

**Classifying Nuclei Shape Heterogeneity in Breast
Tumors with Skeletons**

by

Brian Falkenstein

Submitted to the Graduate Faculty of
the School of Computing and Information in partial fulfillment
of the requirements for the degree of
Master of Sciences

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH
SCHOOL OF COMPUTING AND INFORMATION

This thesis was presented

by

Brian Falkenstein

It was defended on

April 17 2020

and approved by

Adriana Kovashka, Department of Computer Science

Seong Jae Hwang, Department of Computer Science

Chakra Chennubhotla, SpIntelx Inc.

Thesis Advisor: Adriana Kovashka, Department of Computer Science

Classifying Nuclei Shape Heterogeneity in Breast Tumors with Skeletons

Brian Falkenstein, M.S.

University of Pittsburgh, 2020

In this study, we demonstrate the efficacy of medial axis derived scoring statistics for differentiating tumor and non-tumor nuclei in malignant breast tumor histopathology images. Characterizing nuclei shape is a crucial part of diagnosing breast tumors, and these scoring metrics may be integrated into algorithms which aggregate nuclei information across a region to label whole breast lesions. Nuclei and region scoring algorithms such as the one presented here can aid pathologists in the diagnosis of breast tumors.

Table of Contents

1.0 Introduction	1
2.0 Methods	3
2.1 Data	3
2.2 Data Processing	3
2.3 Baselines	4
2.4 Medial Axis Transform Scoring	5
3.0 Results	8
4.0 Conclusions	10
Bibliography	12

1.0 Introduction

Recent advancements in whole slide imaging technology have paved the way for what has been termed Digital Pathology, the digitization of pathology data. Once a biopsy sample has been processed, it can then be scanned into a digital format for viewing by the pathologist. This allows for viewing by the pathologist at any time, while not effecting diagnostic accuracy [Sha+16]. Additional benefits of Digital Pathology include the ability to store large amounts of cases for teaching and research purposes, as well as facilitating telepathology, allowing multiple pathologists to view cases simultaneously and remotely. However, what we are most interested in is how the digitization of pathology data allows for the use of computational algorithms to analyze the data and provide useful information to the pathologist.

Slices of the extracted tissue sample are scanned into WSI's, or whole slide images. These images are far too large and high resolution (around 5 gigapixels) to perform most analytical computation on, so ROI's, or regions of interest, may be extracted via a plethora of methods[Kot+13]. These smaller sub-regions of the whole slide (usually around 1M pixels) are likely to contain information useful to the diagnosis, while ignoring large parts of the image which are not useful (such as white space). Most useful computational analysis is done on the region level. With these manageable ROI's, we can employ a variety of machine learning and computer vision techniques to guide pathologists in the decision making process.

This study focuses on breast tumor analysis in particular for several reasons. Breast cancer is the second most common form of cancer found in women in the US, with an estimated 12% of women developing invasive breast cancer over the course of their life[Bre20]. Annually, roughly 1.6 million breast biopsies are conducted, with around 25% of them showing malignancy[Elm+15]. The remaining cases may be classified along a spectrum of labels, from fully benign up to a very high risk of becoming malignant in the future. The differences in appearance between these classes of pre-malignant tumors can be incredibly subtle, which is reflected in the high rates of discordant diagnoses among pathologists for those high-risk cases[Elm+15]. Further, accurate diagnosis of these lesions is crucial in providing optimal treatment, and early detection and treatment reduces risk of future malignancy[Har+15;

CLW20; Kad+18].

This begs the question: what are pathologists looking for when they are making these diagnoses? Distinguishing between invasive and non-invasive cancer is relatively easy compared to grading high-risk lesions, with discordance rates among invasive cases being as low as 4% [Elm+15] (compared to 52% for high risk benign lesions). However, differentiating between a benign lesion and a low or high risk lesions requires analysis of subtle structural changes in the breast tissue. Complex, high-level structural changes occur in the tumor region as it progresses along the spectrum from benign to malignant [Hal19]. High-risk lesions often show breast ducts being crowded with nuclei in rigid patterns around the lumen (interior opening of the duct). Even the smallest structures present in the tissue, the cell nuclei, demonstrate a change in shape across the disease spectrum.

Higher risk lesions, like FEA (flat epithelial atypia) have rounder epithelial nuclei than lower risk lesions like CCC (columnar cell change) [Yam+16]. ADH (atypical ductal hyperplasia) is one of the highest risk tumors. Epithelial nuclei in ADH tumors are typically completely rounded [PR92]. To get a full picture of the tumor environment in order to make an accurate prediction, one must consider not only the distribution of structures across the image (how dense the nuclei are crowded and in what pattern), but also the shape of the structures themselves, in particular the nuclei. That is the focus of this study, to find an accurate and robust way of characterizing the shape of epithelial nuclei in breast tumors.

2.0 Methods

2.1 Data

Due to the lack of annotated nuclei datasets for high-risk benign lesions, the study is conducted on a dataset of labeled tumor and non-tumor nuclei, particularly *BreCaHAD*[Aks+19]. It consists of 162 breast lesion images, all of which show varying degrees of malignancy. Each nuclei in the images is labeled by a point which is its centroid in the image, and one of the following class labels: mitosis, apoptosis, tumor nuclei, non-tumor nuclei, tubule, and non-tubule. For our purposes, we were just interested in differentiating the tumor and non-tumor nuclei. Theoretically, if we can design a framework to differentiate tumor and non-tumor nuclei in malignant cases based on shape, the same framework should be successful at analyzing high-risk benign tumor nuclei, the more challenging task.

Some issues arise with the dataset. Most glaringly is the fact that all of the cases are malignant. Ideally, we would be studying tumors across a wide distribution of disease levels, including benign and high-risk benign. Similarly, although we know the cases are malignant, there is no fine-grained label for each image. That is, we do not know the grade of the malignancy in the images, and it can vary across the samples. The fact that all tissue samples are malignant naturally leads to a dramatic class imbalance in the data, with almost 92% of nuclei labeled as 'tumor'. Additionally, the ground truth labels (centroid and class label) are not enough for our purposes, so further processing must be done before testing (segmentation, feature extraction).

2.2 Data Processing

If we wish to analyze the shape of the nuclei present in the biopsy images, we need some definition of the boundaries of the nuclei. Ideally, this would come with the dataset, and be a ground truth value. However, in our case, we are only given centroid locations of

the nuclei, and thus must produce our own segmentation of the nuclei.

Many methods exist for segmenting nuclei in histopathology images, however the best results can be gained from using deep learning techniques. Clearly, if our interest is studying the shape of the nuclei in the images, we would desire the best possible segmentation so that our shape representation of the nuclei is as close to reality as possible. For this study, I used the segmentation algorithm proposed by Chen et al.[Che+19], which is a fully convolutional network based off of U-Net[RFB15]. It works similarly to how a standard fully convolutional network does for semantic segmentation, just with some additional smoothing applied after the upsampling.

Once the segmentation was computed, the labeled centroids could be matched to their corresponding nuclei boundary by determining which (masked) boundary the centroid fell within. The labels in the dataset could then be associated with a nuclei boundary. Using this segmentation scheme, 80.18% of the ground truth nuclei in the dataset were found (18149 nuclei). Of these, 92% were positive (16689). Crops of size (81, 81) centered at each nuclei were taken from the original images, and used as the training and testing data.

To account for the class imbalance, training and testing sets were obtained by randomly sampling with replacement from the over-represented class (tumor) until there was an equal number of samples for both classes. Without this step, the model would tend to demonstrate very low specificity (high rate of false positives), while still maintaining a low loss. There were 1460 negative (non-tumor) nuclei after segmentation, so the same number of positive nuclei were randomly selected from the initial set of 16689. This is not ideal, as many negative samples are thrown out, however it is preferable to an overfitted model.

2.3 Baselines

In order to show the efficacy of the proposed method for capturing nuclei shape, the tests will be conducted on features with no information of texture or color in the nuclei. For this reason, I will be testing my method against another method which does not incorporate texture, but has shown to be successful. Yamamoto et al.[Yam+17] obtained an accuracy

of 85% using only a small number of hand-picked morphological features and an SVM. I was able to replicate their accuracy results on the balanced test set using the following set of features: area, eccentricity, extend, major axis, minor axis, convex area, circularity, equivalent diameter, filled area, perimeter, and solidity. Note again how none of these features incorporate any color or texture information.

However, color and texture features have also been used with success[Wan+16]. For this reason, I will also be comparing the proposed shape-based method to a method which includes color, particularly a small CNN. The CNN used was AlexNet[KSH12], pre-trained on ImageNet, and the features were extracted from the last fully-connected layer ($D=1000$). The CNN was tested both on the full color crops, as well as the 3-channel skeleton score image (see below).

2.4 Medial Axis Transform Scoring

Using the medial axis transform (MAT) for the nuclei shape analysis is the novel idea proposed in this study. The MAT is a skeletonization algorithm wherein a closed binary shape is reduced down to a 1 pixel wide branching structure wholly contained within the original shape. This may be conceptualized as a thinning process, where the border of the shape becomes thinner, or erodes, until opposite sides of the shape come together, defining the skeleton. The skeleton points are center points of maximally inscribed discs, circles with more than one point on their surface tangent to the shape boundary. This property makes the MAT skeletonization an interesting candidate for deriving a shape descriptor.

However, the skeleton itself does not contain enough information on its own, but can be used further to derive several scoring statistics. Rezanejad et al.[Rez+19] derive 3 scores from the MAT: ribbon, taper, and separation. They also show that these scores can be successfully used by a CNN to perform scene classification from line images (no color or texture).

Let p be a point on the skeleton, $R(p)$ be the radius function for that point (distance from skeleton point to boundary), and $[\alpha, \beta]$ be a range of points of size k . Finally, let $p \in [\alpha, \beta]$.

Then the three score metrics are defined as follows:

$$S_{ribbon}(p) = R'(p)$$

$$S_{taper}(p) = R''(p)$$

$$S_{separation}(p) = 1 - \int_{\alpha}^{\beta} \frac{1}{R(p)} dp/k$$

Because pixels are discrete, and thus so are the intervals we are integrating and deriving over, we must use numerical gradients. A small value of k was used, as the nuclei are fairly small (roughly 100 pixel perimeter), and the scoring metrics are designed to capture local symmetry. For all of the tests, I used a value of $k=8$.

The ribbon score captures the degree of parallelity of the surrounding contours, and increases as they become more parallel. Taper score is designed to increase as contours resemble the shape of a funnel or railroad tracks, but also has a high value for parallel contours. Separation captures the degree of separation between the contours, and increases with distance.

These algorithms define 3 scores for each skeleton point. Thus, it could be treated as a 3-channel color image, where all non-skeletal pixels have a value of 0 and skeleton points are described with 3 scores. This can then be fed through a CNN to extract features from, which was done.

Another method of extracting a 1 dimensional feature from the skeleton scores is just to extract summery information. This involved taking several averaging statistics of the scores, while also considering structural information of the skeleton itself. The following features were extracted from the scores: min, max, average and standard deviation of each score (ribbon, taper, separation) across the skeleton. The following additional skeleton structural features were extracted: number of branches, average and max branch length, and bend (angle between line connecting furthest 2 skeleton points and major axis). Note that all of these features are independent of any baseline features mentioned, and thus could be used as supplemental information to aid in classification.

Another interesting method for extracting a feature from the skeleton data would be to define the skeleton as a graph, with node features being the 3 scores, and apply graph

convolutions or graph kernels to do the classification. Deep learning on graph structured data has shown promise in recent years and could prove useful in this task. However, the graph analysis is outside of the scope of this study currently.

The skeletonization code was adapted from [Rez19; Sid+02; DDS03; RS13], and the scoring code was written by me.

3.0 Results

Results can be seen in table 1. Classification for all tests was done using a linear SVM with RBF kernel. Recall that there is equal class representation, so a random algorithm would achieve an accuracy of roughly 0.5. Scores were averaged over 20 runs.

Table 1: BreCaHAD Classification Results

Method	Accuracy	Sensitivity	Specificity
Morphological features	0.846 +/- 0.001	0.831 +/- 0.001	0.861 +/- 0.001
CNN on full color crop	0.632 +/- 0.001	0.477 +/- 0.001	0.798 +/- 0.001
CNN on 3-channel skeleton image	0.654 +/- 0.002	0.623 +/- 0.001	0.687 +/- 0.001
MAT Skeleton features	0.850 +/- 0.004	0.844 +/- 0.004	0.852 +/- 0.003

Although the proposed method achieved the highest accuracy, we reject the null hypothesis that the distributions of accuracies for the skeleton features and the morphological features come from different distributions at significance level 5%. So, it is an improvement over the baseline morphological features, but not by much. Additionally, the MAT skeleton feature achieved the highest sensitivity of the methods testing. Sensitivity is especially important in a medical setting, where a false negative (missed disease diagnosis) is much more costly than a false positive (where a patient may receive unnecessary further testing or treatment). This also highlights the failure of the color CNN feature, where its passable accuracy score was achieved by outputting mostly negative labels, resulting in many false negatives. It also demonstrates that the 3-channel skeleton image CNN feature was actually significantly better than the CNN on the full color image, as is seen in the difference in sensitivity.

Scores could likely be improved for both the morphological and skeleton features by taking several steps. First, color features could be incorporated into the skeleton and morphological features. These could be extracted from a CNN, or just be basic statistical measures such as in [Wan+16]. Additionally, a better classifier could have been used, such as a simple

multi-layer neural network. However, with these results we have shown the efficacy of the MAT skeleton feature as a means of distinguishing nuclei shapes.

4.0 Conclusions

Overall, the purpose of the study was not to achieve the best classification metrics and beat out all other state of the art models, but rather to define a useful shape descriptor that accurately captures the change of shape of the nucleus. In that regards, it was a success. The MAT skeleton feature showed improvement over the baseline morphological features. Additionally, features extracted with a CNN on the weighted skeleton image performed better than similar features extracted from the full color image. The MAT skeleton scoring algorithm outlined in this paper is thus a useful shape descriptor in regards to tracking nuclei shape heterogeneity in breast tumors.

Although here applied to distinguishing tumor and non-tumor nuclei in malignant breast tumors, the MAT scores may be robust enough to tackle harder challenges, such as the classification of high-risk tumors, or be applied to other modalities and cancers, such as prostate cancer (which also uses stained slide images for diagnosis). Further work will focus on applying this technique to more challenging tasks and datasets.

We can also apply the nucleus scoring algorithm, where each nuclei receives a set of features, to classify a whole region by considering all the nuclei in the region. One way to do this could be to just take a majority vote. If the majority of nuclei in the region are cancerous, then predict the image as cancerous. Another method could be normalized mean statistics being taken over the graph to return a 1 dimensional vector. These are naive approaches, but they highlights how the MAT scores may extend beyond nuclei classification.

A more interesting way to combine the nuclei features across the image could be to define a graph over the image, where the nodes in the graph are the nuclei, and each node inherits the features describing the nuclei. This would maintain the structure and distribution of nuclei in the original image, while allowing for various graph learning methods to be applied, which can consider both the individual nuclei features, as well as the overall structure of the tumor environment. This is a topic for further study.

I would also like to highlight the importance of model interpretability in medical studies such as this one. Many deep learning models exist which obtain impressive classification

results in this domain, such as in [Alo+18] where they achieve almost 98% accuracy on the BreakHis dataset (benign vs. malignant tumors). However, one thing missing from deep learning models is the ability to easily interpret results. Methods exist for probing deep nets to try to obtain some idea of what its doing, but a deep net will never be as interpretable by a human as a simpler model, such as an SVM. This is especially important when considering who will be using the software, primarily pathologists, who are not experts in data science. Using hand-picked features which can be mapped to real world descriptions (such as a nucleus being rounded) and simpler decision models leads to better potential interpretability and more trust in the model. Further work could include adding interpretability into the proposed feature extraction framework. If our aim is to improve the field of pathology by providing pathologists with tools that will aid them in their diagnostic process, a crucial step is gaining the trust of pathologists, and that can be accomplished through model interpretability.

Bibliography

- [PR92] David L. Page and Lowell W. Rogers. “Combined histologic and cytologic criteria for the diagnosis of mammary atypical ductal hyperplasia”. In: *Human Pathology* 23.10 (1992), pp. 1095–1097. ISSN: 0046-8177. DOI: [https://doi.org/10.1016/0046-8177\(92\)90026-Y](https://doi.org/10.1016/0046-8177(92)90026-Y). URL: <http://www.sciencedirect.com/science/article/pii/004681779290026Y>.
- [Sid+02] Kaleem Siddiqi et al. “Hamilton-Jacobi Skeletons”. In: *Int. J. Comput. Vision* 48.3 (July 2002), pp. 215–231. ISSN: 0920-5691. DOI: 10.1023/A:1016376116653. URL: <https://doi.org/10.1023/A:1016376116653>.
- [DDS03] Pavel Dimitrov, James N Damon, and Kaleem Siddiqi. “Flux invariants for shape”. In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2003, pp. I–I.
- [KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- [Kot+13] Sonal Kothari et al. “Pathology imaging informatics for quantitative analysis of whole-slide images”. In: *Journal of the American Medical Informatics Association : JAMIA* 20 (Aug. 2013). DOI: 10.1136/amiajnl-2012-001540.
- [RS13] Morteza Rezanejad and Kaleem Siddiqi. “Flux graphs for 2D shape analysis”. In: *Shape Perception in Human and Computer Vision*. Springer, 2013, pp. 41–54.
- [Elm+15] Joann G. Elmore et al. “Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens”. In: *JAMA* 313.11 (Mar. 2015), pp. 1122–1132. ISSN: 0098-7484. DOI: 10.1001/jama.2015.1405. eprint: <https://>

jamanetwork.com/journals/jama/articlepdf/2203798/joi150021.pdf.
URL: <https://doi.org/10.1001/jama.2015.1405>.

- [Har+15] Lynn C. Hartmann et al. “Atypical Hyperplasia of the Breast — Risk Assessment and Management Options”. In: *New England Journal of Medicine* 372.1 (2015). PMID: 25551530, pp. 78–89. DOI: 10.1056/NEJMs1407164. eprint: <https://doi.org/10.1056/NEJMs1407164>. URL: <https://doi.org/10.1056/NEJMs1407164>.
- [RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *CoRR* abs/1505.04597 (2015). arXiv: 1505.04597. URL: <http://arxiv.org/abs/1505.04597>.
- [Sha+16] Kabeer K. Shah et al. “Validation of diagnostic accuracy with whole-slide imaging compared with glass slide review in dermatopathology”. In: *Journal of the American Academy of Dermatology* 75.6 (Dec. 2016), pp. 1229–1237. ISSN: 0190-9622. DOI: 10.1016/j.jaad.2016.08.024. URL: <https://doi.org/10.1016/j.jaad.2016.08.024>.
- [Wan+16] Pin Wang et al. “Automatic cell nuclei segmentation and classification of breast cancer histopathology images”. In: *Signal Processing* 122 (2016), pp. 1–13. ISSN: 0165-1684. DOI: <https://doi.org/10.1016/j.sigpro.2015.11.011>. URL: <http://www.sciencedirect.com/science/article/pii/S0165168415003916>.
- [Yam+16] Yoshiko Yamashita et al. “Does flat epithelial atypia have rounder nuclei than columnar cell change/hyperplasia?: A morphometric approach to columnar cell lesions of the breast”. In: *Virchows Archiv* 468.6 (2016), pp. 663–673. ISSN: 1432-2307. DOI: 10.1007/s00428-016-1923-z. URL: <https://doi.org/10.1007/s00428-016-1923-z>.
- [Yam+17] Yoichiro Yamamoto et al. “Quantitative diagnosis of breast tumors by morphometric classification of microenvironmental myoepithelial cells using a machine learning approach”. In: *Scientific reports* 7 (Apr. 2017). srep46732[PII], pp. 46732–46732. ISSN: 2045-2322. DOI: 10.1038/srep46732. URL: <https://pubmed.ncbi.nlm.nih.gov/28440283>.

- [Alo+18] Md. Zahangir Alom et al. “Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network”. In: *CoRR* abs/1811.04241 (2018). arXiv: 1811.04241. URL: <http://arxiv.org/abs/1811.04241>.
- [Kad+18] Tanjina Kader et al. “Atypical ductal hyperplasia: update on diagnosis, management, and molecular landscape”. In: *Breast Cancer Research* 20.1 (2018), p. 39. ISSN: 1465-542X. DOI: 10.1186/s13058-018-0967-1. URL: <https://doi.org/10.1186/s13058-018-0967-1>.
- [Aks+19] Alper Aksac et al. “BreCaHAD: a dataset for breast cancer histopathological annotation and diagnosis”. In: *BMC Research Notes* 12.1 (2019), p. 82. ISSN: 1756-0500. DOI: 10.1186/s13104-019-4121-7. URL: <https://doi.org/10.1186/s13104-019-4121-7>.
- [Che+19] K. Chen et al. “Cell Nuclei Detection and Segmentation for Computational Pathology Using Deep Learning”. In: (2019), pp. 1–6.
- [Hal19] Steven Halls. *Micropapillary breast cancer*. May 2019. URL: <https://breast-cancer.ca/micropap/>.
- [Rez19] Rezanejad. *mrezanejad/AOFSkeletons*. June 2019. URL: <https://github.com/mrezanejad/AOFSkeletons>.
- [Rez+19] Morteza Rezanejad et al. “Scene Categorization From Contours: Medial Axis Based Saliency Measures”. In: June 2019, pp. 4111–4119. DOI: 10.1109/CVPR.2019.00424.
- [Bre20] BreastCancerOrg. *U.S. Breast Cancer Statistics*. Jan. 2020. URL: https://www.breastcancer.org/symptoms/understand_bc/statistics?gclid=CjwKCAjwvtXOBRAFEiwAGWJyZJpuc9fByPVUb04838EcKnRJ4uXckwegdszn0arkrF8EJ0z_2fQYGhoCbWQQA_vD_BwE.
- [CLW20] Laura C Collins, Christine Laronga, and Julia S Wong. *Ductal carcinoma in situ: Treatment and prognosis*. Jan. 2020. URL: <https://www.uptodate.com/contents/ductal-carcinoma-in-situ-treatment-and-prognosis?search=ductal-carcinoma-in-situ-treatment-and-prognosi&source=>

search_result&selectedTitle=1~54&usage_type=default&display_rank=1.