ARCHITECTURAL PATTERNS FOR DIFFERENTIAL DIAGNOSIS OF PROLIFERATIVE BREAST LESIONS FROM HISTOPATHOLOGICAL IMAGES

L. Nguyen[♭⋆†], A.B. Tosun[♭⋆], J.L. Fine[♮], D.L. Taylor[⋆†], S.C. Chennubhotla[⋆†]

[⋆]Department of Computational and Systems Biology, [†]Drug Discovery Institute,
[♮]Magee Womens Hospital of UPMC and Department of Pathology, University of Pittsburgh

ABSTRACT

The differential diagnosis of proliferative breast lesions, benign usual ductal hyperplasia (UDH) versus malignant ductal carcinoma in situ (DCIS) is challenging. This involves a pathologist examining histopathologic sections of a biopsy using a light microscope, evaluating tissue structures for their architecture or size, and assessing individual cell nuclei for their morphology. Imposing diagnostic boundaries on features that otherwise exist on a continuum going from benign to atypia to malignant is a challenge. Current computational pathology methods have focused primarily on nuclear atypia in drawing these boundaries. In this paper, we improve on these approaches by encoding for both cellular morphology and spatial architectural patterns. Using a publicly available breast lesion database consisting of UDH and three different grades of DCIS, we improve the classification accuracy by 10% over the state-of-the-art method for discriminating UDH and DCIS. For the four way classification of UDH and the three grades of DCIS, our method improves the results by 6% in accuracy, 8% in micro-AUC, and 19% in macro-AUC.

Index Terms— Breast Cancer, DCIS, Computational Pathology, Architectural Patterns, Classification

1. INTRODUCTION

A recent and controversial article [1] reported on the issue of diagnostic discordance among pathologists' intrepretations of breast core biopsy specimens. Pathologists tend to agree on invasive carcinoma, but have high level of discordance on atypical ductal hyperplasia and ductal carcinoma in situ (DCIS). For example, usual ductal hyperplasia (UDH) is considered a benign proliferation and no treatment is prescribed. Unlike UDH, DCIS is considered a preinvasive malignant proliferation, and thus requiring aggressive treatment (such as lumpectomy with or without whole breast radiation therapy) [2]. Therefore, there is a great value in developing computational pathology tools that can aid in minimizing diagnostic discordance and providing quantitative measurements for differential diagnosis of proliferative breast lesions.

[♭]Equal contribution. Thanks to funding from UPMC CCA Grant #711077, NIH-NCI U01CA20492601, and NHGRI U54HG008540.

1.1. Background and Related Work

Gold standard practice of pathology involves pathologists examining hematoxylin and eosin (H&E) stained tissue slides under light microscope and making diagnostic and prognostic decisions. Differentiating breast lesions into UDH and DCIS requires careful examination of diagnostic criteria, including cellular features, tissue architectures, and spatial extent [3].

Most notable works on differentiating breast lesions include C-Path Early Breast Neoplasia [4] and DCIS CAD system [5] on whole slide images (WSIs). C-path focuses solely on nuclear features such as morphology, intensity, and texture. DCIS CAD includes some architectural features such as margination of regions of interest (ROIs), nuclei distribution, and stroma/background quantification. However, nuclear and architectural features are treated as independent features by [5], thus ignoring the standard diagnostic practice of examining the spatial distribution of heterogeneous cell types in a lesion [3]. Our method aims to overcome these limitations by incorporating cell phenotypes in building architectural patterns.

1.2. Contributions

In this paper, we focus on classifying UDH from DCIS. Our approach is to first generate a small number of nuclei phenotypes by taking into account nuclear shapes, sizes, intensities and textures. Second, we assign each nucleus to an appropriate phenotype and build a graph connecting each nucleus to its neighbors (Figure 1). Third, we construct neighborhood statistics to capture the spatial distribution of heterogeneous nuclear phenotypes in a lesion. Finally, we cluster the heterogeneous cell distributions into a small number of architectural patterns. For illustration, in Fig. 2 we show examples of architectural patterns discovered in the Early Breast dataset [4] and how they can help distinguish UDH from DCIS.

2. MATERIALS AND METHODS

2.1. Materials

We develop our algorithm using the publicly available Early Breast dataset [4]. This dataset contains 116 cases (80 DCIS

978-1-5090-1172-8/17/$31.00 ©2017 IEEE

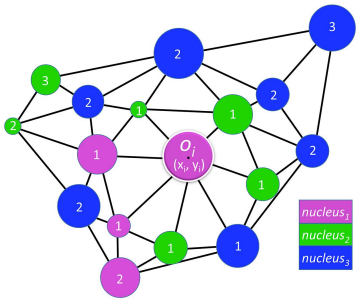

Fig. 1. Delaunay triangulation is used to generate a neighborhood graph for the nuclei on the input image. Each node is assigned to a particular nuclear phenotype (1, 2, or 3 by colors). The center of this neighborhood is nucleus o_i of phenotype 1. The number on each node corresponds to its depth from o_i as derived from a breadth-first traversal. The probability of finding a nucleus of phenotype 1 from the center nucleus o_i at depth level 1 is 28.5% $(= 2/7)$.

Hospitals	UDH	Low DCIS	Med DCIS	High DICS	Total
MGH	79	55	82	60	276
BIDMC	31	4	13	3	51

Table 1. Early Breast dataset [4]: Massachusetts General (MGH) and Beth Israel Deaconess Medical Center (BDMC).

and 36 UDH) from Massachusetts General Hospital (MGH) and 51 cases (20 DCIS, 31 UDH) from Beth Israel Deaconess Medical Center (BIDMC). The pathological grading was obtained from diagnostic pathology reports. Whole slide images were scanned using Philips Ultra Fast Scanner 1.6 (Philips Digital Pathology; Best, Netherlands) at 40$\times$ magnification with a resolution of 0.25μm per pixel. From whole slide image of each case, 1 to 4 diagnostic regions of interest (ROIs) were manually selected for image analysis, resulting in 327 images of various sizes (between 400 pixels and 2000 pixels per dimension). Nuclei masks were included in the dataset and taken as inputs for feature extraction in our paper. Data is summarized in Table 1. Since there is no example of normal tissue in the dataset, we are unable to build classifiers for normal tissue.

2.2. Building nuclei phenotypes

Since we would like to incorporate nuclear phenotypes in deriving architectural patterns, we assign each nucleus to an appropriate phenotype. For that, we compute nuclear features using the nuclei masks provided in [4]. These features include morphology features such as roundness, aspect ratio, bounding box dimensions; intensity features such as means, variance, skewness, and kurtosis in multiple color channels (RGB, HSV, La*b*); and texture features such as Haralick's features and graph run length features for each nuclei. These features were calculated with no additional preprocessing of the raw images. We observe three dominant phenotypes in this data, which we capture using a $k-$means clustering algorithm (in Matlab with $k-$means++ initialization and a warm start option). We speculate that the three dominant phenotypes are a consequence of normal nuclei in UDH, atypical nuclei in low grade DCIS, and pleomorphic nuclei in high grade DCIS.

2.3. Deriving architectural patterns from neighborhood statistics

The proposed method relies on modeling the spatial organization of various nuclei phenotypes within a tissue. Each image I is represented by a set of nuclei $O(I) = \{o_i\}$, where each o_i is represented by its center coordinates (x_i, y_i), together with its phenotype $t_i \in \{\text{nucleus}_1, \text{nucleus}_2, \text{nucleus}_3\}$. To compute the neighborhood statistics, we build a graph on the nuclei phenotypes by constructing a Delaunay triangulation on their center coordinates [6, 7] (Fig. 1).

Then we define a Breadth-First Traversal (B-FiT) for each nucleus o_i, by setting o_i as the root of tree and visiting its neighboring nuclei in breadth-first order using the edges of the generated graph. We set maximum depth for B-FiT to h hops from root. h is set to be small since DCIS nuclei are considered to show local properties such as pleomorphism and forming rigid arches with neighboring nuclei [3]. Next, for each depth level we compute the probabilities of finding each type of nuclei phenotype (Figure 1). Since we have three phenotypes, it makes six different types of edges for each depth level. As the result, for maximum depth of 10 hops, we have a set of 60 probability values describing the neighborhood statistics. In theory, these statistics which capture the spatial distribution of various nuclear phenotypes and their interactions are used to derive different architectural patterns.

Finally, we cluster the neighborhood statistics into q clusters to find representative architectural patterns. For that, the principal components of the training data is calculated first and q is selected such that it will cover 95% of the input variance ($q = 25$ in Fig.2). Following cluster center initialization, each of the nuclei neighborhoods is assigned to its closest cluster. For illustration, consider high grade DCIS in panels (D) and (G) of Fig. 2. We observe that the pattern #17 which is centered on red nuclei in (G) denotes a neighborhood where the center nucleus has a small area of highly concentrated chromatin and is surrounded by nuclei with uniform texture. Each image is represented by the % frequency of occurrence of the q patterns. These feature vectors are then used for classification.

2.4. Classification algorithm

For a fair comparison with C-path method [4], we used logistic regression with Lasso *L1* penalty to build both classifiers: binary (UDH vs. DCIS), and four-way (UDH and three

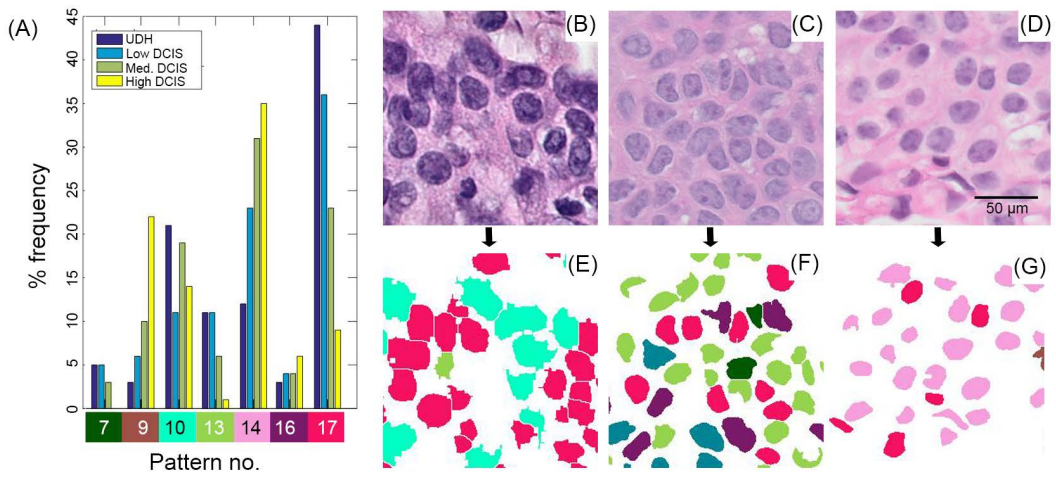

Fig. 2. Architectural patterns derived from nuclei phenotypes can help distinguish proliferative breast lesions. (A) shows four different histograms, each corresponding to a type of breast lesion, merged into one plot for illustration. The *y*-axis shows the % frequency of occurrence of each architectural pattern. In the interest of space only the dominant patterns among the $q = 25$ discovered patterns are shown here. Note that pattern #17 is dominant in UDH, while patterns #9 and #14 are dominant in high grade DCIS. For visualization, we pick the center nucleus for each pattern and paint it with the color shown below the *x*-axis. (B) and (E) show an example UDH, in which patterns #10 and #17 are the majority. (C) and (F) show an example of low grade DCIS, in which #13 and #17 are dominant. (D) and (G) show an example of high grade DCIS where a few of the pattern #17 neighborhood are observed among a large population of pattern #14 neighborhood.

grade of DCIS) [8]. We invoked nine-fold cross-validation to identify the optimal regularization strength from a grid of 50 values chosen on a log scale between $1e^{-4}$ and $1e^{4}$.

3. EXPERIMENTS AND RESULTS

3.1. Experiments

The Early Breast dataset is divided into a train/validation set from MGH and a test set from BIDMC. We trained the models and scanned their parameters on the MGH data and tested them on the BIDMC data. The three models we picked for comparisons are our Nuclei-Spatial (NS), Color-Spatial (CS), and C-path [4]. In [4], each image is characterized by the means and standard deviations of 196 nuclear features, including morphology, intensity, and texture features. These features are used as inputs to a *L1* logistic regression.

In Nuclei-Spatial (NS), we applied *L1* logistic regression on the architectural patterns derived from nuclei phenotypes as described in section 2.3. We tested multiple parameter settings for NS, including combinations of two B-FiT depths ($h = 5$ and $h = 10$) and two levels of variance coverage by principle components (90% and 95%). Among these four combinations, the best results were achieved for $h = 10$ with 95% variance coverage.

In Color-Spatial (CS), instead of deriving architectural patterns on nuclear phenotypes, we derived them on color objects (purple-nuclei, pink-stroma, and white-lumen) as described in [6]. While C-path focuses on nuclear morphology and CS focuses on spatial statistics, NS combines both types of features.

In addition, we also build a AlexNet type deep neural net to classify patches of size 128×128 from these images. However, the current deep net setup resulted in inferior classification performance compared to the other three methods and did worse than chance (<50%). This is potentially due to small number of training examples and thus we did not report these results in this preliminary paper. Since DCIS CAD [5] did not make their code and data available, we do not include their method in our comparisons.

3.2. Performance metrics

The performance of different models are assessed in terms of accuracy and areas under the curve (AUC) [8]. Accuracy is defined as the fraction of correct predictions in the test set. AUC is the area under the receiver operating characteristic (ROC) curve. In the binary classification (UDH vs. DCIS), ROC is created by plotting true positive rate (TPR) against the false positive rate (FPR) at different thresholds where DCIS is the positive class. In the four-way classification (UDH vs. three grades of DCIS), we computed both micro-AUCs and macro-AUCs. Micro-AUC gives each image-label pair an equal contribution to the overall metric, i.e., each pair is counted toward FPR and TPR equally. Macro-AUC calculates the mean of AUCs by treating the four-way classification as four binary classification problems. Better model has higher AUC and accuracy values.

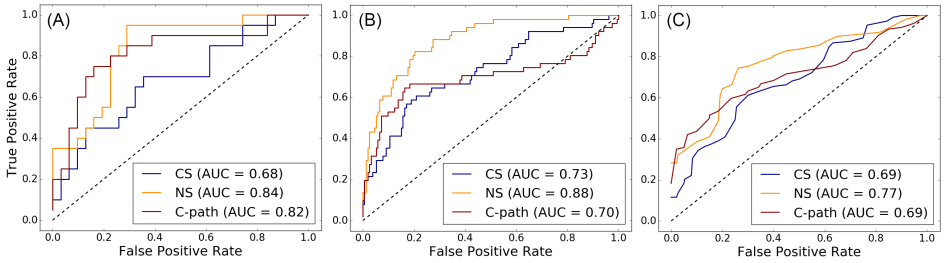

Fig. 3. Classification results. (A) shows AUC for binary classification. Accuracy values are 0.75 for NS, 0.53 for CS, and 0.65 for C-path. (B) and (C) show four-way classification micro- and macro-AUCs. Four-way accuracy values are 0.65 for NS, 0.51 for CS, and 0.59 for C-path. Typical errors by NS include classifying Low and Med DCIS as UDH.

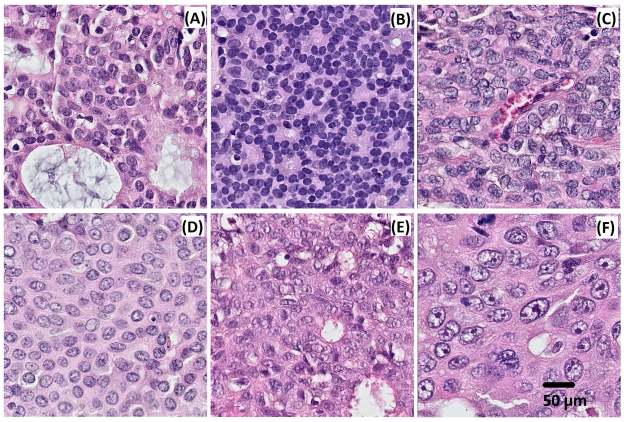

Fig. 4. Example classification results formatted as true class vs. predicted class. (A) Low DCIS vs. UDH, (B) Med DCIS vs. UDH, (C) UDH vs. UDH, (D) Low vs. High DCIS, (E) Med vs. High DCIS, and (F) High vs. High DCIS.

3.3. Results

NS outperforms C-path and CS on both the binary and four-way classification tasks. Note that NS uses both nuclear phenotypes and their spatial distributions, while C-path and CS focus on either one or the other element. This shows the benefit of combining morphometric properties of nuclear atypia and the spatial distributions of the nuclei. For the binary classification (UDH vs. DCIS), NS outperforms C-path by 10% in accuracy. For the four way classification (UDH, Low-DCIS, Medium-DCIS, High-DCIS), the difference is by 6% in accuracy, 8% in micro-AUC, and 19% in macro-AUC. The results are included in Fig. 3, panels (A)-(C). While the classifier NS performs well for UDH and High grade DCIS, it confuses Low and Med grade DCIS with UDH. Fig. 4 shows examples of images correctly and incorrectly classified by NS. In terms of computational speed, it takes 10-15 seconds to calculate the features for each image and roughly 5 seconds for logistic regression on the calculated features of all images.

4. CONCLUSIONS

Our goal in this paper was to build a more complete representation of histopathology images combining nuclear features and neighborhood statistics to discover architectural patterns for differential diagnoses of proliferative breast lesions. The features will have to be further enhanced to correctly identify Low and Med grade DCIS images. Another challenge is to scale our method to whole slide images and tackle the more difficult case of atypical ductal hyperplasia.

5. REFERENCES

[1] J.G. Elmore et al., "Diagnostic concordance among pathologists interpreting breast biopsy specimens," *JAMA*, vol. 313, no. 11, pp. 1122–1132, 2015.

[2] P.A. Carney et al., "Identifying and processing the gap between perceived and actual agreement in breast pathology interpretation," *Modern Pathology*, 2016.

[3] J.F. Simpson and F.I. Boulos, "Differential diagnosis of proliferative breast lesions," *Surgical Pathology Clinics*, vol. 2, no. 2, pp. 235–246, 2009.

[4] F. Dong et al., "Computational pathology to discriminate benign from malignant intraductal proliferations of the breast," *PloS One*, vol. 9, no. 12, pp. e114885, 2014.

[5] B.E. Bejnordi et al., "Automated detection of dcis in whole-slide h&e stained breast histopathology images," *IEEE-TMI*, 2016.

[6] A.B. Tosun and C. Gunduz-Demir, "Graph run-length matrices for histopathological image segmentation," *IEEE TMI*, vol. 30, no. 3, pp. 721–732, 2011.

[7] A.C. Simsek et al., "Multilevel segmentation of histopathological images using cooccurrence of tissue objects," *IEEE TBME*, vol. 59, no. 6, pp. 1681–1690, 2012.

[8] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *JMLR*, vol. 12, pp. 2825–2830, 2011.