High-Risk Breast Lesion Classification With Graphs

Brian Falkenstein

University of Pittsburgh, Pittsburgh PA 15213, USA

Abstract. High risk benign breast lesions are often misclassified by pathologists during biopsy sample evaluation, yet early detection and treatment of these high risk tumors is crucial for improved patient outcome. In this study, I examine the efficacy of spatial network graphs defined on tumor region images for the task of classification. The results of the experiment mirror similar studies on high risk tumor classification in that they highlight the difficulty of the task.

Keywords: Breast cancer · Graphs · Machine learning.

1 Introduction

Breast cancer is not a binary diagnosis, but rather a breast tumor may belong to one or more of a myriad of classifications. Pre-cancerous tumors are masses of cancerous cells that have not migrated beyond their original location, that is, they have not spread to other parts of the body, at which point the tumor would be considered malignant. Early detection and treatment, either by extraction or some other method, can greatly reduce the chances of further invasive cancer [1].

Of concern is the alarming rate of discordance, or disagreement, among pathologists in classifying high risk breast tumors as reported in one study, which showed an average disagreement rate of 52% for high risk lesions, compared to 4% and 13% for invasive and benign cases respectively [2]. A majority (73%) of these discordant high risk classifications were under interpreted. Under interpreted high risk lesions, in a clinical scenario, would mean someone with a heightened risk for developing invasive breast cancer would not be receiving necessary treatment. This begs the question of how we may improve upon the existing framework for detection of high risk breast lesions, and more specifically how we can apply machine learning techniques to aid pathologists in the classification.

Many works have studied various machine learning and image processing methods for assistance in breast cancer diagnosis, ranging all the way back to 1994 when Wolberg et al. [3] classified breast tumor nuclei as benign or malignant using decision trees on human computed features with good success. Features included basic morphological properties, such as area and perimeter, as well as various textural features. This showed that tumor nuclei could be identified by shape and texture. However, the task of classifying whole tumor regions requires

not just reasoning on individual nuclei, but also considering other important biological structures present in the image, such as the lumen (which becomes crowded in high risk tumors) and the duct itself (which may be swollen with extra cells), as well as how these structures are spatially located relative to one another. As a tumor develops, cancerous cells reproduce at a rapid rate, causing a crowding of cells in chaotic patterns when they should be, in the case of epithelial cells, neatly arranged in rows around the outside of the duct.

Many deep learning approaches have been applied to classifying whole regions, with some of them particularly focused on capturing these spatial relationships. Particularly of interest, Chennubhotla et al. [4] define a network over each tumor region, where nodes are defined as nuclei and other biological structures in the image. Nodes are labeled as either one of 3 phenotypes of nuclei (achieved via KNN on morphological and textural features computed from the data set), stroma (space between ducts), or lumen (area in the center of a duct). The whole tumor is then represented by a q dimensional probability vector of the tumor displaying each of the q architectural phenotypes (again derived via KNN on the network graphs). Using this method, they achieved a maximum F score of 0.83 in distinguishing high and low risk lesions, although the recall of the proposed method was lower (0.69). The method proposed in this paper seeks to expand on the spatial network idea found in [4] by incorporating feature representations for nuclei in the graph, as well as using more advanced graph classification methods (Graph Convolutional Networks and Graph Kernels).

Table 1. All of the features extracted for each nuclei.

Feature name	**Description**
Area	The area of the nucleus (number of pixels).
Eccentricity	Eccentricity of the ellipse that has the same second-moment as the region, the ratio of the distance between the foci of the ellipse and its major axis length.
Extent	Ratio of pixels in the region to pixels in the total bounding box.
Major axis length	Length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.
Minor axis	Length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.
Convex area	Number of pixels in the convex hull.
Circularity	Computed as $(4 * Area * pi)/(Perimeter^2)$, how close the shape is to a perfect circle.
Equivalent diameter	Diameter of a circle with the same area as the region.
Filled area	Number of pixels in the filled region.
Perimeter	The perimeter of the boundary of the nucleus.
Solidity	Proportion of the pixels in the convex hull that are also in the region.

2 Methods

The proposed method can be broken up into three distinct phases: nuclei segmentation and feature extraction, graph construction, and graph classification. The segmentation algorithm was provided by Dr. Om Choudhary and is a simple statistical algorithm which segments the nuclei based on PCA of the images color space (since nuclei are dyed purple in the biopsy images, see Figure 1), followed by several post processing and smoothing steps to remove false positive detections and make boundaries smoother. The reason for a simpler statistical approach was the nature of the dataset (discussed further in Experiments), where the images where too large for deep-learning segmentation methods which would have taken too long. However, its important to note that we will be extracting shape based features to describe the nuclei, so a better segmentation algorithm will likely lead to better results in general.

Once we have an estimated boundary for each nuclei, we will isolate each one and extract morphological features to represent the shape of each nuclei. All of the extracted features can be seen in table 1. The goal was to have an accurate descriptor of the shape of the nuclei in the image, independent of their spatial location in the image, as nuclei shape is one indicator of the stage of the tumor.

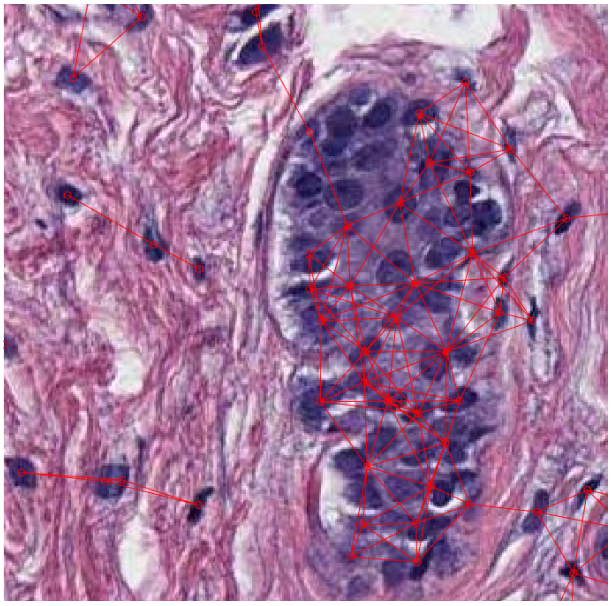

Fig. 1. A spatial network graph defined on a small region of a tumor using distance threshold 50.

All features were extracted using Matlab's *regionprops* function. Next, we define the spatial network graph for the region image. This is done by first

defining a node for each nuclei detected in the image, then drawing edges between nodes based on a distance threshold. For all of my experiments, I used a threshold of 50 pixels, as this seemed to reasonably connect nuclei which were closely bound, such as inside of a duct. Figure 1 shows a zoomed in region of one of the graphs defined. Of notice is how nuclei which are tightly contained within a duct will have a high degree, and nuclei which are outside of the ducts, in the stroma, may have little or no connectivity.

Once we have defined a graph for the image, we discard the original image information, and perform classification using only the graph. Thus, we will be ignoring all color and textural information, and be using exclusively individual nuclei shape, as well as the nuclei spatial relationships to one another. One advantage of this is that we need not care the dimensionality of the input images, as graph methods must function on variable sized graphs, whereas typical deep net approaches for image classification require special processing if images are of variable sizes. Two methods for classifying graphs were explored in this study: Graph Convolutional Networks, and SVM with Graph Attribution Kernel.

2.1 Graph Convolutional Network

Graph Convolutional Networks, or GCNs, seek to extend the convolution operation for use in graph structured data. There are many reasons why you may want to represent data in a graphical manner, and being able to process unstructured graph data in a similar way to how we can process images (which can be thought of as graphs, where each node is a pixel) is important. Two sub problems exist within graph classification, being semi-supervised node classification (where some nodes are unclassified, and we use information from other classified nodes to perform the classification) and whole graph classification. For our purposes, we are interested in defining a graph for a tumor, and classifying the whole thing. For that reason, we need some method of projecting our unstructured graph data into a space where we can perform classification on variably sized and structured graphs.

Following the example in [5], I implemented a simple 2 layer GCN with ReLU activation. During this stage of the network, feature data is propagated between neighboring nuclei, with the intuition being that data will propagate differently for different structured graphs. This is the convolutional operation, and the weights are learnable. Then, global pooling is performed to obtain a 1 dimensional vector representation for the whole graph. Several pooling methods were tested (Min, max, average, sum, top K), but only max pooling proved viable. Averaging tended to show zero separation between the different classes, and sum fails due to it being partially dependent on the number of nodes in the graph, which we'd like to avoid.

Once we have a vector representation for the graph, we perform classification using a simple 2 layer neural network with ReLU activation, and finally a softmax (for 4 way classification) or sigmoid (for binary high vs. low risk) function to obtain class probabilities.

2.2 Graph Attribution Kernel

Due to initial poor performance of the GCN method, I also implemented a graph kernel method, particularly the Attribution Propagation Kernel [6]. A graph kernel, $k(G_1, G_2)$ is a function which compares two graphs based on some similarity measure and produces a single value representing how similar the two graphs are. The Attribution Propagation kernel is a special type of kernel that allows for comparison of graphs which have node level features, where similarity is determined both by the overall structure of the graphs, as well as the individual node features. It does this via iterative diffusion (where information is propagated to neighboring nodes) and definition of histograms of the nuclei features. The histograms thus capture both the original nuclei features, as well as the structure in that the diffusion of information will be different for graphs which are structured different (IE, a more connected graph will be 'smoother' after diffusion than a sparser graph'). Note that there are no learnable parameters in this framework, the diffusion and histogram creation process is nonstochastic.

Once we have our kernel defined, we train an SVM to separate the classes based on their kernel values.

3 Experiments and Results

3.1 Data

The data set used was initially described in [4]. It contains roughly 1700 images of breast tumors of varying sizes (average 1800*1800 pixels) and at varying zoom levels. The images are classified independently by 3 pathologists as either benign, Columnar Cell Change (CCC), Flat Epithelial Atypia (FEA), or Atypical Ductal Hyperplasia (ADH). These labels can then further be broken down into low risk (benign and CC) and high risk (FEA and ADH) for a binary classification.

Several issues arise from the dataset. One being the disparity in class representation, with 49% of all the labels being benign, 13% of them CCC, 7% of them FEA, and 30% ADH. Another big concern is those cases where the 3 pathologists disagree on the label of the image. This occurs very frequently in the dataset, particularly for cases where one or more of the labels are high risk. Only 23% of samples in the data set had fully concordant labels, and of these, over 90% of them are benign. 25% of the data points have all 3 labels as different values, and of these a majority have at least 1 high risk label. For the experiments conducted, the label which is considered the highest risk was taken as the ground truth. The intuition behind this is that perhaps the other pathologists missed a feature which would indiciate a higher risk. The pathologist/pathologists who labeled it as high risk did so for a reason.

3.2 Experiments

Experiments were conducted with both the 4 way classification task, as well as the binary one. The test set consisted of 323 images, 228 low risk, and 95 high risk.

Table 2. Performance of the proposed methods in the binary (high vs. low risk) classification task.

Method	**Weighted F1 Score**	**Precision**	**Recall**
GCN	0.64	0.3636	0.028
Graph Kernel	0.63	0.4242	0.147

Table 3. Confusion matrix for binary classification for the GCN.

	True	
Predicted	391	7
	135	4

4 Discussion and Conclusion

Overall, the proposed method leaves a lot to be desired. Notice the confusion matrix for the GCN, where the few true positive labels it provided were likely by chance. Confusion matrices for the other tests looked similar, and it seems that the method is not working well at all. Previous work, particularly [7] and [4], showed better success in classifying the high risk tumors. Particularly, [7] did so merely by classifying individual nuclei as high risk or not, and then performing a majority vote of nuclei to classify a whole tumor. However, their analysis was conducted on a very small number of images, just 22.

One change that could improve the performance of the model could be a better nuclei segmentation algorithm. Many deep learning approaches exist, and show much better results than the method used here. I was unable to obtain a reasonable set of deep net segmented images within the deadline, but this could be studied further. Another improvement could be simple data augmentation (cropping, rotating, etc.) to give the models, particularly the GCN, more training data. I also had issues with both of the graph classification methods I implemented here. For the GCN, I did not like the global pooling operation, as I feel that this did a bad job distinguishing between different classes due to the lack of separability after pooling over potentially hundreds of nodes. I'd imagine there is a lot of loss of information during the pooling layer. Perhaps, a hierarchical pooling system could be deployed, where subgraphs are pooled and concatenated together with a pooling over the whole graph. Or, perhaps some sort of learnable pooling layer.

My main issue with the graph kernel method was its lack of learnability. As far as I am aware, learnable graph kernels do not exist. Due to the fact that high

Table 4. Performance of the proposed methods in the 4 way classification task.

Method	**Weighted F1 Score**	**Precision**	**Recall**
GCN	0.40	0.39	0.26
Graph Kernel	0.40	0.29	0.28

risk tumors are difficult to differentiate, as evidenced by the high pathologist discordance rates, I'd think a generic kernel may not be robust enough.

In conclusion, more work needs to be done to prove the effectiveness of network graphs for classification of high risk breast tumors. Intuitively, the study of nuclei shape and how they are spatially arranged fits the biological definition of different high risk tumor classifications. However, we may need to do more work to prove that intuition with results.

References

1. 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/NEJMsr1407164.
2. Joann G. Elmore et al. "Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens". In: JAMA313.11(Mar. 2015), pp. 1122–1132.issn: 0098-7484.doi:10.1001/jama.2015.1405.
3. William H. Wolberg, W.Nick Street, and O.L. Mangasarian. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates.Cancer Letters, 77(2):163 – 171, 1994. ISSN 0304-3835.
4. Tosun Akif Burak, Nguyen Luong, Ong Nathan, Navolotskaia Olga, Carter Gloria, Fine Jeffrey L., Taylor D. Lansing, and Chennubhotla, S. Chakra. Histological Detection of High-Risk Benign Breast Lesions from Whole Slide Images. Medical Image Computing and Computer-Assisted Intervention MICCAI 2017 978-3-319-66185-8
5. Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. 2016 1609.02907
6. Marion Neumann and Roman Garnett and Christian Bauckhage and Kristian Kersting. Propagation Kernels, 2014. 1410.3314
7. Yamamoto, Yoichiro et al. "Quantitative diagnosis of breast tumors by morphometric classification of microenvironmental myoepithelial cells using a machine learning approach." Scientific reports vol. 7 46732. 25 Apr. 2017, doi:10.1038/srep46732