Breast Cancer Detection using Deep Learning Neural Networks for Clinical Application

Axel Masquelin, Fangda Li

1. Introduction

In the medical field, pathologist are tasked with providing accurate diagnoses to patients in order to develop optimal treatment strategies. In order to be effective, pathologist require highly accurate, reproducible, and standardized diagnostic tools and are solely limited by the sensitivity, and specificity of such tools. The most notable of these tools is the microscope, utilized to qualitatively detect the presence of bacteria, observed the behavior of cells, and examine cell samples for the presence of cancers/pathogens [1]. However, this approach has its own limitations. The use of microscope to visual inspect slides through images is unstandardized, prone to high error rate based on the pathologist themselves, and require significant time and effort to properly evaluate all aspect of the sample [2]. Due to this, the medical field has slowly shifted toward more automated diagnostic methods that rely on computational analysis of sample images.

In 2016, the International Symposium on Biomedical Imaging (ISBI) hosted the Camelyon Grand Challenge 2016 (Camelyon 16), in which participants were tasked to computationally evaluated histology samples to automatically detect metastatic breast cancer in digital whole slide images [3]. This challenged aimed at utilizing histology samples for breast cancer carcinoma attracted participants from all over the world. The winner of the challenge was a Harvard-MIT combined team, which utilized deep neural networks to examine these whole-slide images.

In 2016 alone, 256 000 new cases of breast cancer were diagnosed worldwide [4]. Furthermore, breast cancer accounts for approximately 6.8% of all cancer related deaths [4]. For a positive diagnostic to be given to a patient, a pathologist will typically follow a set of exams, which will require an initial mammogram to examine the breast for small calcifications, shown in figure 1.

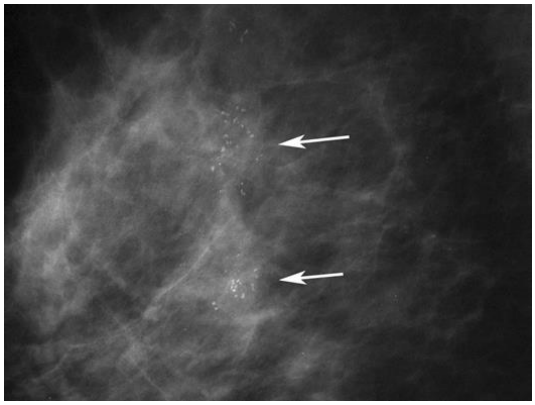

Figure 1: *Mammogram image showing calcifications nodes, potential risk of breast cancer. [5].*

If calcifications are present, then the pathologist will order a biopsy of the breast tissue in order to determine the stage of the cancer, and type (carcinoma vs sarcoma) [6]. From this biopsy, histology samples are generated and examined by trained pathologist. On average, a trained pathologist will only be able to examine 16-20 whole-slide images per day, making this method extremely inefficient and time consuming [7].

Although computer-assisted image analysis have been develop to detect the presence of small metastatic tumors, they are currently not being employed clinically and are utilized purely for research purposed. Hence, there is a need for a system to computationally evaluate and detect the presence of breast cancer carcinoma within histology samples of the patient in order to assist

pathologist in reaching a proper diagnosis. To achieve this, deep neural networks were utilized to detect the presence of breast cancer within whole-slide images.

2. Method

The dataset utilized for this project was acquired from the Camelyon 16 competition. It consisted of 170 whole-slide images, of which 100 were found to be normal histology samples and 70 contained some form of metastases [3]. Each image within the database was approximately 1-2 GB (compressed), and approximately 130k x 80 k pixels for each raw image. This caused issue, as a single whole-slide image could not fit on the RAM of the computer.

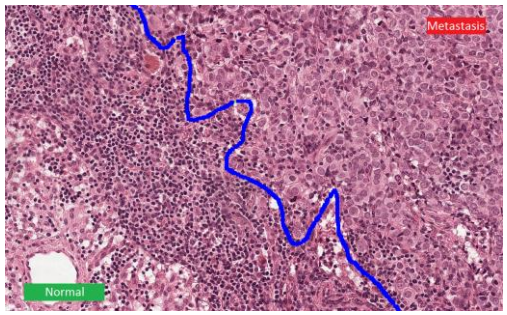

Figure 2: *Histology sample, showing normal cell region (green), and metastatic cell region (red) as seen in whole-slide images [3].*

To solve this issue, the original images were then sampled using Geospatial Data Abstraction Library (GDAL). This allowed for the generations of a new dataset of images based on whole-slide images. Each patches from the whole-slide image were generated to be 224x224 pixels. In order to minimize the amount of blank images generated, the whole-slide images were preprocessed using Otsu's algorithm to generate masks and identify the cell regions vs. white-space region, shown in figure 3.

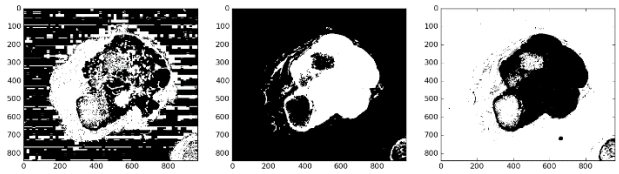

Figure 3: *Segmentation of background and tumor using Otsu algorithm, showing from left to right the background, S mask, and V mask.*

This was done in order to reduce the computation time, and to ensure that the network would be looking at the tissue samples and not white space. Furthermore, two more masks were generated too describe the cancer region and none cancer regions of the cells, as see in figure 4.

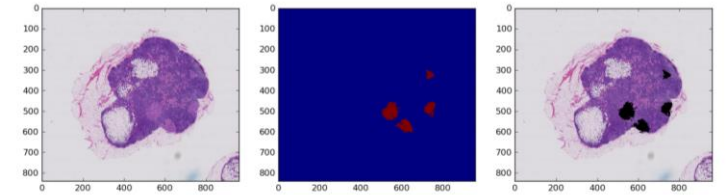

Figure 4: *(Left) raw patched image containing cells, (middle) heat map of tumor area location, (right) tumor mask overlaid on histology sample.*

This process of patch-sampling each imaged allowed for the generation of over 100,000 images to train the neural networks, ResNet18 and ResNet101.

3. Results

To evaluate both ResNet18 and ResNet101, two datasets of different sizes were generated; a small dataset consisting of 18,000 training images (10,000 normal and 8 thousand cancerous) and a 5,000 images testing set, and a larger dataset consisting of 35,000 images (18,000 normal and 17,000 cancerous) and 12,000 images for testing.

Both ResNet18 and ResNet101 were trained, evaluated, and tested on these two datasets.

	Patch Accuracy Small Dataset	*Patch Accuracy Full Dataset*
ResNet18	*92.42%*	*82.41%*
ResNet101	*98.22%*	*96.73%*

Table 1: *Patch accuracy results of the small dataset (trained on 18 thousand images) and full dataset (trained on 35 thousand images.*

Trained pathologist on have an overall diagnostic accuracy ranging from 75-95% [6]. However, examining the results shown in table 1, one can see that in the case of a small dataset, both ResNet18 and ResNet101 outperformed the average trained pathologist. However, when ResNet18 was trained and tested on a larger dataset, its accuracy dropped significantly to 82.41%, while ResNet101's accuracy dropped to 96.73%.

	Patch classification accuracy
GoogLeNet [20]	98.4%
AlexNet [12]	92.1%
VGG16 [19]	97.9%
FaceNet [21]	96.8%

Table 2: *Results from the winning team of the Camelyon 16 [8].*

Comparing the patch classification accuracy of our neural networks to that of the winner of the Camelyon 16 seen in table 2, one can see that on the smaller dataset both ResNet18 and ResNet101 were able to compare to their results. However, on the larger dataset only ResNet101 was able to maintain the same degree of accuracy.

To evaluate both ResNet's diagnostic potential, a receiver operating characteristic (ROC) analysis was conducted. Through this test, it is possible to examine the dynamics the networks' true positive rate (sensitivity) as the false positive rate increases.

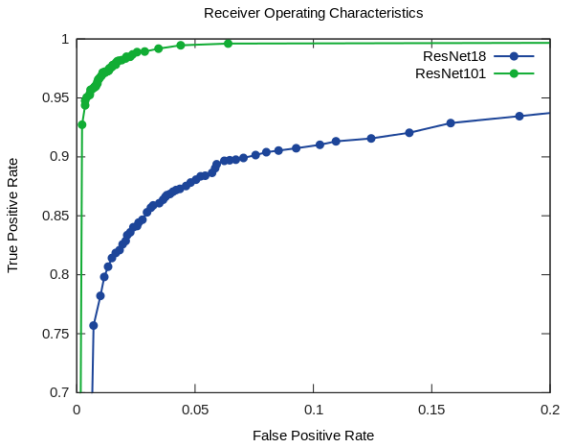

Figure 5: *ROC curve for small data set, showing the ResNet101 in green, and ResNet18 in blue.*

Looking at figure 5, one can clearly see that ResNet101 is the most optimal neural network for the diagnosis of breast cancer when compared to ResNet18. This is due to ResNet101 having a true positive (sensitivity) of 93% at less than 5% false positive rate. This may be attributed to the difference in layers between ResNet101 and ResNet18, where deeper neural network would outperform smaller network at the cost of efficiency.

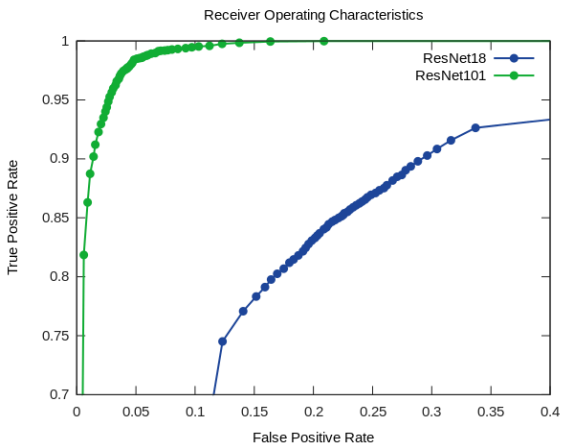

Figure 6: *ROC curve for larger dataset, where ResNet101 is the green curve and ResNet18 is in blue.*

When the ROC curve, figure 6, for the larger dataset, it can be noted that in both neural network performance dropped as a larger dataset was introduced. In the case of ResNet101, at a false positive rate of less than 5%, its sensitivity decreased to approximately 82%.

4. Discussion

Utilizing ResNet18 and ResNet101, it was possible to utilize deep learning network that could detect the presence of tumor cells within a whole slide image at an accuracy rate of 82.41% and 96.73%, respectively. This was achieved through the generation of 224x224 patches of the whole slide image using the GDAL. However, when visually inspecting some of the images present within the dataset, it could be noted that there did exist an error rate of misclassification in the patch generation. This misclassification in the dataset can be explained with the masking method employed to generate the tumor mask, where in some cases the tumor mask extended further then what would be classified as tumor. This issue could be fixed through magnification of the whole-slide image prior to the generation of the mask. This would allow for clearer and distinct border to exist between the two classifications. In its current state, the dataset generated from the Camelyon 16 challenge would not pass FDA regulation which requires the labeling of the dataset to be done by a panel of physicians [9]. Furthermore, the current networks were only trained utilizing two classes; noncancerous (1), and cancerous (2). To improve diagnosis of breast cancer utilizing deep-learning algorithms, further classes should be introduce to increase discrimination between different cancer histology samples, such as breast cancer sarcoma.

In its current state, although a high patch accuracy has been met using ResNet101, the clinical application are limited by the size and efficiency of the neural network. At this point, it is important to note that a deep-learning based approach is not recommended as a stand-alone solution. However, if implemented into the workflow as a pre or post-analysis for pathologist, it would be possible to improve the accuracy of diagnosis worldwide, and increase its reproducibility. This would be especially true in countries such as Thailand which have diagnosis accuracy rates of 91.3% which is lower than the best results from ResNet18 and ReNet101 [6].

5. References

[1] Myron Schultz, "Rudolf Virchow," *Emerg Infect Dis.* vol. 14, no. 9, pp. 1480-1481. Sep. 2008.

[2] N. Farahani, A.V. Parwani, L. Patanowitsz, "Whole slide imaging in pathology: advantages, limitations, and emerging perspectives," *Path. & Laboratory Medicine Internation,* vol. 7, pp. 23-33. 2015.

[3] Consortium for Open Medical Image Computing. (2016). *Data* [Online]. Available: https://camelyon16.grand-challenge.org/

[4] National Health Institute. (2016). *SEER Stat Fact Sheet: Female Breast Cancer* [Online]. Available: https://seer.cancer.gov/statfacts/html/breast.html

[5] Dr. C. Babbs. Lecture slides on Breast Cancer, Oct. 2016.

[6] Y. Gong, "Breast Cancer: Pathology, Cytology, and Core Needle Biopsy Methods for Diagnosis," *Breast and Gynecological Cancer*. pp. 19-37. Feb. 26, 2013.

[7] The American Registry of Radiologic Technologists. (2008). *Mammography Practice Analysis* [Online]. Available: https://www.arrt.org/pdfs/Examinations/MAM-Practice-Analysis.pdf

[7] D. Wang, et al., (2016, Jun. 18). *Deep Learning for Identifying Metastatic Breast Cancer* [Online]. Available: https://arxiv.org/pdf/1606.05718v1.pdf

[8] Food and Drug Administration. (2015, July 9). *Guidance for Industry and Food and Drug Administration Staff- Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions* [Online]. *Available: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187249.htm#4*