# Breast Cancer Classification in Histopathological Images using Convolutional Neural Network

Mohamad Mahmoud Al Rahhal

Information Science
College of Applied Computer Science, King Saud University
Riyadh, KSA

*Abstract*—**Computer based analysis is one of the suggested means that can assist oncologists in the detection and diagnosis of breast cancer. On the other hand, deep learning has been promoted as one of the hottest research directions very recently in the general imaging literature, thanks to its high capability in detection and recognition tasks. Yet, it has not been adequately suited to the problem of breast cancer so far. In this context, I propose in this paper an approach for breast cancer detection and classification in histopathological images. This approach relies on a deep convolutional neural networks (CNN), which is pretrained on an auxiliary domain with very large labelled images, and coupled with an additional network composed of fully connected layers. The network is trained separately with respect to various image magnifications (40x, 100x, 200x and 400x). The results presented in the patient level achieved promising scores compared to the state of the art methods.**

*Keywords—Convolutional neural network (CNN); histopathological images; imagenet; classification*

## I. INTRODUCTION

According to reports of the world health organization (WHO), breast cancer (BC) is the most prevalent type of cancer in women. For instance, incidence rates range from 19.3 per 100,000 women in Eastern Africa to 89.7 per 100,000 women in Western Europe [1]. Current scientific findings indicate that such high variability might be traced back to differences in lifestyle and urbanization. Although early diagnosis is more affordable in developed countries, it is less likely in underdeveloped nations, which implies that undertaking preventive measures only does not offer a cutting-edge solution.

Mammography is a common screening protocol that can help distinguish dubious regions of the breast, followed by a biopsy of potentially cancerous areas in order to determine whether the dubious area is benign or malignant [2], [3]. In order to produce stained histology slides, samples of tissue are taken from the breast during biopsy. In spite of the considerable improvement incurred by such imaging technologies, pathologists tend to visually inspect the histological samples under the microscope for a final diagnosis, including staging and grading [4].

In this context, automatic image analysis is prone to play a pivotal role in facilitating the diagnosis; so far, the relevant processing and machine learning techniques. For instance, the authors in [5] present a comparison of different algorithms of nuclei segmentation, where the cases are categorized into benign or malignant.

Deep CNNs learn mid-level and high-level representations obtained from raw data (e.g., images) in an automatic manner. Recent results on natural images indicate that CNN representations are highly efficient in object recognition and localization applications. This has instigated the adoption of CNNs in the biomedical field, such as breast cancer diagnosis and masses classification [6]–[11], abdominal adipose tissues extraction [12], detection and classification of brain tumour in MR images [13]–[16], skeletal bone age assessment in X-ray Images [17], EEG classification of motor imagery [18], and arrhythmia detection and analysis of the ECG signals [19]–[21]. In particular, in [9], the authors propose a framework for masses classification, which mainly encompasses a CNN and a decision mechanism for breast cancer diagnosis as either benign or malignant in a DDSM mammographic dataset. In [4], the authors propose an improved hybrid active contour model based segmentation method for nuclei segmentation. They adopt both pixel and object-level features in addition to semantic-level features. The semantic-level features are computed using a CNN architecture which can learn additional feature representations that cannot be represented through neither pixel nor object-level features.

Thus, it is to stress the fact that, relatively to other biomedical applications, breast cancer diagnosis has not benefited enough from deep learning, which inspired us to investigate it thoroughly. In particular, I opt for several deep architectures in the context of breast cancer histological image classification, and demonstrate that the common belief that high level deep features are more capable of capturing the contextual as well as spectral attributes in optical images remains valid also in histological breast cancer images. This, in fact, is confirmed by the very satisfactory results reported hereby, which advance late works often by large margins.

The rest of this paper is organized as follows. Material and methods are exposed in Section II. Results are presented in Section III. Finally, conclusion is provided in Section IV.

## II. MATERIAL AND METHODS

### A. Dataset Description

In order to realistically assess any BC diagnosis system, the experiments shall be performed on a large-scale dataset accommodating 1) numerous patients, 2) abundant images. The

latter component is essential to any deep learning model as large data is required for the training phase.

The Breast Cancer Histopathological Image Classification (BreakHis), which was established recently in [22], is an optimal dataset as it meets all the above requirements. Precisely, it is composed of 9,109 microscopic images of breast tumour tissue collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). For convenience, Fig. 1 and 2 display a slide of breast benign and malignant tumour for the same patient seen in different magnification factors.

To date, the dataset contains 2,480 benign (taken from 24 patients) and 5,429 malignant samples (taken from 58 patients) of 700X460 pixels, 3-channel RGB, 8-bit depth in each channel, and PNG format. In its current version, samples present in the dataset were collected by SOB method, also named partial mastectomy or excisional biopsy. This type of procedure, compared to any methods of needle biopsy, removes the larger size of tissue sample and is performed in a hospital with general anaesthetic. This dataset is structured as shown in Table I.

TABLE I.    BREAST CANCER DATABASE USED IN THE EXPERIMENT

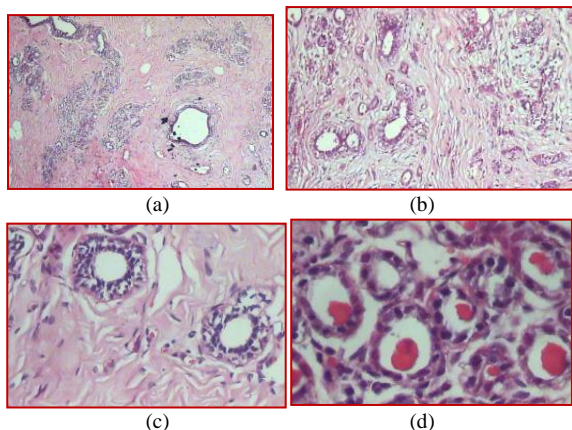| | Magnification | | | | Total | Patient |
|---|---|---|---|---|---|---|
| | 40x | 100 x | 200x | 400x | | |
| Benign | 652 | 644 | 623 | 588 | 2480 | 24 |
| Malignant | 1370 | 1437 | 1390 | 1232 | 5429 | 58 |
| Total | 1995 | 2081 | 2013 | 1820 | 7909 | 82 |



Fig. 1.    A slide of breast benign for the same patient seen in different magnification factors: (a) 40X, (b) 100X, (c) 200X, and (d) 400X.
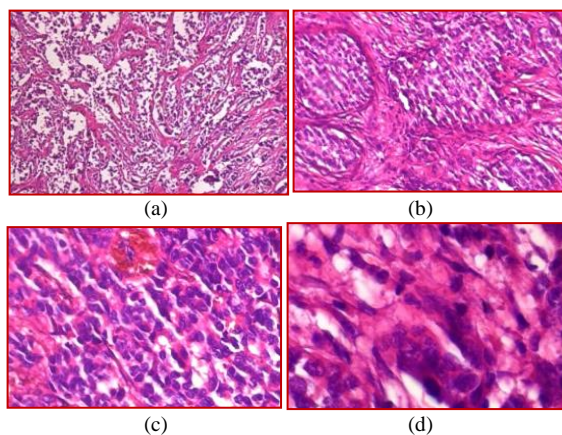


Fig. 2.    A slide of breast malignant tumor for the same patient seen in different magnification factors: (a) 40X, (b) 100X, (c) 200X, and (d) 400X.

### B. Proposed Methodology

Let us consider $D^{(s)} = \{x_i, y_i\}_{i=1}^{n_s}$ as the labeled source data and $y_i \in \{1,2\}$ is its corresponding class label either to be benign or malignant. Similarly, let us refer to $D^{(t)} = \{x_j\}_{j=1}^{n_t}$ as the unseen target data. This paragraph consists of two steps as shown in Fig. 3.

#### 1) Feature Extraction

Deep CNNs are composed of multiple layers of processing which are learnt jointly, in an end-to-end manner, to address specific issues [23]–[25]. Particularly, Deep CNNs are commonly composed of four types of layers namely convolution; normalization, pooling and fully connected. The convolutional layer is considered the main building block of the CNN and its parameters consist of a set of filters (or sometimes referred to as a neuron or a kernel). Every filter is small spatially (along width and height), but extends through the full depth of the input image. The output of this layer is called activation maps or feature maps which are produced via sliding the filters across the input image. The feature maps are then fed to a non-linear gating function such as the Rectified Linear Unit (ReLU). Then the output of this activation function can further be subjected to normalization layer to help in generalization. Regarding the pooling layers, they are usually used immediately after convolutional layers in order to control overfitting and reduce the amount of parameters in the network.
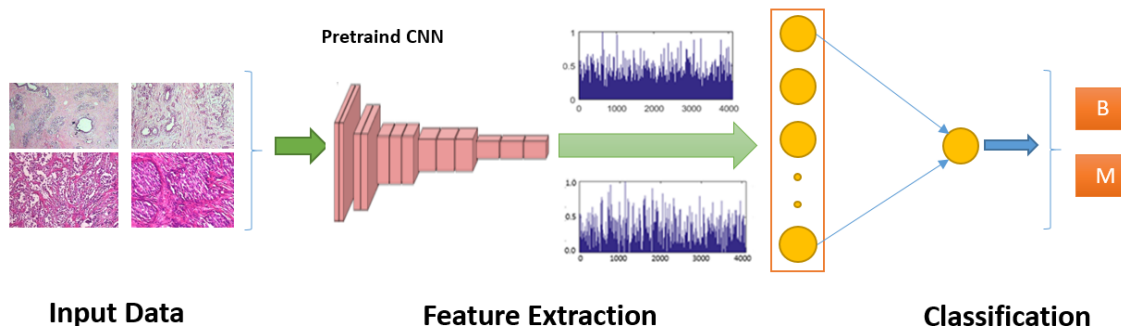

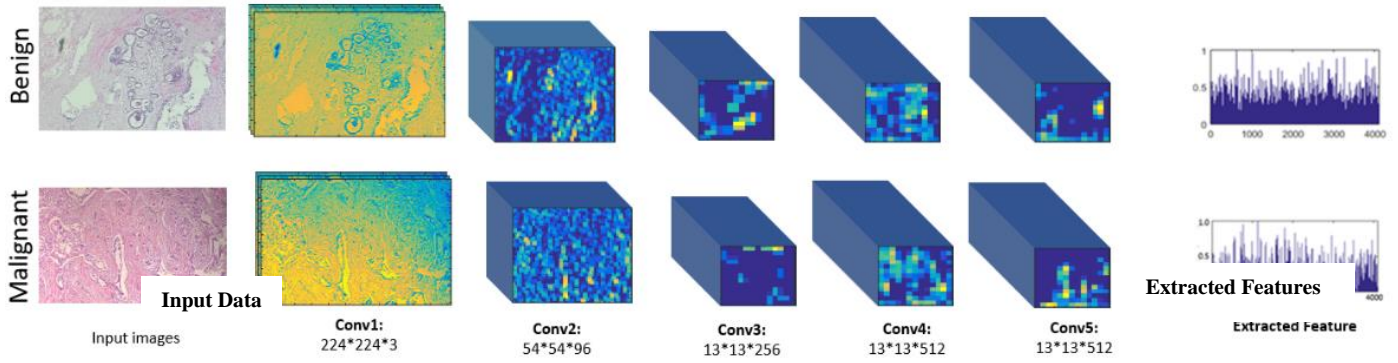
Fig. 3.    Flowchart of the proposed approach.

Fig. 4.   Convolutional layers feature maps with corresponding extracted feature.

In this work, I follow the recent approaches for exploiting pretrained CNN models by taking the output of the last fully connected layer (before the sigmoid layer) to represent the images. That is I feed each image $\mathbf{I}_i$ as input to the network and generate its corresponding CNN feature representation vector $\boldsymbol{x}_i \in \mathfrak{R}^d$ of dimension $d$:

$$\boldsymbol{x}_i = f_{L-1}^{\text{CNN}}\left(\ldots f_2^{\text{CNN}}\left(f_1^{\text{CNN}}(\mathbf{I}_i)\right)\right), \ i = 1, \ldots, n_s \qquad (1)$$

Where $f_j^{\text{CNN}}, j = 1, \ldots, L-1$ represent the functions defining the different layers of CNN, $L$ is the total number of layers, and $n_s$ and $n_t$ represent the number of labeled source images and unlabeled target images, respectively. Fig. 4 shows feature maps with corresponding extracted features.

*2) Classification*

I feed the CNN feature vectors from the previous stage to an extra network placed on the top of the pretrained CNN as shown in Fig. 3. Specifically, this network is composed of two fully-connected layers, a hidden layer followed by binary classification layer: a sigmoid layer. The hidden layer maps the input $\boldsymbol{x}_i$ to another representation $\boldsymbol{h}_i^{(1)} \in \mathfrak{R}^{d^{(1)}}$ of dimension $d^{(1)}$ through the nonlinear activation function $f$ as follows:

$$\boldsymbol{h}_i^{(1)} = f(\boldsymbol{W}^{(1)}\boldsymbol{x}_i) \qquad (2)$$

Where mapping weight matrix referred as $\boldsymbol{W}^{(1)} \in \mathfrak{R}^{d^{(1)} \times d}$. I adopt the sigmoid function i.e. $f(v) = 1/(1 + \exp(-v))$ as a nonlinear activation function. For simplicity, the bias vector in the expression is ignored as it can be incorporated as an additional column vector in the mapping matrix, whereby the feature vector is appended by the value 1.

I adopt the recently introduced dropout technique [26] to increase the generalization ability of the network and prevent it from overfitting.

### III.   RESULTS

*A. Experimental Results*

For the sake of comparison, I followed the protocol proposed in [22], the dataset has been divided into training (70%) and testing (30%) sets taking in consideration that the patient used in training and testing sets are mutually exclusive for training i used 58 patients (17 Benign and 41 Malignant), for test use used 24 patients (7 Benign and 17 Malignant) for test. The adopted protocol was applied independently to each of the four different available magnifications factor (40X, 100X, 200X, and 400X) in the patient level, the recognition rate is computed as follows:

$$\text{Patient Recognition Rate} = \frac{\sum \text{Patient score}}{\text{Total Number of Patient}} \qquad (3)$$

While

$$\text{Patient score} = \frac{C_{rec}}{C_N} \qquad (4)$$

Here, $C_N$ be the number of cancer images of patients, and $C_{rec}$ represent the cancer images that are correctly classified.

For the pretrained CNN, i explore VGGm model [27] which composed of 8 layers, and uses five convolutional filters of dimensions (number of filters $\times$ filter height $\times$ filter depth: $96 \times 7 \times 7$, $256 \times 5 \times 5$, $512 \times 3 \times 3$, $512 \times 3 \times 3$, and $512 \times 3 \times 3$) and three fully connected layers with the following number of hidden nodes (fc1: 4096, fc2: 4096, and softmax: 1000). This network was pretrained on the ILSVRC-12 challenge dataset. I recall that the ImageNet dataset used in this challenge is composed of 1.2 million RGB images of size $224 \times 224$ pixels belonging to 1000 classes and these classes describe general images such as beaches, dogs, cats, cars, shopping carts, minivans, etc. As can be clearly seen, this auxiliary dataset is completely different from the ECG signals used in the experiments.

For training the extra network placed on the top of the pretrained CNN, I follow the recommendations of [28] for training neural networks. I set the dropout probability $p$ to 0.5. I use a sigmoid activation function for the hidden layer. For the backpropagation algorithm, I use a mini-batch gradient optimization method with the following parameters (i.e., learning rate: 0.01, momentum: 0.5, and mini-batch size: 50). The weights of the network are set initially in the range [-0.005 0.005].

TABLE II.    CLASSIFICATION ACCURACIES (%) WITH MAGNIFICATION 40X

| Test | Patient no. | 40x | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Benign | 1 | 100 | 100 | 91.9 | 73.3 | 96.2 |
| | 2 | 65.7 | 30.6 | 100 | 100 | 48.6 |
| | 3 | 96.8 | 53.8 | 100 | 37.9 | 20.0 |
| | 4 | 100 | 100 | 93.3 | 100 | 100 |
| | 5 | 100 | 100 | 51.7 | 91.3 | 100 |
| | 6 | 84.6 | 38.5 | 100 | 94.6 | 92.9 |
| | 7 | 18.2 | 48.3 | 0.0 | 9.1 | 0.0 |
| Malignant | 8 | 100 | 96.9 | 90.0 | 100 | 85.7 |
| | 9 | 100 | 100 | 100 | 100 | 68.0 |
| | 10 | 100 | 82.8 | 82.8 | 78.3 | 100 |
| | 11 | 96.9 | 95.2 | 100 | 94.7 | 100 |
| | 12 | 100 | 100 | 90.6 | 100 | 100 |
| | 13 | 100 | 100 | 100 | 97.3 | 100 |
| | 14 | 82.6 | 95.0 | 100 | 100 | 100 |
| | 15 | 92.3 | 57.1 | 74.2 | 91.4 | 100 |
| | 16 | 33.3 | 96.2 | 78.3 | 100 | 100 |
| | 17 | 100 | 100 | 100 | 100 | 100 |
| | 18 | 97.2 | 90.0 | 100 | 100 | 88.5 |
| | 19 | 98.0 | 100 | 100 | 100 | 100 |
| | 20 | 93.8 | 72.0 | 26.7 | 100 | 97.4 |
| | 21 | 23.5 | 94.4 | 23.5 | 100 | 100 |
| | 22 | 100 | 100 | 100 | 92.3 | 100 |
| | 23 | 100 | 100 | 71.9 | 90.0 | 100 |
| | 24 | 100 | 100 | 100 | 100 | 91.7 |
| Average | | 86.8 | 85.4 | 82.3 | 89.6 | 87.0 |

TABLE IV.    CLASSIFICATION ACCURACIES (%) WITH MAGNIFICATION 200X

| Test | Patient no. | 200x | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Benign | 1 | 100 | 100 | 100 | 0.0 | 84.2 |
| | 2 | 48.4 | 52.6 | 100 | 86.8 | 25.8 |
| | 3 | 100 | 73.7 | 100 | 100 | 6.3 |
| | 4 | 100 | 88.1 | 91.7 | 92.9 | 100 |
| | 5 | 100 | 94.6 | 100 | 60.0 | 100 |
| | 6 | 57.1 | 45.7 | 97.3 | 88.2 | 100 |
| | 7 | 12.5 | 87.5 | 12.5 | 97.3 | 0.0 |
| Malignant | 8 | 93.3 | 100 | 100 | 100 | 92.9 |
| | 9 | 100 | 100 | 100 | 100 | 77.8 |
| | 10 | 95.8 | 100 | 78.7 | 90.0 | 86.7 |
| | 11 | 97.1 | 94.7 | 86.7 | 92.9 | 100 |
| | 12 | 100 | 100 | 100 | 100 | 100 |
| | 13 | 100 | 96.4 | 94.1 | 95.5 | 100 |
| | 14 | 23.8 | 92.9 | 93.8 | 100 | 100 |
| | 15 | 100 | 85.7 | 84.8 | 100 | 97.1 |
| | 16 | 85.7 | 100 | 100 | 85.7 | 100 |
| | 17 | 100 | 86.7 | 100 | 100 | 100 |
| | 18 | 100 | 100 | 100 | 100 | 100 |
| | 19 | 94.5 | 80.0 | 100 | 91.2 | 100 |
| | 20 | 90.6 | 92.6 | 100 | 88.9 | 100 |
| | 21 | 0.0 | 96.9 | 100 | 76.2 | 100 |
| | 22 | 100 | 100 | 100 | 82.8 | 100 |
| | 23 | 100 | 23.8 | 94.4 | 40.9 | 100 |
| | 24 | 100 | 89.3 | 100 | 100 | 100 |
| Average | | 86.8 | 83.3 | 86.7 | 93.1 | 86.2 |

TABLE III.    CLASSIFICATION ACCURACIES (%) WITH MAGNIFICATION 100X

| Test | Patient no. | 100x | | | | |
| | | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|
| Benign | 1 | 72.7 | 72.7 | 72.7 | 72.7 | 72.7 |
| | 2 | 66.7 | 66.7 | 66.7 | 66.7 | 66.7 |
| | 3 | 100 | 100 | 100 | 100 | 100 |
| | 4 | 100 | 100 | 100 | 100 | 100 |
| | 5 | 100 | 100 | 100 | 100 | 100 |
| | 6 | 70.6 | 70.6 | 70.6 | 70.6 | 70.6 |
| | 7 | 21.2 | 21.2 | 21.2 | 21.2 | 21.2 |
| Malignant | 8 | 97.1 | 97.1 | 97.1 | 97.1 | 97.1 |
| | 9 | 100 | 100 | 100 | 100 | 100 |
| | 10 | 100 | 100 | 100 | 100 | 100 |
| | 11 | 92.7 | 92.7 | 92.7 | 92.7 | 92.7 |
| | 12 | 100 | 100 | 100 | 100 | 100 |
| | 13 | 100 | 100 | 100 | 100 | 100 |
| | 14 | 40.9 | 40.9 | 40.9 | 40.9 | 40.9 |
| | 15 | 100 | 100 | 100 | 100 | 100 |
| | 16 | 26.7 | 26.7 | 26.7 | 26.7 | 26.7 |
| | 17 | 95.2 | 95.2 | 95.2 | 95.2 | 95.2 |
| | 18 | 100 | 100 | 100 | 100 | 100 |
| | 19 | 90.7 | 90.7 | 90.7 | 90.7 | 90.7 |
| | 20 | 93.9 | 93.9 | 93.9 | 93.9 | 93.9 |
| | 21 | 87.5 | 87.5 | 87.5 | 87.5 | 87.5 |
| | 22 | 100 | 100 | 100 | 100 | 100 |
| | 23 | 100 | 100 | 100 | 100 | 100 |
| | 24 | 100 | 100 | 100 | 100 | 100 |
| Average | | 86.8 | 85.7 | 85.7 | 85.7 | 85.7 |

TABLE V.    CLASSIFICATION ACCURACIES (%) WITH MAGNIFICATION 400X

| Test | Patient no. | 400x | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Benign | 1 | 93.8 | 94.1 | 40.0 | 52.9 | 100 |
| | 2 | 66.7 | 89.2 | 88.2 | 88.5 | 50.0 |
| | 3 | 100 | 69.2 | 91.7 | 80.0 | 20.7 |
| | 4 | 94.1 | 79.5 | 53.3 | 84.6 | 88.9 |
| | 5 | 93.3 | 96.9 | 100 | 50.0 | 88.2 |
| | 6 | 70.8 | 45.8 | 96.9 | 71.4 | 100 |
| | 7 | 3.0 | 60.0 | 9.1 | 96.9 | 11.8 |
| Malignant | 8 | 100 | 92.3 | 93.1 | 100 | 80.0 |
| | 9 | 100 | 100 | 100 | 100 | 81.8 |
| | 10 | 100 | 100 | 85.7 | 100 | 73.3 |
| | 11 | 92.1 | 100 | 60.0 | 100 | 100 |
| | 12 | 100 | 100 | 100 | 100 | 100 |
| | 13 | 100 | 95.1 | 100 | 100 | 100 |
| | 14 | 54.2 | 92.9 | 100 | 100 | 100 |
| | 15 | 100 | 53.3 | 86.1 | 96.6 | 100 |
| | 16 | 45.5 | 100 | 100 | 92.9 | 100 |
| | 17 | 100 | 80.0 | 100 | 100 | 100 |
| | 18 | 100 | 100 | 100 | 100 | 93.5 |
| | 19 | 90.2 | 100 | 96.4 | 100 | 100 |
| | 20 | 96.2 | 100 | 95.8 | 100 | 100 |
| | 21 | 26.7 | 100 | 100 | 66.7 | 100 |
| | 22 | 100 | 100 | 100 | 71.4 | 100 |
| | 23 | 100 | 75.0 | 93.8 | 26.7 | 100 |
| | 24 | 100 | 100 | 100 | 60.0 | 100 |
| Average | | 86.8 | 84.4 | 88.5 | 87.1 | 84.9 |

TABLE VI. THE PATIENT-LEVEL CLASSIFICATION ACCURACIES (%), COMPARING OUR METHOD WITH EXISTING RESULTS ON THE BREAKHIS DATASET

| Retrained CNN | Magnification | | | | Average |
|---|---|---|---|---|---|
| | 40x | 100x | 200x | 400x | |
| [22] | 83.8±2.0 | 82.1±4.9 | 85.1±3.1 | 82.3±3.8 | 83.33 |
| CNN [29] | 83.1±2.1 | 83.2±3.5 | 84.6±2.7 | 82.1±4.4 | 83.25 |
| Vgg_m | 86.2±2.7 | 85.9±0.5 | 87.2±3.6 | 86.3±1.7 | 86.80 |

To present our results, I train the CNN networks depending on their magnifications (40x, 100x, 200x, and 400x) separately. The experiment was repeated five times as shown in Tables II-V, the average accuracy for five cases of the proposed CNN methods has been reported in Table VI which shows the superior accuracy of the proposed methods at patient level against state of the art methods in different magnification factors.

## IV. CONCLUSION

This paper proposed a deep learning framework for breast cancer detection and classification. The yielded results confirm that deep learning can incur large margin improvements with respect to handcrafted features. Although the presented method achieves plausible scores, it can benefit from further improvements, potentially by 1) customizing more deep models; and 2) fusing several deep architectures in order to elevate the performance. Another direction to undertake is to adopt active learning in order to raise the classification scores. Ultimately, domain adaptation is another research line that can introduce tangible improvements.

## ACKNOWLEDGMENT

### REFERENCES

[1] A. Jemal et al., "Cancer statistics, 2008," CA. Cancer J. Clin., vol. 58, no. 2, pp. 71–96, Apr. 2008.

[2] L. Adepoju, W. Qu, V. Kazan, M. Nazzal, M. Williams, and J. Sferra, "The evaluation of national time trends, quality of care, and factors affecting the use of minimally invasive breast biopsy and open biopsy for diagnosis of breast lesions," Am. J. Surg., vol. 208, no. 3, pp. 382–390, Sep. 2014.

[3] J. M. Eberth et al., "Surgeon Influence on Use of Needle Biopsy in Patients With Breast Cancer: A National Medicare Study," J. Clin. Oncol., vol. 32, no. 21, pp. 2206–2216, Jul. 2014.

[4] T. Wan, J. Cao, J. Chen, and Z. Qin, "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," Neurocomputing, vol. 229, pp. 34–44, Mar. 2017.

[5] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," Comput. Biol. Med., vol. 43, no. 10, pp. 1563–1572, Oct. 2013.

[6] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," Neurocomputing, vol. 197, pp. 221–231, Jul. 2016.

[7] W. Sun, T.-L. (Bill) Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," Comput. Med. Imaging Graph.

[8] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "AggNet: Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1313–1321, May 2016.

[9] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," in 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 2560–2567.

[10] K. Sirinukunwattana, S. E. A. Raza, Y. W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1196–1206, May 2016.

[11] J. Xu et al., "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images," IEEE Trans. Med. Imaging, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[12] F. Jiang et al., "Abdominal adipose tissues extraction using multi-scale deep neural network," Neurocomputing.

[13] M. Havaei et al., "Brain tumor segmentation with Deep Neural Networks," Med. Image Anal., vol. 35, pp. 18–31, Jan. 2017.

[14] X. W. Gao, R. Hui, and Z. Tian, "Classification of CT brain images based on deep learning networks," Comput. Methods Programs Biomed., vol. 138, pp. 49–56, Jan. 2017.

[15] J. Kleesiek et al., "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," NeuroImage, vol. 129, pp. 460–469, Apr. 2016.

[16] K. Kamnitsas et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," Med. Image Anal., vol. 36, pp. 61–78, Feb. 2017.

[17] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in X-ray images," Med. Image Anal., vol. 36, pp. 41–51, Feb. 2017.

[18] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," Opt. - Int. J. Light Electron Opt., vol. 130, pp. 11–18, Feb. 2017.

[19] M. M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," Inf. Sci., vol. 345, pp. 340–354, Jun. 2016.

[20] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675, Mar. 2016.

[21] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu, "ECG signal enhancement based on improved denoising auto-encoder," Eng. Appl. Artif. Intell., vol. 52, pp. 194–202, Jun. 2016.

[22] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," IEEE Trans. Biomed. Eng., vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[24] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning Hierarchical Features for Scene Labeling," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

[25] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," Dec. 2013.

[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," J. Mach. Learn. Res., vol. 15, pp. 1929–1958, 2014.

[27] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.

[28] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, 2013.

[29] Neslihan Bayramoglu, Juho Kannala, and Janne Heikkilä, "Deep Learning for Magnification Independent Breast Cancer Histopathology Image Classification," in 23rd International Conference on Pattern Recognition, 2016.