

**Ilchenko Vladyslav**

*Orthopedic Surgeon and Pathology Researcher  
Clinical Diagnostic Centre of Svyatoshynskiy District*

## **USING DEEP LEARNING ALGORITHMS FOR BREAST CANCER DIAGNOSIS IN PATHOLOGY**

***Summary.** Breast cancer is one of the main causes of cancer death worldwide. The diagnosis of biopsy tissue with hematoxylin and eosin stained images is non-trivial and specialists often disagree on the final diagnosis. Computer-aided Diagnosis systems contribute to reduce the cost and increase the efficiency of this process. We propose a deep learning architecture, based on VGG-16 network for learning high-level image representation to achieve high classification accuracy with low variance in medical image binary classification tasks by dividing whole-slide images on 50x50 pixel histology image patches. We aim to learn discriminant compact features at beginning of our deep convolutional neural network.*

***Key words:** breast cancer, biopsy, deep learning, machine learning, VGG16, convolutional neural network.*

**Background.** Breast cancer is the leading cause of cancerous death in women in less-developed countries, and is the second leading cause of cancerous deaths in developed countries, accounting for 29% of all cancers in women within the U.S. [1]. Survival rates increase as early detection increases, giving incentive for pathologists and the medical world at large to develop improved methods for even earlier detection [2]. Breast cancer diagnosis usually consists in an initial

detection via palpation and regular check-ups using mammography or ultrasound imaging. The diagnosis is then followed by breast tissue biopsy if the check-up exam indicates the possibility of malignant tissue growth [3].

Breast tissue biopsies allow the pathologists to histologically assess the microscopic structure and elements of the tissue. The histology allows to distinguish between normal tissue, non-malignant (benign) and malignant lesions and to perform a prognostic evaluation [4, pp. 403-410]. Benign lesions represent changes in normal structures of breast parenchyma that are not directly related with progression to malignancy. Carcinomas can be classified as in situ or invasive. There are many forms of breast cancer including Ductal Carcinoma in Situ (DCIS), Invasive Ductal Carcinoma (IDC), Tubular Carcinoma of the Breast, Medullary Carcinoma of the Breast, Invasive Lobular Carcinoma, Inflammatory Breast Cancer and several others [5].

In in situ carcinoma the cells are restrained inside the mammary ductal-lobular system, whereas in invasive carcinoma the cells spread beyond that structure. The tissue collected during the biopsy is commonly stained with hematoxylin and eosin (H&E) prior to the visual analysis performed by the specialists. During this procedure, relevant regions of whole-slide tissue scans are assessed [6, pp. 147-171].

The diagnosis process using H&E stained biopsies is not trivial, and the average diagnostic concordance between specialists is approximately 75% [7, pp. 313]. The manual examination of histology images requires intense workload of highly specialized pathologists. The subjectivity of the application of morphological criteria in usual classification motivates the use of computer-aided diagnosis (CAD) systems to improve the diagnosis efficiency and increase the level of inter-observer agreement [8, pp. 236-251].

Deep learning has recently seen enormous success in challenging problems such as object recognition in natural images, automatic speech recognition and machine translation [9, p. 521]. This success has prompted a surge of interest in applying deep convolutional networks (DCN) to medical imaging. Many recent studies have shown the potential of applying such networks to medical imaging, including breast screening mammography; however, without investigating the fundamental differences between medical and natural images and their impact on the design choices and performance of proposed models. For instance, many recent works have either significantly downscaled a whole image or focused on classifying a small region of interest. Furthermore, the potential of DCNs has only been assessed in limited settings of small data sets often consisting of less than 1000 images, while the success of such networks in natural object recognition is largely attributed to the availability of more than 1M annotated images. This further hinders our understanding of the true potential of DCNs in medical imaging, particularly in breast cancer screening.

A deep convolutional neural network [10, pp. 541-551; 11, pp. 2278–2324] is a classifier that takes an image  $x$  as input, often with multiple channels corresponding to different colors (e.g., RGB), and outputs the conditional probability distribution over the categories. This is done by a series of nonlinear functions that gradually transform the input pixel-level image. A major property of the deep convolutional network, which distinguishes it from a multi-layer perceptron, is that it heavily relies on convolutional and pooling layers, which make the network invariant to local translation of visual features in the input.

## **Proposed Approach**

### Datasets

We evaluate our DCNs model on this data sets:

1. 500 breast cancer cases from The Cancer Genome Atlas(Whole-slide images are stored in the Aperio .svs file format as multi-resolution pyramid structures (the size of the highest resolution image exceed 50000 by 50000 pixels)) [12].
2. 23 breast cancer cases from Department of Pathology at the University Medical Center in Utrecht, The Netherlands(images in the TIFF format with resolution 5657x5657 pixels) [13].
3. 162 breast cancer cases from Case Western Reserve University in Cleveland, Ohio(preprocessed by nice people at Kaagle and divided on 275215 50x50 pixel patches) [14]

Slides from the 1-st and 2-d sources was also divided on 50x50 pixel patches.

The final dataset consists of 975717 50x50 pixel patches. These images are labeled with two classes:

1. Normal tissue or benign lesions (label 0) – 578085 patches;
2. In situ carcinoma, and invasive carcinoma (label 1) – 397632 patches.

We divided the data into training and validation sets, with an 80-20 splitting rule:

1. Training set – 780574 patches
  - a. Normal tissue or benign lesions (label 0) – 462468 patches;
  - b. In situ carcinoma and invasive carcinoma (label 1) – 318106 patches;
2. Validation set – 195143 patches
  - a. Normal tissue or benign lesions (label 0) – 115617 patches;
  - b. In situ carcinoma and invasive carcinoma (label 1) – 79526 patches.

### Model

We using Keras for model training (currently transfer learning by fine-tuning a modified Vgg16 model). Ultimately, we aim to develop a model that is

sufficiently stronger than existing approaches for the task of breast cancer tumor proliferation score classification.

Our model takes as input the pixel values of the individual samples and is trained to predict the probability is this sample from the malignant tumor or from normal tissue.

We chose a VGG-16 network to classify the  $50 \times 50$  histology image patches in order to explore the scale and organization features of nuclei and the scale features of the overall structure. Because they do not have complicated high-level semantic information, a 16-layer structure suffices.

To leverage contextual information from the cropped images, we added global context to the last convolutional layer of the VGG networks. The input images are passed to two independent branches, our VGG network and a global average pooling layer. The transformed output of the global pooling layer is unpooled to the same shape as that of the feature maps after the last convolutional layer of the VGG network and is then concatenated with the feature maps. These two feature maps are then fused by another  $1 \times 1$  convolutional layer and then passed through three fully-connected (FC) layers for classification.

Magnification is an important factor for analyzing microscopic images for diagnosis. The most informative magnification level is still debatable, so we've included two possible scales in our work for comparison.

### Methodology

We utilized the Amazon AWS GPU resources to run the whole program. The AWS g3 instance has NVIDIA Tesla M60 GPUs, each with 2048 parallel processing cores and 8 GiB of video memory which enables to run a complicated network structure. The framework for deep learning is the latest slim version of Tensorflow and Keras above it. We applied the transfer learning method to our model which is by using the pre-trained VGG16 model as a checkpoint and

continue to train the neural network. The reason that we can do so is because the ImageNet has very large amount of dataset and is trained quite well. We can make full use of the pre-trained model and get the feature based on the ImageNet dataset.

### Performance measurements

The performance of a method, in medical image analysis, is typically measured by specificity, sensitivity, and F1 score. Sensitivity measures the proportion of actual positives patches (with in situ carcinoma, and invasive carcinoma) that correctly identify. Therefore, it is computed according to the following definition:

$$sensitivity = \frac{TP}{TP+FN} [15, pp. 198-201]$$

where TP(true positive) is the number of positives patches with carcinoma which have been successfully detected, and FN(false negative) is the number of samples with carcinoma which have not been detected by the method. In contrast, specificity measures the proportion of identified negatives patches, in which the percentage without carcinoma is correctly classified as non cancerous. In this manner, specificity is computed as:

$$specificity = \frac{TN}{TN+FP} [16, p. 14]$$

where TN(true negative) is the number of non-cancer patches(normal tissue or benign lesions) which have been successfully classified, and FP(false positive) is the number of non-cancer patches which have been wrongly classified as cancer.

And, F1-score measures the average F1 score through different class labels which is computed as:

$$F1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} [17, p.16]$$

where PPV:

$$PPV = \frac{TP}{TP+FP} [18, pp. 445-448]$$

and TPR:

$$TPR = \frac{TP}{TP+FN} [19, p. 13]$$

**Results.** In medical image analysis, the performances of a method is typically measured by using sensitivity (the true positive rate) and specificity (the true negative rate) instead of measuring direct accuracy of the classification.

In the data science competitions (like Kaggle competitions), the competitors were evaluated their results using the log-loss metric. To assess overall classification relevance, we therefore compute the F1-score, too.

The results of our proposed approach illustrated in 2 tables. Table 1 shows the overall results and Table 2 shows results in three metric measurements.

Table 1

**The overall results**

	<b>In situ carcinoma and invasive carcinoma</b>	<b>Normal tissue or benign lesions</b>
Predicted in situ carcinoma and invasive carcinoma	69,983 (TP)	2,312 (FP)
Predicted normal tissue or benign lesions	9,543 (FN)	113,305 (TN)

Table 2

**The results in three metric measurements**

	<b>Sensitivity</b>	<b>Specificity</b>	<b>F1</b>
Results	0.88	0.98	0,92

**Conclusion.** In this paper, we proposed a deep convolutional neural network architecture for binary classification of breast tissue biopsies that had aimed to learn high-order convolutional features at the initial layers of the model.

The proposed DCNs architecture not only proved the power of itself, but it also proved that the features in the breast tissue biopsies can be learned and compacted at the preliminary layers of a deep model.

In terms of future work, we plan to extend the proposed deep models capabilities by integrating the automatic segmentation of breast tissue biopsies in the above-mentioned data set.

### **References**

1. [Electronic resource]. - Access mode: <http://emedicine.medscape.com/article/1947145-overview#a3>
2. [Electronic resource]. - Access mode: <http://emedicine.medscape.com/article/1947145-overview#a7>
3. [Electronic resource]. - Access mode: NationalBreastCancerFoundation. Breast Cancer Diagnosis; 2015. Available from: <http://www.nationalbreastcancer.org/breast-cancer-diagnosis>.
4. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991 [Electronic resource]. - Access mode: <https://doi.org/10.1111/j.1365-2559.1991.tb00229.x> PMID: 1757079
5. [Electronic resource]. - Access mode: <http://emedicine.medscape.com/article/1947145-workup#c12>
6. Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological Image Analysis: A Review. *IEEE Reviews on Biomedical Engineering*. 2009; <https://doi.org/10.1109/RBME.2009.2034865>
7. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson ANA, et al. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. *Jama*. 2015; <https://doi.org/10.1001/jama.2015.1405> PMID: 25781441



8. Tang J, Rangayyan RM, Xu J, Naqa IE, Yang Y. Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. *IEEE Transactions on Information Technology in Biomedicine*. 2009; <https://doi.org/10.1109/TITB.2008.2009441> PMID: 19171527
9. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015.
10. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, 1989.
11. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, 1998.
12. [Electronic resource]. - Access mode: <http://cancergenome.nih.gov/>
13. [Electronic resource]. - Access mode: <https://www.umcutrecht.nl/en/Subsites/UMC-Utrecht-Lab/About-us/Pathology>
14. [Electronic resource]. - Access mode: <https://www.kaggle.com/paultimothymooney/predict-idc-in-breast-cancer-histology-images/data>
15. K. Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4), 2007.
16. M. J. DCruz, M. A. Jadhav, M. A. Dighe, M. V. Chavan, and J. Chaudhari. Detection of lung cancer using backpropagation neural networks and genetic algorithm.
17. J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. Parkin, D. Forman, and F. Bray. Cancer incidence and mortality

worldwide: Iarc cancerbase no. 11 [internet]. lyon, france: International agency for research on cancer. globocan. 2013; 2012 v1. 0. Available from: <http://globocan.iarc.fr>, 2015.

18.R. Hecht-Nielsen et al. Theory of the backpropagation neural network. *Neural Networks*, 1(Supplement-1), 1988.

19.A. Jemal, P. Vineis, F. Bray, L. Torre, and D. Forman. *The cancer atlas*. American Cancer Society, 2016.