Modeling Histological Patterns for Differential Diagnosis of Atypical Breast Lesions

Akash Parvatikar[1], Om Choudhary[1], Arvind Ramanathan[2], Olga Navolotskaia[3], Gloria Carter[3], Akif Burak Tosun[4], Jeffrey L. Fine[3], and S. Chakra Chennubhotla[1,4]

[1] Dept. of Computational and Systems Biology, Univ. of Pittsburgh, Pittsburgh, USA
{akp47,opc3,chakracs}@pitt.edu
[2] Data Science and Learning, Argonne National Laboratory, Lemont, Illinois, USA
ramanathana@anl.gov
[3] Dept. of Pathology, UPMC Magee-Womens Hospital, Pittsburgh, USA
{navolotskaiao, finejl}@upmc.edu, {cartgj}@mail.magee.edu
[4] SpIntellx Inc., Pittsburgh, USA
{burak, chakra}@spintellx.com

Abstract. Our goal in this paper is to build parametric models for a dictionary of histological patterns that aid in the differential diagnosis of atypical breast lesions and evaluate the inferential power of these hand-crafted features. Diagnosis of high-risk atypical breast lesions is challenging and remains a critical component of breast cancer screening, presenting even for experienced pathologists a more difficult classification problem than the binary detection task of cancer *vs* not-cancer. Following guidelines in the WHO classification of the tumors of the breast (an essential reference for pathologists, clinicians and researchers) and in consultation with our team of breast sub-specialists ($N = 3$), we assembled a visual dictionary of sixteen histological patterns (e.g., cribriform, picket-fence), a subset that pathologists frequently use in making complex diagnostic decisions of atypical breast lesions. We invoke parametric models for each pattern using a mix of *unary*, *binary* and *ternary* features that account for morphological and architectural tissue properties. We use 1441 ductal regions of interest (ROIs) extracted automatically from 93 whole slide images (WSIs) with a computational pathology pipeline. We collected diagnostic labels for all of the ROIs: normal and columnar cell changes (CCC) as low-risk benign lesions (=1124), and flat epithelium atypia (FEA) and atypical ductal hyperplasia (ADH) as high-risk benign lesions (=317). We generate likelihood maps for each dictionary pattern across a given ROI and integrate this information to determine a diagnostic label of high- or low-risk. Our method has comparable classification accuracies to the pool of breast pathology sub-specialists. Our study enables a deeper understanding of the discordance among pathologists in diagnosing atypical breast lesions.

Keywords: Atypical breast lesions · Visual feature dictionary · Parametric models · Computational pathology

1 Introduction

Benign breast lesion diagnoses account for approximately a million cases annually [1]. The patients are subjected to additional screening procedures depending upon the relative risk associated with the diagnostic subtypes of the benign lesions (e.g., high-risk is associated with atypical hyperplasia) [2,3]. Over half of the patients diagnosed with atypical hyperplasia, which is histologically further classified into atypical ductal hyperplasia (ADH) and atypical lobular hyperplasia (ALH), contract breast cancer within 10 years of screening, thereby demanding an accurate diagnosis of these precursor lesions.

On the contrary, a recent clinical study showed significant levels of disagreement in differential diagnosis of cases with atypia (48 - 56%) resulting in *overinterpretation* (subjecting patients to unnecessary medical procedures) and *underinterpretation* (subjecting patients to no treatment) [4]. The underlying difficulty in classifying atypia from benign lesions stems from the fact that diagnostically relevant histopathological patterns overlap in the spectrum of low- to high-risk lesions, complicating the decision making process (Fig. 1). In the interest of patient management, it is convenient to stratify patients into "low- / high-risk" categories based on their histological evidence and associated risk-factor [3].

Our Approach: Following guidelines in the WHO classification of the tumors of the breast [5] (an essential reference for pathologists, clinicians and researchers) and in consultation with our team of breast pathology sub-specialists ($N = 3$), we assembled a visual dictionary of a *subset* of histological patterns that aid pathologists in undertaking differential diagnoses of atypical breast lesions (Fig. 1). Our goal in this differential diagnostic study is to build parametric models for each pattern using a mix of unary, binary and ternary features that account for cytological (nuclear shape and orientation, lumen shape), architectural (intraductal), and spatial-extent details of low- and high-risk lesions.

Prior Work: Previously, we have approached this problem in an unsupervised manner by simply encoding cytological properties of nuclear atypia and integrating them with the spatial distribution of the nuclei in relationship to stroma and lumen components of breast tissue (i.e., architectural patterns) [6]. Measured in terms of *recall* of high-risk lesions, *the classification performance reported here (0.76) is a significant improvement over our previous approach (0.69).* Although there are studies in the machine classification of breast tumors [7–10], many of these do not include diagnostically challenging ADH cases nor provide directions for a *computational* understanding of the structural changes in the breast tissue triggered by atypia and other malignancies. To the best of our knowledge, our work in analytically modeling a visual pattern dictionary that traditionally defines the standards on tumor classification/ nomenclature for pathologists worldwide is the first of its kind.

2 Methodology

2.1 Segmenting Ducts, Lumen and Nuclei

We designed a new algorithm for segmenting ducts, lumen and nuclei on large scale WSIs. To start with, WSI images stored in RGB format are color deconvolved into their respective stain intensities namely, hematoxylin and eosin by using the color deconvolution plugin in ImageJ [11]. The stain colors are further normalized with a reference dataset to standardize color variations for downstream processing (see supplementary material for more details). To ease the computational burden of detecting ducts in a WSI, we build a Gaussian pyramid of the hematoxylin intensity WSI. The hematoxylin intensity image at the coarsest level of the pyramid is broken into non-overlapping superpixels (area = 300 pixels), which are sets of connected pixels with similar intensity values, using simple linear iterative clustering (SLIC) algorithm [12]. The innovative steps of our algorithm are in assigning probabilities for the presence of a duct given a pair of nearby superpixels ("context- ML") and further identifying all those superpixels that are "moderate-to-heavily" stained as the ones inside a duct ("stain-ML"). Using the superpixels identified as initial guesses, we perform a region-based active contour segmentation [13] that separates foreground (ducts/lumen) from the background (rest of the image). For hematoxylin and eosin stained images, the cost-function for the active contour is driven by the difference in the mean of the hematoxylin stain in the foreground and background regions. For example, two superpixels that have a high probability of being inside a duct have roughly the same stain ("moderate to heavy stain") and their boundaries are merged iteratively by the active contour optimization. Often ducts appear as "clusters" and to segment these we run the region-based active contour on the *probability* map returned by the context and stain-ML models. The probability maps impute non-zero probabilities to ducts and regions bridging them, and a region-based active contour model run on the probability map is more successful in delineating a cluster of ducts.

To identify lumen we use context- and stain-based ML models to select image regions that are not part of the ducts – non-tissue areas on the WSI, connective tissue areas and lumen. We perform connected-component analysis to select and exclude large components, likely to correspond to non-tissue and connective tissue areas. The remaining components highlight lumen regions that lie inside ducts and are verified visually in our training images. To identify and segment nuclei inside a duct, we first select parts of image lying inside a duct, then use ImageJ to threshold intensities and finally run watershed to delineate the nuclear boundaries (see supplementary material for more details).

2.2 Building Parametric Models of Histological Patterns

We invoke parametric models for histological patterns using a mix of *unary*, *binary* and *ternary* features as shown in Fig. 1. The colorbars over each feature in Fig. 1 indicate the lesion where the feature is most likely to be found, e.g.,
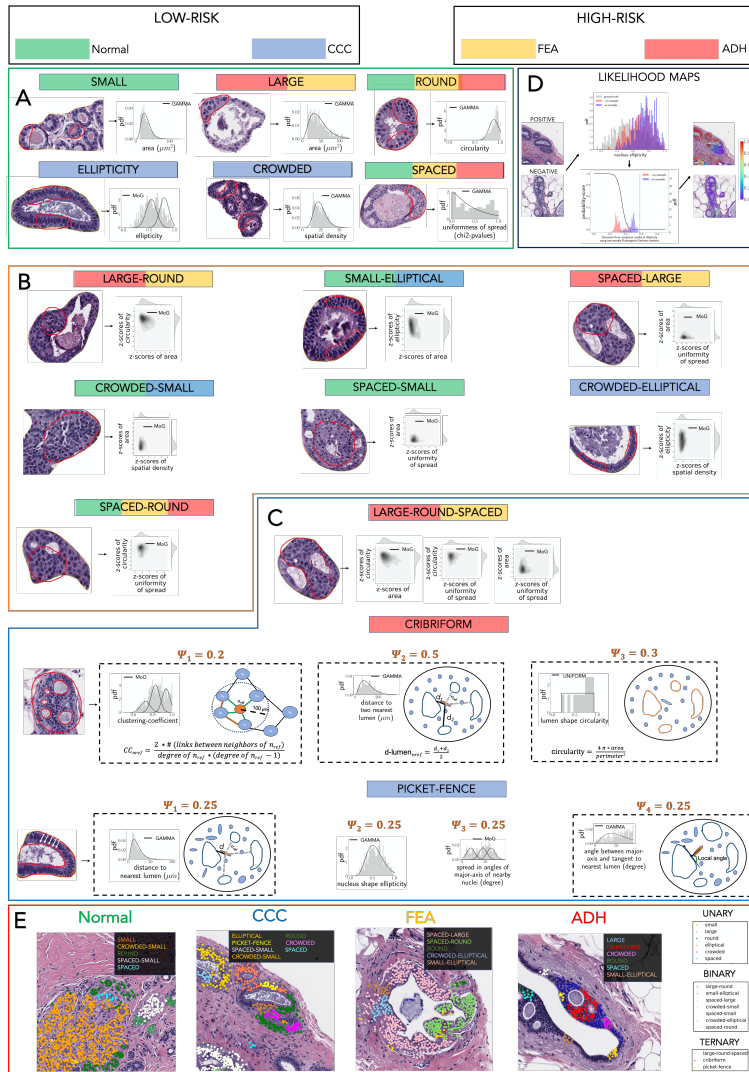
Fig. 1: Parametric models of histological patterns in the form of (A) unary, (B) binary and (C) ternary features. (D) Computing likelihood scores to reveal (E) dominant patterns in representative images of low- and high-risk lesions.

large and round nuclei are often found in high-risk lesions, small and elliptical nuclei in low-risk lesions, and cribriform pattern tends to be exclusive to ADH.

Unary Features: In consultation with the breast pathologists on our team, we selected a spectrum of morphological features on the basis of size, shape, and spatial spread around each nucleus. Nuclear size (quantified using area) is known to provide diagnostic cues in pathological grading [14–17], with groups of small

and large nuclei having a propensity to belong to low-risk and high-risk lesions respectively [18]. To build analytical models of *small* and *large*, we first construct a histogram of nuclear areas obtained from an ensemble of ROIs showing prototypical example regions within a duct containing small and large nuclei (Fig. 1A) and model this histogram with a Gamma distribution (see supplementary material for a listing of all the model parameters derived in this section).

Next, nuclear shape has been identified as diagnostically meaningful, e.g., CCC lesion shows dominant elliptical nuclei [19]. We quantitate this feature with *roundness* measured as $(4\pi \times \text{area})/\text{perimeter}^2$ and *ellipticity* given by the ratio of length of minor-axis to the length of major-axis. Roundness ranges from 0 (irregular star-like appearance) to 1 (perfect circle), while ellipticity characterizes the "flatness" of an object with lower values denoting highly elliptical nuclei (Fig. 1A). In each case, because of the intrinsic heterogeneity of these measurements, we consider a spatial neighborhood around each nucleus, and model the distributions of roundness with a Gamma distribution and ellipticity with a 2-component mixture of Gaussians (MoG) model (Fig. 1A).

Finally, several studies have shown that studying the spatial organization of nuclei provides insights into the abnormalities of cells which might eventually lead to malignancy. For instance, the nuclei arrangement in a CCC lesion frequently exhibits crowding and/or overlapping [20, 21].

However, for cases belonging to high-risk atypical lesions (FEA and ADH) the nuclei tends to be uniform and evenly-spaced [20, 22]. To quantify "crowding" around each nucleus, its average distance to 10 nearest nuclei is computed. An analytical model of *crowdedness* is constructed by considering local ROIs within a duct where clusters of nuclei show significant crowding behavior and then computing its spatial density. To capture evenly spaced/ uniform dispersion pattern around a nucleus, we start by placing a regular grid of size 3×3 centered at a reference nucleus and measure the density of 20 neighboring nuclei by counting the population of nuclei in each grid cell as described in [23]. We then compare this observed population against expected number of nuclei under the *complete spatial randomness* hypothesis which asserts the occurrence of points (here nuclei) within grids in a random fashion analogous to a Poisson point process using a χ^2-test statistic and acquiring the corresponding p-value using the χ^2 distribution table. Larger the p-value, greater is the likelihood of observing a uniform/ evenly spaced dispersion of nuclei around the reference nucleus.

Binary Features: Although, the unary features show some inferential strength (indicated by the color bars on top of each feature in Fig. 1), a pathologist typically makes an informed decision by paying attention to the pairwise combinations of such features. For instance, a CCC lesion (low-risk) exhibits crowded and elliptical nuclei arrangement. A high-risk lesion tends to display a greater likelihood of large-round, spaced-large, and spaced-round nuclei. A lesion showing majority regions of small nuclei coupled with crowded and/ or spaced behavior is representative of a normal duct. In our study, we considered 7 such binary features obtained from pairwise combinations of unary features which

is shown in panel Fig. 1B. We take z-scores for each unary feature, and model the joint distribution of z-scores from the feature pair with a two-component, two-dimensional mixture of Gaussian distribution.

Ternary Features: Some of the diagnostically relevant histological patterns are best represented by a combination of more than two unary features. *I. Large-Round-Spaced*: We take z-scores from each feature, i.e., large, round and spaced, and build a three-component, three-dimensional mixture of Gaussian model using ground-truth examples. *II. Cribriform*: This pattern is characterized by polarization of epithelial cells within spaces formed by "almost" circular multiple lumen (> 2) which are 5-6 cells wide and whose appearance closely resembles to "holes in Swiss cheese". This complex architectural pattern can be identified by analytically modeling three sub-features: *clustering coefficient*, distance of the nucleus from two nearest lumen, and circularity of the lumen (computed using ImageJ) adjacent to the nucleus. The polarization of epithelial cells around lumen is characterized by clustering-coefficient and is computed by following the method described in [24] and is illustrated in the second row of Fig. 1C. A group of nuclei occupying the spacing between two lumen has a tendency to show cribriform pattern around them. Thus, we measure the average distance between each nucleus to the nearest two lumen and model its distribution using gamma function (see middle row of Fig. 1C). The final likelihood for cribriform pattern is obtained from the weighted sum of the likelihood scores of sub-features. We performed grid search on the mixing coefficients to learn that the likelihood scores from the three sub-features should be mixed in the proportion of 0.2, 0.5, and 0.3 respectively. *III. Picket-Fence*: This pattern is recognized from a group of crowded elliptical nuclei oriented perpendicular to the basement membrane (lumen). The analytical model of this high-order visual feature can be obtained by constructing parametric models of four simple sub-features: distance of a nucleus to nearest lumen, nuclear ellipticity, a spread in the angle of major-axis of 10 nearby nuclei, and its local angle with respect to the basement membrane as shown in the last row of Fig. 1C. Since, each sub-feature contributes equally to observing this ternary feature, we chose to assign a mixing coefficient of 0.25 in combining the likelihood scores from the four sub-features to determine the presence of a picket-fence pattern.

2.3   Computing Likelihood Scores to Assigning Diagnostic Labels

Computing Likelihood Scores: As discussed in the previous section, the parametric models for the histological patterns are probability distributions. For example, a cytological feature like nuclear ellipticity for a given nucleus inside a ROI can be assigned a probability value under the mixture of Gaussian model for the template ($\mathcal{G}_t$) image derived in Fig. 1A. However, accurate measurements of ellipticity values are greatly influenced by the precision with which nuclei boundaries are segmented. This naturally leads to heterogeneity in the estimates of ellipticity. To account for this heterogeneity, we chose to compare the neighborhood around the reference nucleus to the neighborhood in the template image. In particular, we model the ellipticity values in the neighborhood

of the reference nucleus with a new mixture of Gaussian model ($\mathcal{G}_n$), just like in Fig. 1A, and then compare model parameters of $\mathcal{G}_n$ with $\mathcal{G}_t$ as shown in Fig. 1D. We used two different distance measures for comparing the model parameters: *Kullback-Leibler divergence* for mixture of Gaussians and *two-sample Kolmogorov Smirnov test* for unimodal Gamma distributions. Small distances imply greater evidence for the pattern. We turn the distances into a likelihood score by an inverted S-function as shown in Fig. 1D. This process is carried out in a similar fashion for every feature present in the visual dictionary.

Strategy for Differential Diagnosis: We adopt a non-linear strategy here, similar to what expert pathologists do, in that we find sub-regions within ROI by non-maxima suppression (threshold value of 0.85 on the likelihood scores) where the evidence for one or more of the unary, binary or ternary feature is dominating. Fig. 1E provides a visual illustration of the likelihood maps of dominant patterns in representative images of low- and high-risk lesions. Low-risk lesions show dominant islands of *round*, *small*, *spaced*, and *spaced-small* in a normal ROI and *elliptical*, *round*, *spaced-small*, *crowded-small*, and *picket-fence* neighborhoods in a CCC ROI. In comparison, high-risk lesions show dominant regions of *spaced-large*, and *spaced-round* in a FEA labeled ROI and compelling strengths for *large* and *cribriform* patterns along with traces of *crowded* and *spaced* in ADH labeled ROI. These patterns validate the canonical forms shown in Fig. 1A-C.

Having identified dominant unary, binary and ternary feature regions, we use 3 descriptive statistics: median value of the likelihood scores of all the nuclei found in each sub-region, median number of nuclei found in each sub-region and the number of sub-regions.

This is calculated for each one of the unary, binary and ternary features (total = 16), thereby obtaining a 48 column feature vector for a single image. We computed feature vectors for all 1441 labeled duct ROIs which resulted in 834×48 size feature map used to train the classifier and 607×48 data matrix for testing. To analyze the benefit of including binary and ternary features we further slice the 48 column feature vector to be suitable for three scenarios: unary (U) only, unary and binary (U-B), and unary, binary, and ternary features (U-B-T). Due to inherent training and testing class imbalance, which reflects the real-world prevalence statistics of atypical lesions, we upsampled high-risk examples using SMOTE technique [25].

Prior to classifying the lesions, we pay close attention to the presence of cribriform pattern, a symbolic visual primitive of ADH (a high-risk) category [22, 26, 27]. ROIs predicted to show cribriform pattern are classified as high-risk, if the number of nuclei forming the cribriform sub-region is greater than 8 (hyperparameter optimized over the training data). The reduced dataset, devoid of cribriform, is tested for each of the scenarios (U, U-B, and U-B-T) with logistic regression (LR), support vector machine (SVM), random forest (RF), and gradient boosted classifier algorithms. The best model was chosen by optimizing the parameters using GridSearchCV based on precision, recall, and F-scores and then performed a 10-fold stratified cross-validation to check for overfitting. In

optimizing the hyperparameters, the operating point was selected to value recall over precision reflecting the clinical decision objective where a false negative outcome is penalized higher than a false positive.

3 Results and Discussion

Models	Baseline		Comparisons			U	U-B	U-B-T
	Majority	Expert	Lenet	Overfeat	Alexnet	LR		
Recall	0	**0.77**	0.23	0.31	0.4	0.56	0.59	**0.76**
Specificity			0.88	0.84	0.86	0.64	0.69	0.63
TN			475	451	462	345	373	336
FN			54	48	42	31	29	17

Table 1: Performance measures with U, U-B and U-B-T feature sets and comparison with other baseline strategies (including majority classification and average single expert pathologist assessment) and deep-learning models.

Dataset: We used 1441 ductal ROIs extracted automatically from a computational pathology pipeline (see Section 2.1) from 93 WSIs which were scanned at $0.5 \mu m$/pixel resolution at $20\times$ magnification captured using Aperio ScanScope XT microscope. Among these, the *training* set constituting 834 ROIs were diagnostically labeled by a single sub-specialist pathologist (SP1), while a consensus diagnosis was achieved for the remaining 607 *testing* set ROIs with a pool of 3 breast pathology sub-specialists (SP1, SP2, and SP3). The diagnostic labels include: "Normal", "CCC", "FEA", or "ADH", which were further regrouped into two classes: low-risk (Normal and CCC) and high-risk (FEA and ADH). While the training set comprised of 587 low-risk and 247 high-risk examples, the test set included 537 low-risk and only 70 high-risk cases, leading to the issue of class-imbalance and the choice of *recall* of high-risk lesions as a performance metric for the classification strategy. We are reporting recall to emphasize correct detection of high-risk lesions, as the consequence of misdiagnoses (false negative) implies increased chance of developing cancer for lack of providing early treatment. The concordance among the 3 pathologists in labeling the test set was moderate (Fleiss' kappa score of ≈ 0.55 [6]).

Results: Table 1 shows the outcome of the differential diagnosis strategy that we implemented using the three feature sets: U, U-B, and U-B-T. The average performance of the three pathologists informs the baseline with single expert pathologist [6]. We tested with Logistic Regression (LR), Random Forest, and SVM with SMOTE and cross-validation parameter scanning. LR performed the best. SVM and Random Forest misclassified high-risk images containing large/round/spaced nuclei (a high-risk feature, see Fig. 1) as low-risk. This resulted in lower recall compared to LR, which was successful in capturing these features. Additionally, we tested approaches with deep learning: Lenet [28], Alexnet [29], and Overfeat [30]. For training deep learning networks the ROIs

obtained from duct segmentation were downscaled to 512×512 and the class imbalance was handled by performing data augmentation through rotations and reflections. Further, these class-balanced batches were trained using 3 networks for 3,000 epochs.

Discussion: We find progressive improvement in the performance from U to U-B to U-B-T feature sets, achieving highest recall of 0.76 which outperforms the majority classification (obtained by assigning all cases to the majority label of low-risk, thereby having a recall of 0) and has a comparable performance to the assessment made by single breast pathology sub-specialist (SP1). Our approach with ~150 parameters (see supplementary material) is readily amenable to explainability which cannot be delivered by current deep learning (DL) methods (~10-50 million parameters and large training data). To the best of our knowledge, there are no widely reported DL methods for borderline of atypical breast lesions, but an abundance of these algorithms for cancer vs no-cancer datasets. To further promote research in the use of DL for borderline cases, we chose to continue working with the same set of networks as used in our previous work [6], with one exception of incorporating improved duct segmentation component (see supplementary material). The average computation time to obtain likelihood scores and return a diagnostic label is 1 minute for an image with 1000 nuclei on a single 2.4 GHz processor. In some low-risk examples, the accurate identification of class specific histological patterns (e.g. small, crowded-small) is missed due to the under-segmentation of overlapping nuclei resulting in a wrong classification (false positive). However, we observed that U-B-T features (best recall) misclassified 24% of the high-risk images as low-risk (false negative). Upon investigation, we found that majority of the wrongly classified images had rigid cellular bars and micropapillae (club-shaped lumina) architecture, two additional distinguishing characteristics of ADH [5] not included in the dictionary for the present study. As future work, we anticipate successful inclusion and analytical modeling of the missing patterns from the WHO breast cancer resource, to further reduce the number of false positives and false negatives. Our strategy has the potential to extend to other organ systems and act as a surrogate in the case review and quality assurance discussions for reducing discordance between pathologists.

References

1. Figueroa, J. D., et al.: Standardized measures of lobular involution and subsequent breast cancer risk among women with benign breast disease: a nested case–control study. Breast Cancer Research and Treatment **159**(1), 163–172 (2016)

2. Santen, R. J.: Benign breast disease in women. In Endotext [Internet]. MD-Text.com, Inc. (2018)

3. Dyrstad, S. W., et al.: Breast cancer risk associated with benign breast disease: systematic review and meta-analysis. Breast Cancer Research and Treatment **149**(3), 569–575 (2015)

4. Elmore, J. G., et al.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA **313**(11), 1122–1132 (2015)

5. Lakhani, S. R.: WHO Classification of Tumours of the Breast. International Agency for Research on Cancer (2012)

6. Tosun, A. B., et al.: Histological detection of high-risk benign breast lesions from whole slide images. In International Conference on MICCAI, pp. 144–152. (2017)

7. Mercan, E., et al.: Assessment of Machine Learning of Breast Pathology Structures for Automated Differentiation of Breast Cancer and High-Risk Proliferative Lesions. JAMA Network Open **2**(8), e198777 (2019)

8. Bejnordi, B. E., et al.: Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. J Med Imaging (Bellingham) **4**(4), 044504 (2017)

9. Li, H., et al.: Quantitative nuclear histomorphometric features are predictive of Oncotype DX risk categories in ductal carcinoma in situ: preliminary findings. Breast Cancer Res. **21**(1), 114 (2019)

10. Dong, F., et al.: Computational pathology to discriminate benign from malignant intraductal proliferations of the breast. PLoS ONE **9**(12), e114885 (2014)

11. Schindelin, J., et al.: Fiji: an open-source platform for biological-image analysis. Nature Methods **9**(7), 676–682 (2012)

12. Achanta, R., et al.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell **34**(11), 2274–2282 (2012)

13. Chan, T. F., et al.: Active contours without edges. IEEE Trans Image Process **10**(2), 266–277 (2001)

14. Pienta, K. J., et al.: Correlation of nuclear morphometry with progression of breast cancer. Cancer **68**(9), 2012–2016 (1991)

15. Cui, Y., et al.: Nuclear morphometric features in benign breast tissue and risk of subsequent breast cancer. Breast Cancer Research and Treatment **104**(1), 103–107 (2007)

16. Kashyap, A., et al.: Study of nuclear morphometry on cytology specimens of benign and malignant breast lesions: A study of 122 cases. Journal of Cytology **34**(1), 10 (2017)

17. Narasimha, A., et al.: Significance of nuclear morphometry in benign and malignant breast aspirates. International Journal of Applied and Basic Medical Research **3**(1), 22 (2013)

18. Mommers, E. CM., et al.: Prognostic value of morphometry in patients with normal breast tissue or usual ductal hyperplasia of the breast. International Journal of Cancer **95**(5), 282–285 (2001)

19. Yamashita, Y., et al.: Does flat epithelial atypia have rounder nuclei than columnar cell change/hyperplasia? A morphometric approach to columnar cell lesions of the breast. Virchows Archiv **468**(6), 663–673 (2016)

20. Logullo, A. F., et al.: Columnar cell lesions of the breast: a practical review for the pathologist. Surgical and Experimental Pathology **2**(1), 1–8 (2019)

21. Pinder, S. E., et al.: Non-operative breast pathology: columnar cell lesions. Journal of Clinical Pathology **60**(12), 1307–1312 (2007)

22. Allison, K. H., et al.: Histological features associated with diagnostic agreement in atypical ductal hyperplasia of the breast: illustrative cases from the B-Path study. Histopathology **69**(6), 1028–1046 (2016)
23. Sergio R., et al.: pysal/pointpats: pointpats 2.1.0. https://doi.org/10.5281/zenodo.3265637, (2019)
24. Zhou, N., et al.: Large scale digital prostate pathology image analysis combining feature extraction and deep neural network. arXiv:1705.02678 (2017)
25. Chawla, N. V., et al.: SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002)
26. Ely, K. A., et al.: Core biopsy of the breast with atypical ductal hyperplasia: a probabilistic approach to reporting. The American Journal of Surgical Pathology **25**(8), 1017–1021 (2001)
27. Chen, L., et al.: Diagnostic upgrade of atypical ductal hyperplasia of the breast based on evaluation of histopathological features and calcification on core needle biopsy. Histopathology **75**(3), 320–328 (2019)
28. LeCun, Y., et al.: LeNet-5, convolutional neural networks. http://yann.lecun.com/exdb/lenet **20**, 5 (2015)
29. Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. In: NeurIPS, pp. 1097–1105. (2012)
30. Sermanet, P., et al.: Overfeat: Integrated Recognition, Localization and Detection using Convolutional Networks. In ICLR, CBLS. (2014)