

# Written Review — Deep Learning for Identifying Metastatic Breast Cancer

Adithya Ganesh, Armin Pourshafeie, Behrooz Ghorbani, Karen Yang, Philip Hwang, Wendy Liu

December 5, 2016

## 1 Overview

This paper describes the highest performing algorithm in the International Symposium on Biomedical Imaging (ISBI) Camelyon Grand Challenge [7]. This research competition focuses on using machine learning to improve cancer detection from digital whole slide images (WSIs). To detect cancerous regions, current practice requires manual evaluation of patient images. This practice is not quantitative, it is subject to manual errors, and is highly labor intensive. The authors attempt to leverage recent advances in image classification to build an automatic, low-cost and robust pipeline that could provide an alternative for the traditional approach.

## 2 Methods

### 2.1 Data

In the competition, the participants all used Camelyon16 dataset provided by the organizers. This dataset contains 400 whole slide images (WSIs) split into 270 training and 130 test examples. The data is provided by two institutions: Radboud UMC and UMC Utrecht. From the batch coming from Radboud UMC 160 are used for training (90 cancer, 70 normal) and 80 slides are kept for testing. From the UMC Utrecht batch, 110 are used for training (70 cancer and 40 normal) and 130 are used for testing.

### 2.2 Evaluation Metric

The competition entails two stages. In the first stage, the competitors are evaluated on their ability to identify cancer in the slide. For this section the participants assign a cancer probability to each slide and get evaluated based on the AUC of the ROC curve (slide-based evaluation). In the second part of the competitions, the participants were evaluated on their ability to identify the correct lesions (lesion-based evaluation). In this part, each participant submitted a probability and a location for the cancerous region within a WSI. The scoring metric was computed as the average sensitivity at 6 predefined false positive rates:  $\frac{1}{4}$ ,  $\frac{1}{2}$ , 1, 2, 4, and 8 false positives per WSI.

### 2.3 Algorithms

To reduce computation time and extract the most informative portions of the image, the algorithm first excludes background white space using a threshold-based segmentation method. The cancer metastasis detection framework consists of a patch-based classification stage and a heatmap-based post-processing stage.

#### Patch-based Classification:

To train the patch-based classifier, 256x256 pixel patches were extracted from the positive and negative regions of the WSI. Subsequently, four well-known deep learning networks were trained for the patch classification task. After evaluation, the two deeper networks, GoogLeNet [6] and VGG16 [4] outperformed the other two. They adopted GoogLeNet as their deep network structure due to its speed and stability. GoogLeNet is a deep convolutional architecture consisted of 27 layers and more than 6 million parameters.

Furthermore, various magnification levels, including 40x, 20x and 10x were evaluated. The best results was obtained from 40x magnification, which is the only one used in the report for the Camelyon competition.

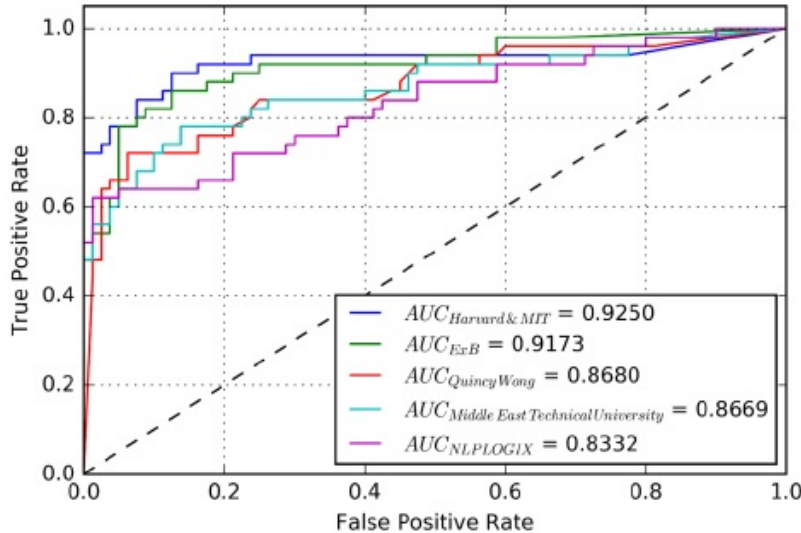


Figure 1: ROC curve of Slide-based Classification

Furthermore, the authors noticed that a significant proportion of errors were due to false positive classification of histologic mimics of cancer. To remedy this, additional training examples from these negative regions were added to the training dataset. Therefore, the final model was retrained on a dataset enriched for these *hard negative* examples.

### Heatmap-based Post-processing:

Patch-based classification generates a tumor probability heatmap for each WSI where each pixel contains a value between 0 and 1, indicating the probability that the pixel contains tumor. Post-processing is then performed to compute slide-based and lesion-based scores for each heatmap.

**a. Slide-based Classification:** 28 geometrical and morphological features were extracted from each heatmap, including the percentage of tumor region over the whole tissue region, the area ratio between tumor region and the minimum surrounding convex region, the average prediction values, and the longest axis of the tumor region. These features are computed over the tumor probability heatmaps across all training cases. A random forest classifier is then trained to discriminate the WSIs with metastases from the negative WSIs. On the test cases, their slide-based classification method achieved an AUC of 0.925, making it the top-performing system for the slide-based classification task in the Camelyon grand challenge.

**b. Lesion-based Detection:** For this task two deep models were trained: the first model (D-I) uses the initial training dataset described above and the second model (D-II) trained on a set enriched for tumor-adjacent negative regions. D-II produces fewer false positives than D-I but has reduced sensitivity. First the heatmap produced from D-I was thresholded at 0.900, which creates a binary heatmap. Connected components are identified within the tumor binary mask, and the central point is used as the tumor location for each connected component. To estimate the probability of tumor at each of these  $(x, y)$  locations, they take the average of the tumor probability predictions generated by D-I and D-II across each connected component.

## 3 Results

Figure 1 shows the achieved ROC curve of the top performing competitors. It can be seen that in the low false-positive regime, the winning algorithm heavily outperforms the competitors. Figure 2 shows the sensitivity-false positive curve. We can see that the pathologist did not make any false positive predictions. Also if we are willing to make some false positive predictions, it is possible to marginally improve the pathologists sensitivity.

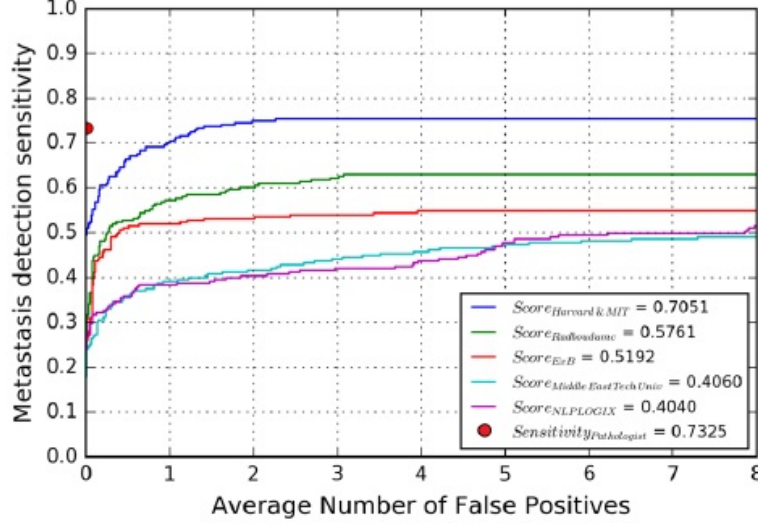


Figure 2: Free-Response Receiver Operator Characteristic (FROC) Curve of the Lesion-based Detection

## 4 Discussion and Future Research

The authors built a promising model that performed well in detecting metastatic breast cancer in digital whole slide images (WSIs) of sentinel lymph node biopsies. However, there are a few limitations to their experiments. Hence we propose the following extensions and improvements.

For the patch-based classification task, the researchers have applied traditional convolutional neural network. An important extension could apply new architectures such as the residual neural network [1], which introduces “shortcut connections” that compute the mapping  $\mathcal{F}(\mathbf{x}) + \mathbf{x}$ . This model demonstrates the tractability of training extremely deep neural networks, and has achieved state-of-the-art performance on the ImageNet research benchmark.

The batch normalization technique, which reduces internal covariate shift for each mini-batch, has recently been shown to accelerate training times and act as a regularizer [2]. Moreover, this strategy could reduce the impact of imbalanced training data, which is a common scenario in the biomedical setting.

Generative adversarial networks [3] have been shown to learn robust, general image representations. Since obtaining labeled data can be expensive, using a generative approach can facilitate neural network pretraining and training set expansion [5].

Finally, the authors have trained nets using 4 different convolutional architectures all of which performed reasonably well. Using an ensemble of these classifiers (and different resolutions) may improve the performance of the system.

## References

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.
- [2] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

- [3] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *arXiv preprint arXiv:1409.1556*, 2014.
- [5] Leon Sixt, Benjamin Wild, and Tim Landgraf. Rendergan: Generating realistic labeled data. *arXiv preprint arXiv:1611.01331*, 2016.
- [6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [7] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*, 2016.