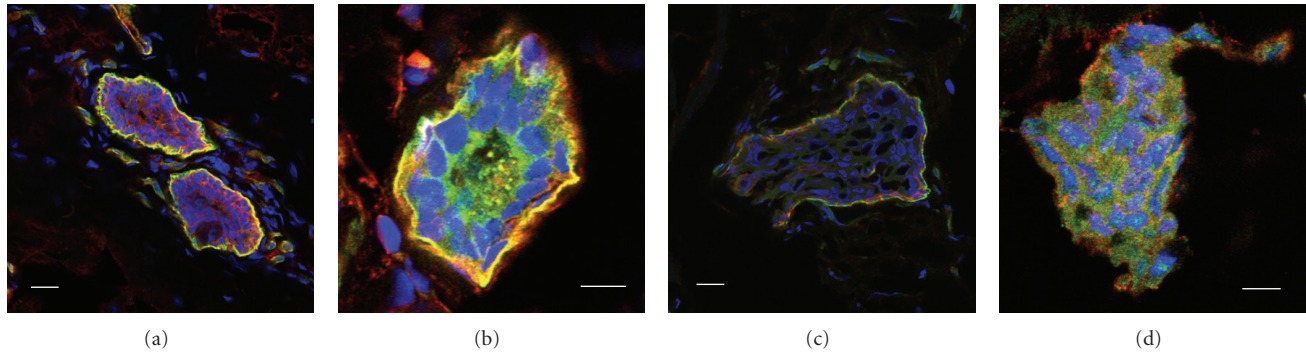


FIGURE 9: Tissue samples from human patients exhibit variations in acini morphology along the metastatic cascade similar to the 3D cultures. In the nontumorigenic glands from a healthy patient shown in (a), integrin  $\alpha 3$  can be observed along the lateral surface of the epithelia, while integrin  $\alpha 6$  shows strong staining across the basal surface. Precancerous gland shown in (b) exhibits a clear reduction in the amount of integrin  $\alpha 3$  expression along the lateral membranes of the epithelia, while basal integrin  $\alpha 6$  expression remains strong. Noninvasive carcinoma gland shown in (c) could be easily identified by the loss of a proper hollow lumen and stratification of the epithelial layer. In these large regions of tissue, integrin  $\alpha 3$  is absent from cells in the interior of the gland; however, along with integrin  $\alpha 6$ , there can be observed along the basal surfaces of cells in contact with the basal lamina. Invasive carcinoma gland shown in (d) includes cords or clusters of cells exhibiting faint labeling of both proteins distributed along the cell membrane compared to background and control tissue. Scale bars represent  $20 \mu\text{m}$  in (a) and (c), and  $10 \mu\text{m}$  in (b) and (d).

hollow lumen morphology with these robust quantitative measurements enabled us to characterize the data more objectively.

The density of basal membrane protein integrin  $\alpha 6$  is quantified through multiple features as described in Section 2.4.2 and found to reflect the metastatic potential of acinar structures. The quantitative and statistically significant changes identified in the analysis of integrin  $\alpha 6$  distribution strongly correlate with our expectations based on the visual inspection of the acinar structures. It is seen that the 10A and AT acini exhibited the highest amount of integrin  $\alpha 6$  localization along the basal membrane, reflecting the presence of an intact, extracellular matrix basement membrane on the basal surface of these cells. In addition, it is observed that the acinar structures of more malignant cell lines showed dramatic loss in the localization of integrin  $\alpha 6$  along the basal membrane of the cells and an increase in the internal localization of this protein, reflecting loss of the basement membrane. Features extracted from CA1A acinar structures are significantly different from some of the features extracted from the CA1H and DCIS acinar structures that can be explained by the degree of malignancy of this cell line. CA1A cells exhibit the highest malignancy of all the other cell lines with the advanced ability to metastasize. When cells are able to metastasize, cells may form secondary tumors called macrometastases. At this stage, these cells are able to adapt to their new environment in order to improve their abilities of survival. It is possible that the CA1A cells are behaving as a secondary tumor when placed in a basement membrane protein rich environment causing them to acquire some epithelial like phenotypes, such as increasing the basal membrane localization of integrin  $\alpha 6$ . Thus, these features thoroughly capture the distribution of

integrin  $\alpha 6$  throughout the acinar structure as visualized and are functionally expected across the different grades of cancer.

The features based on the density of integrin  $\alpha 3$  as described in Section 2.4.3 yield both expected and unexpected results. As reported in a previous study [34], we expect to observe different levels of total integrin  $\alpha 3$  expression across the cell lines of varying metastatic potential. Our expectations are confirmed by our results that the total integrin  $\alpha 3$  expressions exhibit a decrease between the nonmalignant and noninvasive carcinomas and a higher expression is observed in the invasive carcinoma cell lines (compared to the noninvasive KCL and DCIS cell line features). These features support the idea that integrin  $\alpha 3$  switches its function from acting in a cell-cell adhesion manner to cell-ECM adhesion manner with increasing metastatic ability. As shown in Figure 6(a), the basal continuity of integrin  $\alpha 3$  is the lowest for the DCIS cell lines as expected due to the low density of integrin  $\alpha 3$  in this cell line. The continuity at the basal membrane displayed increased localization in the invasive carcinoma cell lines suggesting a cell-ECM adhesion role. The measurement of integrin  $\alpha 3$  density in the exterior of the hollow lumen as shown in Figure 6(d) determined that the progression of cancer yields increased localization of this protein within the hollow lumen. It was at similar levels for the 10A, AT, and KCL cell lines however that the expression ratio increased in expression between the DCIS through CA1A cells. This feature was developed to quantify the functional switch of integrin  $\alpha 3$  from cell-cell adhesion to cell-ECM adhesion with progressing metastatic state. The increase in this feature suggests a switch from cell-cell adhesion to cell-ECM adhesions; however, this change may be influenced by the reduction in the size of the hollow

lumen and increased integrin  $\alpha 3$  expression between cells throughout the acinar structure, thus hindering the robust characterization of the integrin  $\alpha 3$  lateral localization.

The amount of an integrin subunit that colocalizes with the other varies across the cell lines. Basal membrane integrin  $\alpha 6$  colocalizes with integrin  $\alpha 3$  more in the tumorigenic cell lines (except the DCIS) than the nontumorigenic cell line 10A. This could be due to the increasing internal expression of integrin  $\alpha 6$  indicating a loss in cell polarity. Another reason could be due to integrin  $\alpha 3$  changing localization from the lateral membrane to the basal membrane as it switches functions from cell-cell adhesion to cell-ECM adhesions. It could also be a combination of the two since loss in cell polarity results in irregular localization of the proteins; thus both integrin subunits are expressed throughout the cell membrane. Interestingly, DCIS acini exhibit the lowest colocalization of integrin  $\alpha 6$  with integrin  $\alpha 3$ . This could be due to the low expression of integrin  $\alpha 3$  in DCIS cells as shown in Figure 6(b) that yield higher amounts of isolated integrin  $\alpha 6$  subunits in the acinar structures.

The alternative comparison is the amount of integrin  $\alpha 3$  that colocalizes with integrin  $\alpha 6$ . We observe that the non-invasive and the invasive cell lines exhibit higher colocalization than the nonmalignant cell lines. Interestingly, DCIS cell line has the highest amount of colocalization. This confirms our suggestion that low expression of integrin  $\alpha 3$  is the cause of the lowest colocalization of integrin  $\alpha 6$  with integrin  $\alpha 3$  in this cell line. In this case almost all of the integrin  $\alpha 3$  is colocalized with integrin  $\alpha 6$  due to its low expression levels and loss of cell polarity in the DCIS acinar structures. The nonmalignant cell lines exhibit higher levels of colocalization than the noninvasive cell lines. By comparing the two colocalization features, we infer that this is likely due to integrin  $\alpha 3$  localization at the basal and lateral membranes while integrin  $\alpha 6$  localizes primarily at the basal membrane alone. This is reflected in 80% colocalization of integrin  $\alpha 6$  with integrin  $\alpha 3$  and 30% colocalization of integrin  $\alpha 3$  with integrin  $\alpha 6$  as shown in Figure 6(c). On the other hand, the invasive carcinoma cell lines exhibit approximately 70% colocalization of the basal membrane integrin with the lateral membrane integrin and approximately 50% colocalization of the lateral membrane integrin with the basal membrane integrin. Due to the approximately same values, this suggests that both integrin subunits have lost their specific localization and are expressed throughout. We also can further confirm this with the internal densities of both integrin  $\alpha 6$  and integrin  $\alpha 3$  having higher values.

Proposed features display statistical significance across the cell lines with varying metastatic abilities. This indicates that the proposed features capture and quantify biologically relevant morphological changes. These features have the potential for studying structure-function relationships in a controlled and quantitative system. This application is useful in identifying underlying mechanisms of cancer and the role of specific protein functions on acinar structures. In addition, this approach could be used to test the effects of potential chemotherapy targets and combinations of drugs on the structures of both nontumorigenic, precancerous,

noninvasive carcinoma and invasive carcinoma cells in preliminary studies. Also, with increasing medical imaging technology the future holds potential for many quantitative and computerized diagnostic tools and systems to aid doctors in diagnosis, treatment options, and prognosis. Finally, these features could be applied to current histology samples when tagged with fluorescent antibodies to quantify complex structural features.

In order to demonstrate the relevance of our approach to tissue samples obtained from human patients, we performed immunohistochemistry on frozen sections as described in Section 2.6 following the same procedure for cell culture experiments as outlined from Section 2.2 to Section 2.5. Images of nontumorigenic, precancerous, noninvasive, and invasive mammary gland tissue were collected using confocal microscopy. The random orientation of glands in sectioned material presented a challenge as the acinar structures in *in vitro* cultures are typically spherical in shape. In order to eliminate the longitudinal or oblique planes of sections from the present analysis, single optical sections of glands oriented roughly in crosssection that resemble the acinar formations in the *in vitro* cultures were cropped manually. Figure 9 shows examples of glands analyzed in our study. Although nonspecific Alexa568 secondary antibody labeling of luminal contents was observed in some cases, visual inspection of these images indicates that the pattern of both integrin  $\alpha 6$  and integrin  $\alpha 3$  exhibit similar staining patterns to *in vitro* acinar structures. In our preliminary study, we identified 12 glands in the 9 images analyzed in this study. For each gland, we extracted the proposed features and performed grading using the SVM-based classifier trained with the *in vitro* feature set. The *in vivo* test set was graded accurately except one nontumorigenic gland that was graded as non-invasive carcinoma. Nevertheless, we note that developing a computer-aided grading system for the grading of *in vivo* tissue samples is beyond the scope of this paper and left as future work.

#### 4. Conclusion

In this paper, we present a method that enables quantitative characterization of 3D breast culture acini with varying metastatic potentials. Specifically, we propose statistically significant features based on acinar structure morphology that capture differences between different grades of cancer that are difficult to assess under microscopic inspection. The experimental results demonstrate the efficacy of the proposed features to differentiate between the nonmalignant, noninvasive carcinoma, and invasive carcinoma grades of breast cancer with 89.0% accuracy. In addition, our preliminary studies indicate that our methodology can also be used for the grading of cancer in *in vivo* tissues provided that the captured tissue samples include cross-sectional portions of the glands. Hence, our method demonstrates great promise to model morphology-function relationships within controlled 3D systems and hold potential as an automatic breast cancer prognostic tool in current histology samples.

## Authors' Contribution

L. M. Polizzotti and B. Oztan contributed equally to this work.

## Acknowledgments

This work was supported in part by NIH Grant RO1 EB008016. The authors would like to thank Dr. C. Cagatay Bilgin for helping with stitching the image tiles and K. M. Henderson and Nimit Dhulekar for discussions and suggestions.

## References

- [1] "Cancer Facts and Figures," American Cancer Society, 2011.
- [2] M. A. Hayat, *Methods of Cancer Diagnosis, Therapy and Prognosis: Breast Carcinoma*, Springer, New York, NY, USA, 2008.
- [3] A. J. Jensen, A. M. Naik, R. F. Pommier, J. T. Vetto, and M. L. Troxell, "Factors influencing accuracy of axillary sentinel lymph node frozen section for breast cancer," *American Journal of Surgery*, vol. 199, no. 5, pp. 629–635, 2010.
- [4] J. S. Suri and R. M. Rangayyan, *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*, SPIE, Bellingham, Wash, USA, 2006.
- [5] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: a review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.
- [6] K. J. Martin, D. R. Patrick, M. J. Bissell, and M. V. Fournier, "Prognostic breast cancer signature identified from 3D culture model accurately predicts clinical outcome across independent datasets," *PLoS One*, vol. 3, no. 8, Article ID e2994, 2008.
- [7] V. M. Weaver, A. H. Fischer, O. W. Peterson, and M. J. Bissell, "The importance of the microenvironment in breast cancer progression: recapitulation of mammary tumorigenesis using a unique human mammary epithelial cell model and a three-dimensional culture assay," *Biochemistry and Cell Biology*, vol. 74, no. 6, pp. 833–851, 1996.
- [8] B. Weigelt and M. J. Bissell, "Unraveling the microenvironmental influences on the normal mammary gland and breast cancer," *Seminars in Cancer Biology*, vol. 18, no. 5, pp. 311–321, 2008.
- [9] C. Hebner, V. M. Weaver, and J. Debnath, "Modeling morphogenesis and oncogenesis in three-dimensional breast epithelial cultures," *Annual Review of Pathology: Mechanisms of Disease*, vol. 3, pp. 313–339, 2008.
- [10] H. Chang, C. Park, and B. Parvin, "Quantitative representation of three-dimensional cell culture models," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '07)*, pp. 89–92, April 2007.
- [11] K. A. Rejniak, S. E. Wang, N. S. Bryce et al., "Linking changes in epithelial morphogenesis to cancer mutations using computational modeling," *PLoS Computational Biology*, vol. 6, no. 8, Article ID e1000900, 2010.
- [12] J. Tang, H. Enderling, S. Becker-Weimann et al., "Phenotypic transition maps of 3D breast acini obtained by imaging-guided agent-based modeling," *Integrative Biology*, vol. 3, no. 4, pp. 408–421, 2011.
- [13] R. Fernandez-Gonzalez, M. H. Barcellos-Hoff, and C. Ortizde-Solórzano, "A tool for the quantitative spatial analysis of complex cellular systems," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1300–1313, 2005.
- [14] L. McKeen, C. C. Bilgin, G. E. Plopper, and B. Yener, "Investigation of Structure-Function Relationships of Breast Cancer's Metastatic Potential," in *49th American Society for Cell Biology Annual Meeting*, 2009.
- [15] L. McKeen-Polizzotti, K. M. Henderson, B. Oztan, C. C. Bilgin, B. Yener, and G. E. Plopper, "Quantitative metric profiles capture three-dimensional temporospatial architecture to discriminate cellular functional states," *Bmc Medical Imaging*, vol. 11, article no. 11, 2011.
- [16] B. Oztan, L. McKeen-Polizzotti, C. C. Bilgin, K. M. Henderson, B. Yener, and G. E. Plopper, "Classification of breast cancer grades through quantitative characterization of ductal structure morphology in three-dimensional cultures," in *ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB '11)*, Chicago, Ill, USA, 2011.
- [17] F. R. Miller, H. D. Soule, L. Tait et al., "Xenograft model of progressive human proliferative breast disease," *Journal of the National Cancer Institute*, vol. 85, no. 21, pp. 1725–1732, 1993.
- [18] M. J. Worsham, G. Pals, J. P. Schouten et al., "High-resolution mapping of molecular events associated with immortalization, transformation, and progression to breast cancer in the MCF10 model," *Breast Cancer Research and Treatment*, vol. 96, no. 2, pp. 177–186, 2006.
- [19] F. R. Miller, "Xenograft models of premalignant breast disease," *Journal of Mammary Gland Biology and Neoplasia*, vol. 5, no. 4, pp. 379–391, 2000.
- [20] G. Y. Lee, P. A. Kenny, E. H. Lee, and M. J. Bissell, "Three-dimensional culture models of normal and malignant breast epithelial cells," *Nature Methods*, vol. 4, no. 4, pp. 359–365, 2007.
- [21] S. Preibisch, S. Saalfeld, and P. Tomancak, "Globally optimal stitching of tiled 3D microscopic image acquisitions," *Bioinformatics*, vol. 25, no. 11, pp. 1463–1465, 2009.
- [22] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with imageJ," *Biophotonics International*, vol. 11, no. 7, pp. 36–41, 2004.
- [23] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, 1994.
- [24] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans Syst Man Cybern*, vol. 9, no. 1, pp. 62–66, 1979.
- [25] B. Oztan, G. Sharma, A. Pasupuleti, and P. R. Mukund, "Processing of degraded documents for long-term archival using Waferfiche technology," in *Preservation Strategies and Imaging Technologies for Cultural Heritage Institutions and Memory Organizations (Archiving '09)*, pp. 197–202, May 2009.
- [26] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer, New York, NY, USA, 2nd edition, 2002.
- [27] C. R. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 265–270, 2003.
- [28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2000.
- [29] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, New York, NY, USA, 2004.
- [30] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods*, Cambridge University Press, New York, NY, 2000.

- [31] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [32] Z. R. Yang, "Biological applications of support vector machines," *Briefings in bioinformatics*, vol. 5, no. 4, pp. 328–338, 2004.
- [33] Y. W. Chen and C. J. Lin, "Combining SVMs with various feature selection strategies," *Studies in Fuzziness and Soft Computing*, vol. 207, pp. 315–324, 2006.
- [34] D. K. Rhee, S. H. Park, and Y. K. Jang, "Molecular signatures associated with transformation and progression to breast cancer in the isogenic MCF10 model," *Genomics*, vol. 92, no. 6, pp. 419–428, 2008.