Infiltrating duct carcinoma and Lobular carcinoma classification in whole slide image with Convolutional Neural Networks

Weituo Hao 109241801
Computer Science Department
Of Stony Brook University
whao@cs.stonybrook.edu

ABSTRACT
This report focuses on the issues raised by the analysis and evaluation of invasive ductal carcinoma tissue regions in whole slide images (WSI) of breast cancer (BCa). With convolution neural networks (CNNs) which is a method based on learn-from-data, we have experimented on the classification problem of infiltrating duct carcinoma (IDC) and lobular carcinoma (LC). We compared performance of different architecture and verified that deeper network will work better especially on the dataset which is harder to classify than others. Our method can achieve average accuracy of 0.6532 on patch-level and average accuracy of 0.72 on image level.

Categories and Subject Descriptors
J.3.2 [**Computer Applications**]: Life and Medical Sciences-Medical Information Systems H.3.3 [**information Search and Retrieval**]: Clustering

General Terms
Medical imaging, deep learning

Keywords
Computer vision, medical imaging, convolutional neural networks.

1. INTRODUCTION
According to the American Cancer Society, more than 180,000 women in the United States find out they have invasive breast cancer each year. Most of them are diagnosed with invasive ductal carcinoma. Invasive means that the cancer has "invaded" or spread to the surrounding breast tissues. Ductal means that the cancer began in the milk ducts, which are the "pipes" that carry milk from the milk-producing lobules to the nipple. Carcinoma refers to any cancer that begins in the skin or other tissues that cover internal organs — such as breast tissue. All together, "invasive ductal carcinoma" refers to cancer that has broken through the wall of the milk duct and begun to invade the tissues of the breast. Over time, invasive ductal carcinoma can spread to the lymph nodes and possibly to other areas of the body.

One way to reduce the bad effect of breast cancer is early diagnosis. Precise delineation of IDC in WSI is crucial to the subsequent estimation of grading tumor aggressiveness and predicting patient outcome. But it is very time consuming for pathologists to scan large parts of benign parts to finally identify the areas of malignancy.

In addition, most of previous approaches involve segmentation of histologic primitives (e.g. nuclei) and then extracting features represent the appearance or arrangement of these primitives to distinguish malignant from benign regions. These approaches typically involve a series of pre-processing steps including detection, segmentation with the result that the final classification result is dependent on the accuracy of the preceding steps.

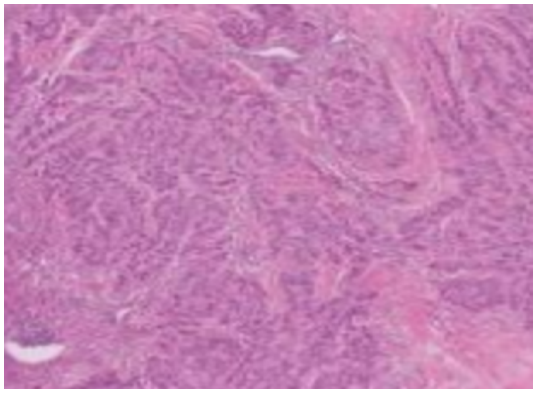
On the other hand, deep learning approach has been proved to outpace traditional approaches of most challenging problems like computer vision and image classification. So in this report we adopt extended convolution neural networks to solve the challenging problem of cancer image classification. The main advantage of deep learning models is that they learn the most appropriate representation in a hierarchical manner as part of the training process. However, most deep learning strategies involving histopathology involve applied a pixel level classification over relative small images. Our approach on the other hand employs a deep learning framework based classification on square tissue regions from WSI obtained by regular sampling, thereby enabling the application of the classifier over the entire canvas of the WSI. Specifically, we first let the CNN be trained over a large amount of image patches (tissue regions)

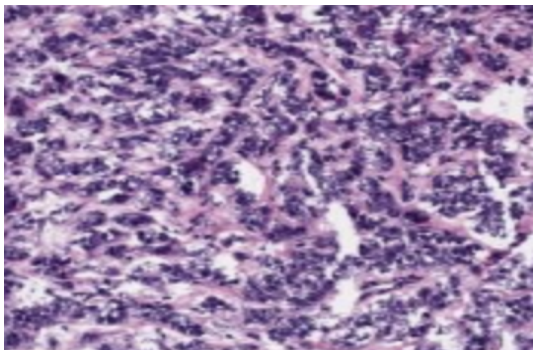
from WSI to learn a hierarchical part-based representation. Based on the scores of each image patch we combine them use the average score to vote for the WSI's label thus convert the result back to image-level.

The rest of this paper is organized as follows: Section 2 describes the specific problem; Section 3 presents the methodology. Section 4 shows the experimental results and discussions. Section 5 concludes with the main findings and future directions.

2. PROBLEM DEFINITION

In this section, we assume that we are given a set of whole slide image of different patient ids. These whole slide images are labeled with different cancers. For instance, the WSI labeled 0 means the image is from a patient with Infiltrating duct carcinoma and 1 means Lobular carcinoma. So the problem is given a new WSI without any prior knowledge, we need to figure out how to predict the new image's label. The figure 1 shows some screenshots of the WSI of infiltrating duct carcinoma.





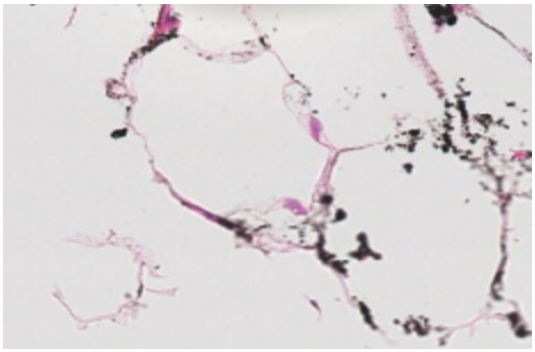


Figure 1 screenshots of infiltrating duct carcinoma WSI

From top to down, the patches are carcinoma part, normal parts and fat tissue. Note that the cancer turns to turn down the diversity of cells and change all of them into cancer cells. In addition, fat tissue in WSI is a hard problem to deal with if no image segmentation done according to pathologist's instruction.

3. METHODOLOGY

The basic idea of our method can be illustrated as following picture. We first sample the whole slide image into different patches. Then we use CNNs to learn these patches. For a new whole slide image we can also sample it into different patches, and use the model in the training phase to predict these patches. Finally we get the average score of each predicted labels.

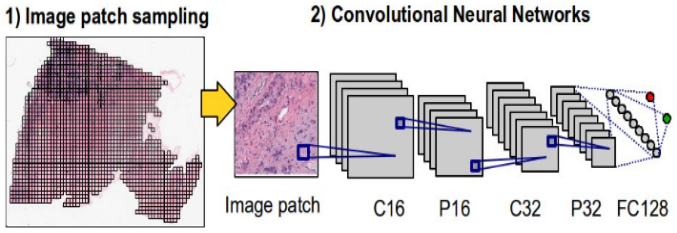


Figure 2 Overall framework

3.1 Image patch sampling

Note that each original digitalized WSI is so large that it is not unusual that computers fail to operate on the WSI. So we first sample the WSI into small patches. Each WSI is split into non-overlapping image patches of 400*400 pixels. Patches with mostly fatty tissue or slide background are discarded. Since we don't use any pathological information to segment the WSI, the label for the whole slide image is used to

label each image patch. In addition, different resolution will result in different results. Here we mainly use the 5X resolution datasets. And some sample image patch under this patch can be seen as follows. We also carry out experiments on the 20X datasets. But it turns out this dataset contains too much fatty tissue patches which effects the final learning accuracy. We will discuss it in the Section 4

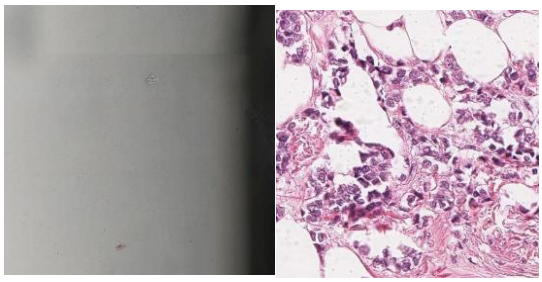


Figure 3 Image patches. The left one is background remaining in the dataset and the right one is the tissue patch. The left one is from TCGA-A2-A0SY-01Z-00-DX1 and the right one is from the same image id

3.2 Convolutional Neural Networks

CNN involve applications of local features detectors or filters over the whole image to measure the correspondence between individual image patches and signature patterns within the training set. Then, an aggregation or pooling function is applied to reduce the dimensionality of the feature space. The image patch in Section 3.1 are used as inputs to a 4-layer CNN and 7-layer CNN separately. (Note that we only count the convolution layer here and ignore other layers such as pooling layer or RELU layer. The real layer of 4-layer CNN is 14 layers and 19 layers for 7-layer CNN).

Specifically, for convolution layer we use the Gaussian kernel for each image patch which is a good strategy to present larger image. To model the real biological process we add RELU layer to each convolution layer to rectify non-linearities in the convolution output.

For pooling layer, we use the max pooling strategy. This stage allows us to reduce the original large dimension of image representation through a subsampling strategy which support local space invariances. Then, this layer applies a max pooling function over feature maps by a spatial windows without overlapping. The stride is 2 for each pooling layer.

After several convolution and RELU layers we add two inner product layers with dropout layer as the end. The dropout probability is 0.5. Finally, the output will go through softmax classification layer which satisfies the function:$\sigma(\boldsymbol{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$.

The exponential function is applied to each output value obtained from positive class neuron of classified patches by CNN to get values between 0 and 1, so that they could be interpreted as the scores for the predicting label. Also we can consider the score is the probability for each predicting label. For the WSI's label, we just get the average score of each label. The following picture shows the 4-layer architecture.

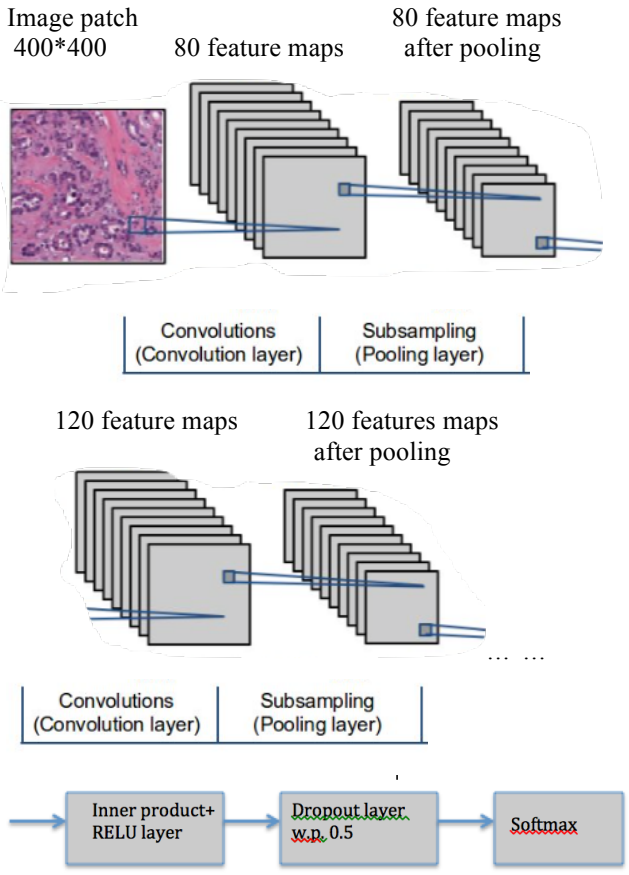


Figure 4 The CNN architecture of 4-layer

Since the 7-layer architecture is similar to the 4-layer except for some parameters and more convolution layer. We don't list the picture here. Note that the 4-layer only means the convolution layer not all the layers we have used.

4. EXPERIMENTAL RESULTS

All whole slide images come from the National Cancer Institute TCGA data portal. We first carry out the experiments on the 20X datasets. And the result is not as good as 5X datasets since there are more fatty tissue patches. So all the experiments are done under 5X datasets.

4.1 Infiltrating duct carcinoma and Lobular carcinoma

The 4-layer CNN and 7-layer CNN are used to classify the Infiltrating duct carcinoma patches labeled by 0 and Lobular carcinoma patches labeled by 1. We sample nearly 700,000 patches under these two kinds of cancer. For training set, we extract nearly 35,000 patches and for test set, we extract nearly 10,000 patches which include about 240 whole slide image ids. Among these 240 whole slide image ids, nearly 60% are labeled as 0 and the rest are labeled as 1.

To see the learning effect of these two architecture, we extract the accuracy of each 1,000 iteration during the learning process. If the architecture has learned something from the training dataset, the trend of the accuracy should be increasing. The accuracy of each architecture is displayed in the following pictures. The y-axis means the accuracy in patch level. The value of x-axis is proportional to the real iteration times.

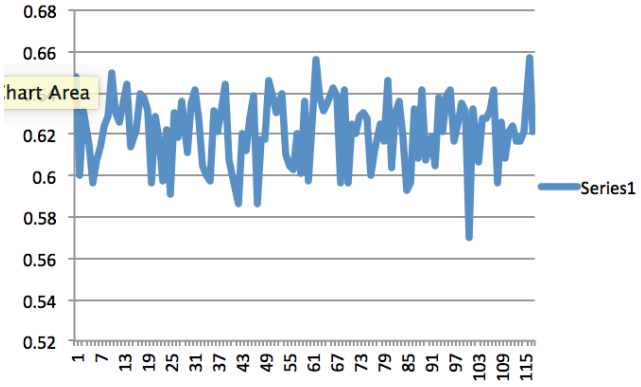


Figure 5 Accuracy trend of 4-layer CNN after 96,000 iterations

Figure 5 shows the accuracy trends of 4-CNN architecture after 96,000 iterations. The average accuracy is 0. 6296. From this figure we may not see any increasing trend during the training testing process.

We may assume that the phenomenon is caused by not enough training iteration times. So the result of 132,000 iteration is as follows:

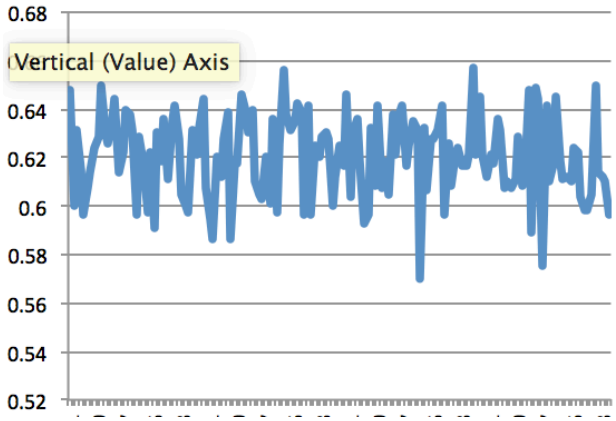


Figure 6 Accuracy trend of 4-layer CNN after 132,000 iterations

Similarly as the figure 5, the trend is not increasing currently. The average accuracy is about 0.6247.

To classify the infiltrating duct carcinoma and Lobular carcinoma is the most challenging task among the classification problems of breast cancer images. Note that the deep learning architecture should work better with more layers. So we continue to use 7-layer CNN.

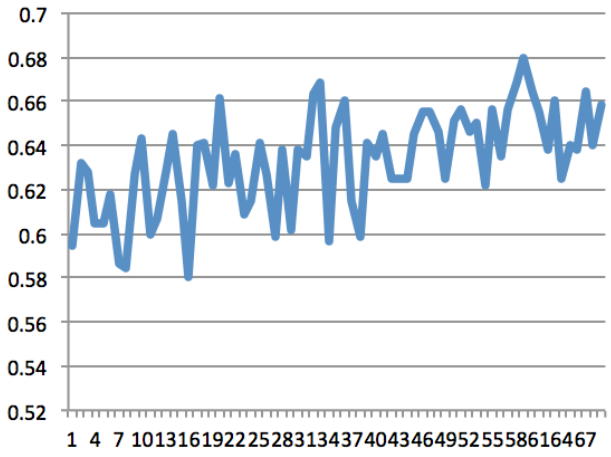


Figure 7 Accuracy trend of 7-layer CNN after 96,000 iterations

Figure 7 shows the accuracy trend of 7-layer CNN after 96,000 iterations. It displays a obvious increasing trend during the training iteration. And the

average accuracy is 0.6532, higher than all the results of 4-layer CNN. So 7-layer CNN works better than the 4-layer CNN here.

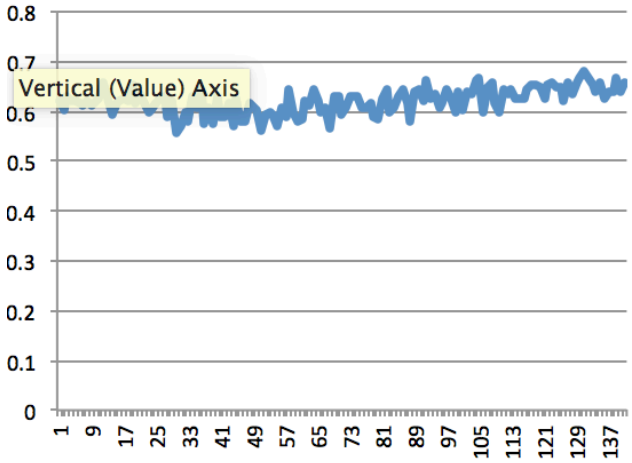


Figure 8 Accuracy trend of 7-layer CNN after 132,000 iterations

Figure 8 shows a long run of the patch accuracy of the 7-layer CNN architecture. It seems that the front part of the training plays a role of pre-training. After certain number of iterations, the architecture begins to classify the patches accurately.

4.2 Infiltrating duct carcinoma and other kinds of carcinoma

Then we continue to use the training model to predict other kinds of cancer images. They are Infiltrating duct and lobular carcinoma labeled as 2, Infiltrating duct mixed with other types of carcinoma labeled as 3, Mucinous adenocarcinoma labeled as 4, Metaplastic carcinoma labeled as 5, Inflammatory carcinoma labeled as 6, Infiltrating lobular mixed with other types of carcinoma labeled as 7, Medullary carcinoma labeled as 8, Duct micropapillary carcinoma labeled as 9, Medullary carcinoma labeled as 10. There are 537,292 image patches in training data and 148,702 image patches in testing data.

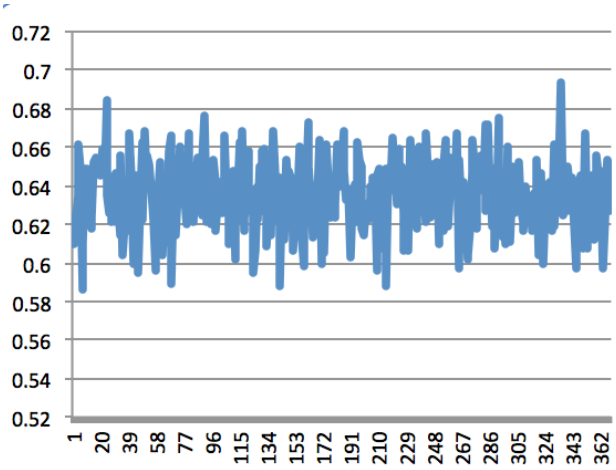


Figure 9 Accuracy trend of 4-layer CNN after 400,000 iterations

Figure 9 shows that 4-layer CNN still doesn't work even we increase the iteration time to 400,000. It can't learn anything from the training data set.

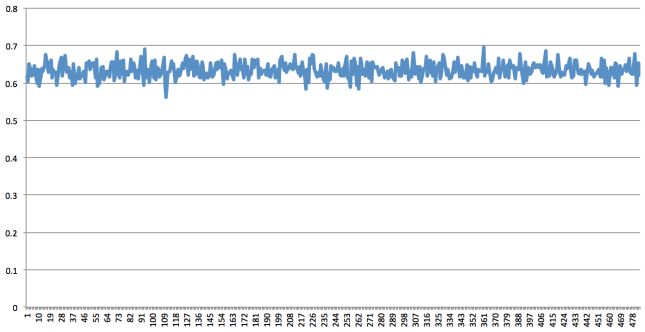


Figure 10 Accuracy trend of 7-layer CNN after 400,000 iterations

Figure 10 shows that 7-layer CNN doesn't work as good as previous experiments. The whole accuracy trend seems to increase only a little bit.

4.3 Whole slide image accuracy

According to the result of section 4.1 and 4.2, we carry out the whole slide image accuracy test on the classification of Infiltrating duct carcinoma and lobular carcinoma by 7-layer CNN model. The testing image patches are 10,000 in total that come from about 240 whole slide image ids. Almost each WSI comes from one patient. We use the trained model to predict each patch and consider the average the score as the final result of the WSI. There are nearly 62% patches of Infiltrating duct carcinoma and 38%

patches of Lobular carcinoma. Our model give an accuracy of nearly 72% on the whole image level.

In summary, we can conclude the following statements. 1. Infiltrating duct carcinoma and Lobular carcinoma are more difficult to classify than other kinds of carcinoma. 2. 4-layer CNN is not deep enough to get the satisfied result but 7-layer CNN does. 3. To classify Infiltrating duct carcinoma out of all other kinds of carcinoma we need new learning architecture. 4. 7-layer CNN can classify both the image patches and whole slide image well.

4.4 Weights presentaion

In this section we will give a brief presentation of the weights of each trained model to check the detail of each filters. Usually the weights or filters should differ from each other to guarantee that the whole learning framework indeed learn something from the training data. So we compare layer 1 of 4-layer CNN and 7-layer CNN architecture in the following two pictures.

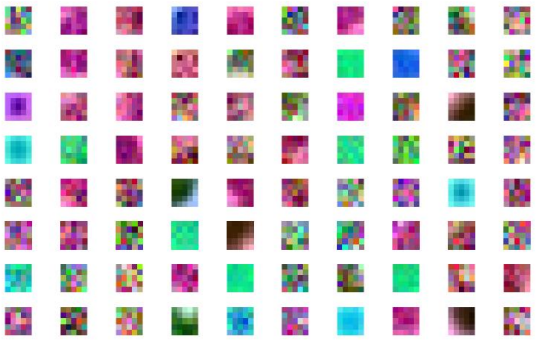


Figure 11 Weights of 7-layer CNN

Figure 11 shows the weights we have got from the trained model. These filters are from the layer 1 convolution and 80 in total. We can see enough variance of each filter.

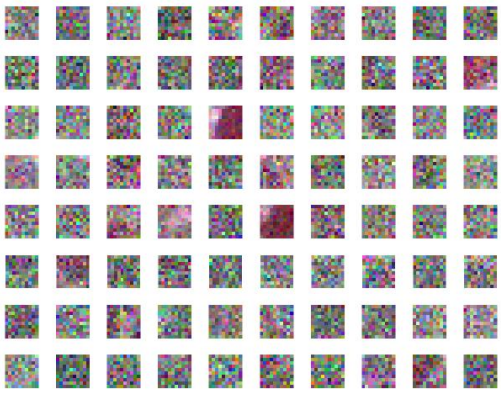


Figure 12 Weights of 4-layer CNN

Figure 12 displays the 4-layer CNN's first convolution filters. There is not enough information about this batch of filters since every filter seems similar to each other. This picture may explain 4-layer CNN doesn't work for the datasets.

5. CONCLUSIONS

We adopt two kinds of CNNs architecture to classify the Infiltrating duct carcinoma and Lobular carcinoma whole slide image that is most challenging task to classify other kinds of breast carcinoma images. 7-layer CNN architecture succeeds to finish the task and we get accuracy of 0.72 without any prior pathological information.

In future we will continue to investigate the reason why 4-layer CNN doesn't work well as 7-layer CNN, the relation between the depth of the CNN and its capability on different datasets. In addition, we may try to first unsupervised clustering [2] during the preprocessing phase and investigate the multi-supervised nets method mentioned in [3]

6. ACKNOWLEDGEMENT

Many thanks to Prof. Dimitris Samaras's lecture on advanced computer vision during this semester and Le's preprocessing on image patches.

7. REFERENCES

[1] Angel, C., Ajay, B., et al, "Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks," Medical Imaging 2014: Digital Pathology, Vol.9041

[2] Scott Doyle , Michael Feldman.,et al, “Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features” Biomedical Imaging [J]496 – 499, 14-17 May 2008, Paris, France

[3] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeplysupervised nets. arXiv:1409.5185, 2014.

[4] LeCun, Y., Bottou, L., Bengio, Y., and Ha↵ner, P., “Gradient-based learning applied to document recogni- tion,” Proceedings of the IEEE 86(11), 2278–2324 (1998).