Predicting Invasive Ductal Carcinoma in breast histology images using Convolutional Neural Network

Conference Paper · July 2019

DOI: 10.1109/NAECON46414.2019.9057822

3 authors, including:

Hesham Alghodhaifi
University of Texas MD Anderson Cancer Center
12 PUBLICATIONS **128** CITATIONS
SEE PROFILE

Mohammed Naif Alghodhaifi
Sana'a University
2 PUBLICATIONS **40** CITATIONS
SEE PROFILE

Predicting Invasive Ductal Carcinoma in breast histology images using Convolutional Neural Network

Hesham Alghodhaifi
ECE Department
University of Michigan-Dearborn
Dearborn, USA
halghodh@umich.edu

Abdulmajeed Alghodhaifi
School of Medicine
Amran University
Amran, Yemen
al_hesham2010@yahoo.com

Mohammed Alghodhaifi
School of Medicine
Sana'a University
Sana'a, Yemen
Mmsnaif99@gmail.com

***Abstract*— Over the past ten years, there has been a rise in using deep learning for medical image analysis such as CNN. Deep learning is used extensively in the field of healthcare to identify patterns, classify and segment tumors and so on. The classification of breast cancer is a well-known problem that attracts the attention of many researchers in the field of healthcare because breast cancer is the second major cause of cancer-related deaths in women. The most common subtype of all breast cancers is the Invasive Ductal Carcinoma (IDC). There are many ways to identify this type of breast cancer such as a biopsy where tissue is removed from patient and studied under microscope. The biopsy is followed by a diagnosis which is based on the qualification of the pathologists, who will look for abnormal cells. The next task for pathologists is to assign an aggressiveness grade to a whole mount sample. To do this, pathologists focus on the region of interest which contain the IDC. Therefore, one of the popular pre-processing steps for automatic aggressiveness grading is to delineate the exact regions of IDC inside of a whole mount slide. In this paper, we have experimentally tested two CNN models using depthwise separable convolution and standard convolution to enhance the accuracy of the convolutional neural network. We tested different types of activation functions such as ReLU, Sigmoid, and Tanh. As well as applying gaussian noise to test the robustness of the two models. The results show convolutional neural networks outperformed the softmax classifier, with standard convolution and ReLU where we achieved ~87.5% classification accuracy, ~93.5% sensitivity, and ~71.5% specificity.**

Keywords—Breast cancer, Classification, IDC, CNN, activation, classifier, prediction

I. INTRODUCTION

Cancer with all types has become one of the major health issues these days. According to statistics by the IARC (International Agency for Research on Cancer) from the WHO (World Health Organization), and GBD (Global Burden of Disease Cancer Collaboration), cancer cases increased by 28% between 2006 and 2016 [1] [2]. In addition, there will be 2.7 million new cancer cases emerging in 2030 [1] [2]. Among the various types of cancer, breast cancer is considered one of the major cause of death in women aged between 20 and 59 years and the second cause of death for women aged more than 59 years [3]. There are more than 1.7 million incident cases, 535,000 deaths, and 14.9 million disability-adjusted life years [2]. The early detection and diagnosis of cancer pathology is essential to reduce its morbidity rates [4].

Breast cancer diagnosis consists of an initial screening by physicians and regular check-ups using mammography or ultrasound imaging [5]. To complete the diagnosis process, the check-ups screening is then followed by breast tissue biopsy if the check-up exam indicates the possibility of malignant tissue growth [8]. The biopsy methods contains collecting samples of cells, looking and fixing them on the microscope, and then marking them [6]. The advantages of taking the breast tissue biopsies are to allow the pathologists to histologically estimate the microscopic structure of the tissue [7].

The histology analysis allows to differentiate between benign and malignant tissue. Moreover, this analysis helps to perform a prognostic evaluation [9]. Benign lesions represent changes in normal structures of breast tissue. One type of the benign lesion is the Invasive ductal carcinoma (IDC).

Invasive ductal carcinoma (IDC) is the most popular subtype of all Breast cancers [10]. About 80% of all breast cancers are invasive ductal carcinomas [11]. This is usually identified by pathologists through visual analysis of tissue slides. Pathologists usually focus on the region that contains the invasive ductal carcinomas in the tissue slides to identify the aggressiveness of this type of cancer [12]. Therefore, the first step in the histopathological analysis is to distinguish between tissue regions corresponding to invasive ductal carcinomas and non – invasive ductal carcinoma. Identifying IDC in the tissue slides allows for more analysis of tumor aggressiveness via the well-known Bloom-Richardson and Nottingham grading schemes [13]. Invasive ductal carcinomas breast cancer detection is a common challenging task because pathologist inspects large benign region to finally define the IDC. Precise identification of IDC region in histology images is an essential step towards the accurate estimation of grading tumor aggressiveness.

Breast cancer detection in histopathology images has been studied by many researchers [14–19]. Most of these methods apply segmentation of histologic elements and followed by features extraction for characterization purposes. As a result of feature extraction, these researchers are able to distinguish between IDC and non-IDC regions. Most of the segmentation and feature extraction approaches involve a large number of different types and complicated handcrafted features to represent the visual content of IDC histopathology images. These approaches usually follow a series of pre-processing steps including detection, segmentation and the result of the final classification is relied on the accuracy of the pre-processing steps.

Recently, deep learning (DL) algorithms are introduced to the public to solve different problems in computer vision and pattern recognition. These algorithms are a result of the gradual development of conventional neural networks [10, 12]. Many of these algorithms are designed based on different concepts from many disciplines such as neuroscience, physics, math, medicine, and engineering. Deep learning approaches typically apply different computer vision

techniques and multiple linear and non-linear transformations of the data, with an ultimate goal of getting a useful representation [20]. Deep learning approaches are so powerful in learning tasks and outperform the traditional methods that include features extraction and machine learning [21–26]. The success of deep learning methods comes from the availability of big data and the powerful computation using graphics processing unit (GPU) machines [27].

Pathology images are captured from taking images of histopathology glass slides by a scanner [28]. Applying deep learning to pathology images provides a fruitful chances to learn hidden patterns in these images. Deep learning methods shows incredible results in the automated segmentation and classification of diseases on histopathology, MRI, CT, X-rays and ultrasound images. One type of deep learning models is convolutional neural networks (CNN), which are a group of multi-layer neural networks designed for use on two-dimensional data, such as images and videos [29]. In recent years, CNN models had been also applied widely for automatic mitosis detection in breast cancer histopathology images [30–32].

The main advantage of deep learning models is that they learn the best representation of data as part of the training process [33]. In our paper, we designed two CNN models that automatically analyze IDC histology images for cancer risk factors, a task that took pathologists hours to complete.

II. METHODOLOGY

A. Dataset Preparation

First, the original dataset is used from [34, 35]. The original images dataset consist of 162 slide images scanned at 40x. These slide images are massive in terms of spatial dimensions. From these slide images, a total of 277524 patches of 50×50 pixels were extracted, including 198738 no IDC examples (no IDC in the patch) and 78786 IDC examples (IDC in the patch). The ground truth of IDC regions over histopathology image slides was acquired via manual delineation of the IDC region by a skilled pathologist.
Fig.1 shows the breast histopathology images dataset examples.
From the above numbers, it is clearly that our data is imbalanced. That means we have more than 2x the number of non-IDC data points than IDC data points. Each image in our dataset has a unique filename structure. This filename structure contains patient ID, x-coordinate of the crop, y-coordinate of the crop, and the class label 0 or 1. Zero class label indicates healthy tissue (no IDC in the image) and the one class label indicates IDC in the image. From fig. 1, healthy tissue (no IDC) examples look brighter in color while the IDC examples tend to be darker.

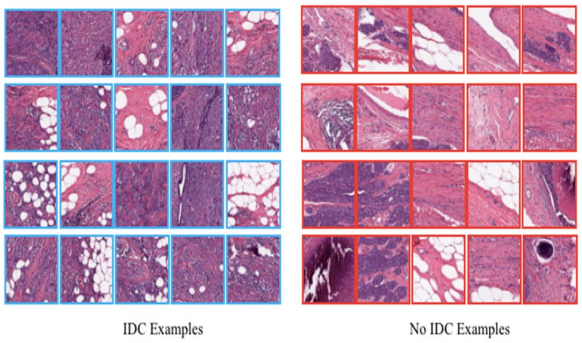

Fig. 1. Example of image patches for positive (blue) and negative (red) categories of the training set

Our breast cancer image dataset consists of 277524 patches. The dataset split into training and testing set, 80% and 20% respectively (fig.2). 10% of the training set is splitted up as a validation set.

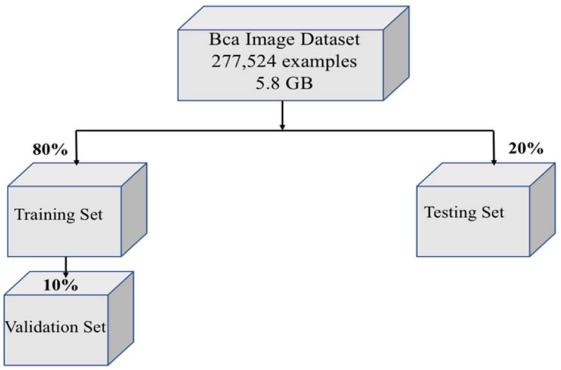

Fig. 2. Building the breast cancer images dataset

B. Convolutional Neural Networks using Depthwise Separable Convolution (IDCDNet)

Convolutional Neural Networks include application of local feature detectors, or filters over the whole image to measure the similarities between individual image patches and signature patterns within the training set [36]. After that, an aggregation or subsampling (pooling) function is applied to reduce feature space dimension. The name of our model is Invasive ductal carcinoma Depthwise Networks (IDCDNet). IDCDNet involves 7-layers CNN. In IDCDNet, 6-layers are used for performing depthwise separable convolution and pooling and the remaining layer is fully-connected. Depthwise separable convolution is used here because it is more efficient, requires less memory and less computation. Depthwise separable convolutions consist of two stages. The first stage is depthwise convolution which comprises the filtering process. In this stage, the convolution is applied to a single input channel at a time and this is different from the standard convolution that applies convolution to all channels. The second stage is the pointwise convolution which involves the combination process. In the combination process, the resulting output channels are mixed together.

IDCDNet uses 3×3 convolution filters, similar to VGGNet, and stacks multiple 3×3 convolution filters on top of each other before applying max-pooling (similar to

VGGNet). Unlike VGGNet, IDCDNet uses depthwise separable 2D convolution. The output feature map from depthwise separable 2D convolution looks like a notable map where learned features was detected. Different activation functions are used to rectify non-linearities in the convolution output. These functions are Rectified Linear Unit (ReLU), sigmoid, and hyperbolic tangent function (tanh). A batch normalization step is applied to each output feature map to reduce overfitting in the input layer by adjusting and scaling the activations [37]. Then, a pooling layer is applied to reduce original large dimension of image representation through a max-pooling strategy. Then, a dropout regularization is used to randomly ignore selected neurons during training. This technique is used to make the network less sensitive to specific weight of neurons [38]. Then, a fully-connected layer is applied in the top of IDCDNet architecture in order to capture all the sophisticated relationships between features. The input of the fully-connected layer is the output of the pooling layer. The output of the fully-connected layer is a feature vector. The final layer is the classification layer. The classification layer is a fully-connected layer with one neuron per each of the two classes activated by softmax classifier. The whole IDCDNet model is trained using Adagrad optimizer to minimize the binary cross-entropy loss function. Gaussian noise of 0.1 and 0.2 are added to every layer to see the accuracy of the IDCDNet while using depthwise separable convolution.

C. Convolutional Neural Networks using Standard Convolution (IDCNet)

In this section, the IDCDNet model is used with some modifications such as using the standard convolution instead of depthwise separable convolution. The modified IDCDNet model uses 7-layers CNN. The new model name is IDCNet. The first 6-layers are used for performing standard convolution and pooling and the last layer is the fully-connected layer. Different activation functions such as Rectified Linear Unit (ReLU), sigmoid, and hyperbolic tangent function (tanh) are used to test the performance of the model. Then, batch normalization, pooling and fully-connected layer with softmax classifier are applied. Adagrad optimizer is used to minimize the binary cross-entropy loss function. Then, gaussian noise of 0.1 and 0.2 are included in each layer to test the behavior of the IDCNet while using standard convolution.

III. EXPERIMENT PERFORMANCE EVALUATION

All patches on the dataset are classified as IDC (1 class label) or Non-IDC (0 class label). Therefore, classification outcomes are evaluated over a validation dataset and over test dataset to compare the prediction of image patches and ground truth. A set of performance measures is obtained by calculating true positives, false positives, false negatives and true negatives. Precision (P1) is calculated to estimate the proportion of IDC detected from total actual IDC patches. Recall (R1) and Sensitivity (S1) are obtained to estimate the proportion of IDC correctly predicted from whole IDC automatically predicted. Specificity (S2) is defined as the proportion of Non-IDC correctly predicted from total actual Non-IDC patches. F-measure (F1) and balanced accuracy are captured to show the trade off when minimizing the false positive and false negative.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The IDCNet model approach for automated IDC tissue region detection was quantitatively evaluated on the testing dataset, yielding a classification performance of 76% (F1-score) and 87.13% (Accuracy). This performance is detailed in the following table including precision, recall or sensitivity and specificity with 81%, 93.44%, and 71.14% respectively.

TABLE I. QUANTITATIVE PERFORMANCE FOR CLASSIFICATION OF IDC AND NO IDC TISSUES. THE PERFORMANCE MEASURES SHOWED ARE PRECISION (P1), RECALL (R1) OR SENSITIVITY (S1), SPECIFICITY (S2), F-MEASURE (F1) AND ACCURACY (ACC).

	P1	S1	S2	F1	Acc
IDCDNet	0.81	0.938	0.6617	0.73	0.8598
IDCNet	**0.81**	**0.9344**	**0.7114**	**0.76**	**0.8713**

Table I indicates that using standard convolution in our model is better than using depthwise separable convolution with an accuracy of 87.13% compared to 85.98%.

The follwing table presents the results of testing both models IDCDNet and IDCNet using different activation functions such as ReLU, sigmoid, and tanh.

TABLE II. QUANTITATIVE PERFORMANCE FOR CLASSIFICATION OF IDC AND NO IDC TISSUES USINH DIFFERENT ACTIVATION FUNCTIONS.

	P1	S1	S2	F1	Acc
IDCDNet (ReLU)	0.81	0.938	0.6617	0.73	0.8598
IDCDNet(sigmoid)	0.53	0.66	0.93	0.67	0.7427
IDCDNet (tanh)	0.60	0.7665	0.8746	0.71	0.7971
IDCNet (ReLU)	**0.81**	**0.9344**	**0.7114**	**0.76**	**0.8713**
IDCNet (sigmoid)	0.78	0.9181	0.7270	0.75	0.8640
IDCNet (tanh)	0.79	0.9335	0.6522	0.72	0.8539

Table II shows that using ReLu activation function with the standared convolution in IDCNet model gives the best performance with an accuracy of 87.13%.

In the following sections, A gaussian noise of 0.1 and 0.2 is applied to each layer in IDCDNet and IDCNet models and then the outcomes are presented to show the performance of the two suggested models.

Table III explains the perfomance measures of the two models.

TABLE III. QUANTITATIVE PERFORMANCE FOR CLASSIFICATION OF IDC AND NO IDC TISSUES WHEN APPLYING 0.1 GAUSSIAN NOISE

	P1	S1	S2	F1	Acc
IDCDNet (ReLU)	0.30	0.080	0.9772	0.45	0.3346
IDCNet (ReLU)	**0.85**	**0.9824**	**0.2468**	**0.38**	**0.7742**

From Table III, the IDCDNet model accuracy is dropped down from 85.98% to 33.46% when applying 0.1 Gaussian noise and the IDCNet accuracy is dropped down from 87.13% to 77.42%. This indicates that the IDCNet model is robust comparing to IDCDNet.

Table IV demonstrates the performance measures of the two models when having a 0.2 gaussian noise.

TABLE IV. QUANTITATIVE PERFORMANCE FOR CLASSIFICATION OF IDC AND NO IDC TISSUES WHEN APPLYING 0.2 GAUSSIAN NOISE

	P1	S1	S2	F1	Acc
IDCDNet (ReLU)	0.24	0.1853	0.640	0.35	0.3140
IDCNet (ReLU)	**0.48**	**0.8582**	**0.3358**	**0.40**	**0.7104**

Table IV reveals the same trend as table III with more decrease in the accuracy in the two models. The values of P1 and F1in all tables are for the IDC detection (1 class). The training loss and accuracy on dataset using IDCNet model are shown in fig. 3.

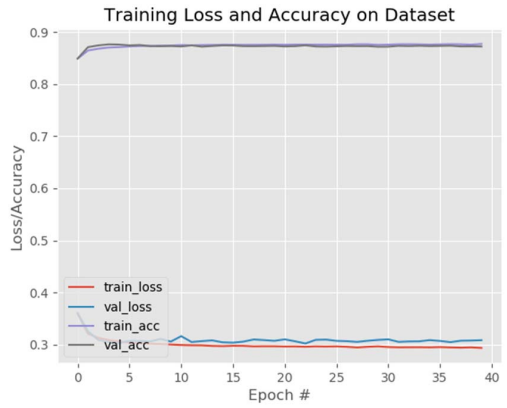

Fig. 3. Training loss and Accuracy on Dataset using IDCNet

With IDCNet, an accuracy of 87.13% (fig.3) is achieved as compared to IDCDNet model accuracy which is 85.98% (fig. 4).

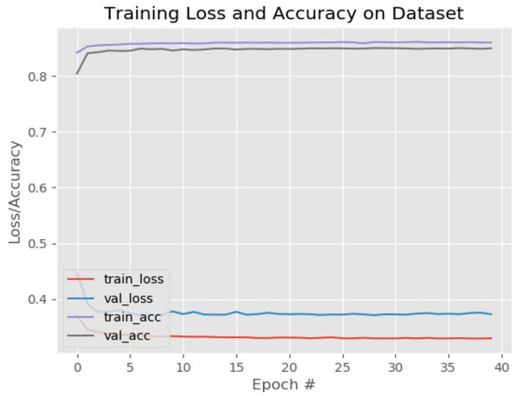

Fig. 4. Training loss and Accuracy on Dataset using IDCDNet (40 Epochs)

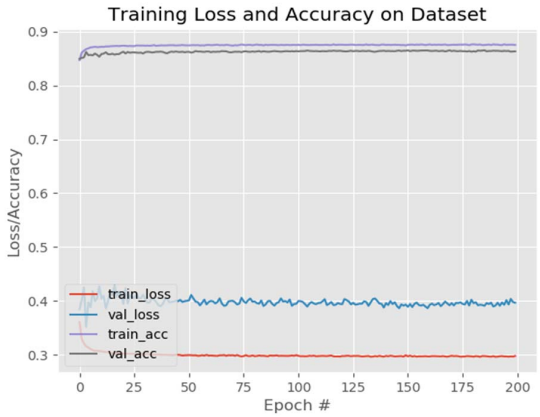

Fig. 5. Training loss and Accuracy on Dataset using IDCDNet (200 Epochs)

V. Conclusion

IDC breast cancer is a threat to women around the globe and is responsible for increasing the female death rate. The improvement of the treatment of IDC cancer can be achieved by proper investigation and diagnosis. Identification of IDC in the earlier stages can save many lives. The status of IDC cancer changes with time, like the appearance, spread, and structural geometry of the cells are changing so rapidly because of the chemical changes inside the cell. These changes in the structure of cells can be captured by analyzing biomedical images of the biopsy such as the histopathology images. However, these images are complicated and require expert pathologists to analyze malignancy. However, deep learning diagnosis techniques can glean a significant amount of details from the images and provide a decision based on the gained information, such as cancer identification, by classifying the images. Most of research works in histopathology tumor detection address this problem by combining different types of handcrafted features and machine learning algorithms. In this paper work, we presented a deep learning approach for automated detection of IDC regions in histopathology images.

In this work, we presented two CNN models to predict IDC in histopathology image. In addition, we showed that two methods of convolutions are able to accurately identify IDC on histpathology images. We found that using the standard convolution gave the best performance with an accuracy of 87.13%. We tested the performance of the two models in detection IDC while using different activation functions such as ReLU, sigmoid, and ranh. We found that IDCNet and IDCDNet models perform so well when using ReLU comapring to other activation functions.

One of the most interesting findings was that adding gaussian noise to each layer of the IDCNet and IDCDNet models was more influential on the robustness of the IDCDNet model comparing to IDCNet. Future work will explore the effects of IDCNet and IDCDNet models with advanced architectures and validation on larger dataset. However, the most interesting future directions are designing a rubost CNN models and finally designing an automatic IDC aggressevness grading system.

References

[1] Bradley, Andrew P. "The use of the area under the ROC curve in the evaluation of machine learning algorithms." Pattern recognition 30.7 (1997): 1145-1159.

[2] Fitzmaurice, Christina, et al. "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study." JAMA oncology 4.11 (2018): 1553-1568.

[3] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA: A Cancer Journal for Clinicians. 2016;66(1):7–30.

[4] Smith Ra, Cokkinides V, Eyre HJ. American Cancer Society Guidelines for the Early Detection of Cancer, 2004. CA: A Cancer Journal for Clinicians. 2004;54(1):41–52.

[5] Stenkvist, Björn, Sighild Westman-Naeser, Jan Holmquist, Bo Nordin, Ewert Bengtsson, Jan Vegelius, Olle Eriksson, and Cecil H. Fox. "Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations." Cancer research 38, no. 12 (1978): 4688-4697.

[6] Veta, Mitko, et al. "Breast cancer histopathology image analysis: A review." IEEE Transactions on Biomedical Engineering 61.5 (2014): 1400-1411.

[7] Spanhol, Fabio A., et al. "A dataset for breast cancer histopathological image classification." IEEE Transactions on Biomedical Engineering 63.7 (2015): 1455-1462.

[8] NationalBreastCancerFoundation. Breast Cancer Diagnosis; 2015. Available from: http://www.nationalbreastcancer.org/breast-cancer-diagnosis.

[9] 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;19(5):403–410. 10.1111/j.1365-2559.1991.tb00229.

[10] DeSantis, C., Siegel, R., Bandi, P., and Jemal, A., "Breast cancer statistics, 2011," CA: A Cancer Journal for Clinicians 61(6), 408–418 (2011).

[11] "Invasive Ductal Carcinoma: Diagnosis, Treatment, and More." Breastcancer.org, 9 Mar. 2019, www.breastcancer.org/symptoms/types/idc.

[12] Elston, C. and Ellis, I., "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," Histopath. 19(5), 403–410 (1991).

[13] Genestie, C., Zafrani, B., Asselain, B., Fourquet, A., Rozan, S., Validire, P., Vincent-Salomon, A., and Sastre-Garau, X., "Comparison of the prognostic value of scarff-bloom-richardson and nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems.," Anticancer Res 18(1B), 571–6 (1998).

[14] Petushi, S., Garcia, F. U., Haber, M. M., Katsinis, C., and Tozeren, A., "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer.," BMC medical imaging 6, 14 (Jan. 2006).

[15] Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., and Tomaszewski, J., "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in [2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro], 284–287, IEEE (May 2008).

[16] Doyle, S., Agner, S., Madabhushi, A., Feldman, M., and Tomaszewski, J., "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in [2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro], 496–499, IEEE (May 2008).

[17] Dundar, M. M., Badve, S., Bilgin, G., Raykar, V., Jain, R., Sertel, O., and Gurcan, M. N., "Computerized classification of intraductal breast lesions using histopathological images," IEEE Transactions on Biomedical Engineering 58, 1977–1984 (July 2011).

[18] Niwas, S. I., Palanisamy, P., Zhang, W., Mat Isa, N. A., and Chibbar, R., "Log-gabor wavelets based breast carcinoma classification using least square support vector machine," in [2011 IEEE International Conference on Imaging Systems and Techniques], 219–223, IEEE (May 2011).

[19] Basavanhally, A., Ganesan, S., Feldman, M. D., Shih, N., Mies, C., Tomaszewski, J. E., and Madabhushi, A., "Multi-field-of-view framework for distinguishing tumor grade in er+ breast cancer from entire histopathology slides," IEEE transactions on biomedical engineering 60, 2089–2099 (Aug 2013).

[20] Bengio, Y., Courville, A., and Vincent, P., "Representation learning: A review and new perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8), 1798–1828 (2013).

[21] Weston, J., Bengio, S., and Usunier, N., "Large scale image annotation: Learning to rank with joint wordimage embeddings," Machine Learning 81, 21–35 (Oct. 2010).

[22] Seide, F., Li, G., and Yu, D., "Conversational speech transcription using context-dependent deep neural networks," in [Proc Interspeech], 437–440 (2011).

[23] Glorot, X., Bordes, A., and Bengio, Y., "Domain adaptation for large-scale sentiment classification: A deep learning approach," in [Proceedings of the Twenty-eight International Conference on Machine Learning (ICML'11)], 27, 97–110 (June 2011).

[24] Boulanger-Lewandowski, N., Bengio, Y., and Vincent, P., "Modeling temporal dependencies in highdimensional sequences: Application to polyphonic music generation and transcription," in [Proceedings of the Twenty-nine International Conference on Machine Learning (ICML'12)], ACM (2012).

[25] Ciresan, D. C., Meier, U., and Schmidhuber, J., "Multi-column deep neural networks for image classification," in [Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)], CVPR '12, 3642–3649, IEEE Computer Society, Washington, DC, USA (2012).

[26] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [Advances in Neural Information Processing Systems 25], 1106–1114 (2012).

[27] Hey, T. and Trefethen, A., [The Data Deluge: An e-Science Perspective], 809–824, John Wiley & Sons, Ltd (2003).

[28] Ghaznavi, F., Evans, A., Madabhushi, A., and Feldman, M., "Digital imaging in pathology: Whole-slide imaging and beyond," Annual Review of Pathology: Mechanisms of Disease 8(1), 331–359 (2013). PMID: 23157334.

[29] Arel, I., Rose, D. C., and Karnowski, T. P., "Research frontier: Deep machine learning–a new frontier in artificial intelligence research," Comp. Intell. Mag. 5, 13–18 (Nov. 2010).

[30] Malon, C., Miller, M., Burger, H. C., Cosatto, E., and Graf, H. P., "Identifying histological elements with convolutional neural networks," in [Proceedings of the 5th International Conference on Soft Computing As Transdisciplinary Science and Technology], CSTST '08, 450–456, ACM, New York, NY, USA (2008).

[31] Malon, C. and Cosatto, E., "Classification of mitotic figures with convolutional neural networks and seeded blob features," Journal of Pathology Informatics 4(1), 9 (2013).

[32] Ciresan, D., Giusti, A., Gambardella, L., and Schmidhuber, J., "Mitosis detection in breast cancer histology images with deep neural networks," in [Medical Image Computing and Computer-Assisted Intervention MICCAI 2013], Lecture Notes in Computer Science 8150, 411–418, Springer Berlin Heidelberg (2013).

[33] Cruz-Roa, Angel, et al. "Automatic Detection of Invasive Ductal Carcinoma in Whole Slide Images with Convolutional Neural Networks." Medical Imaging 2014: Digital Pathology, 2014, doi:10.1117/12.2043872.

[34] Janowczyk, Andrew, and Anant Madabhushi. "Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial with Selected Use Cases." Journal of Pathology Informatics, vol. 7, no. 1, 2016, p. 29., doi:10.4103/2153-3539.186902.

[35] "Breast Cancer Classification with Keras and Deep Learning." PyImageSearch, 23 Apr. 2019, www.pyimagesearch.com/2019/02/18/breast-cancer-classification-with-keras-and-deep-learning/.

[36] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-based learning applied to document recognition," Proceedings of the IEEE 86(11), 2278–2324 (1998).

[37] Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y., "What is the best multi-stage architecture for object recognition?," in [Computer Vision, 2009 IEEE 12th International Conference on], 2146–2153 (Sept 2009).

[38] "Dropout Regularization in Deep Learning Models With Keras." Machine Learning Mastery, 2 May 2019, machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/.