Advances and Applications in Mathematical Sciences
Volume 20, Issue 3, January 2021, Pages 409-426

A CONVOLUTIONAL NEURAL NETWORK TO PREDICT BREAST CANCER FOR HISTOPATHOLOGICAL IMAGES

SONIKA NAGAR, MRADUL KUMAR JAIN, NIRVIKAR and AMIT KUMAR AGARWAL

CSE Department
ABES Engineering
College, Ghaziabad, India.
E-mail: nagar.sonika1994@gmail.com

CSE Department
ABES Engineering
College, Ghaziabad, India
E-mail: mradul.jain@abes.ac.in

CSE Department
College of Engineering
Roorkee, Roorkee, India
E0mail: nirvikarlohan@yahoo.com

CSE Department
ABES Engineering
College, Ghaziabad, India.
E-mail: amitkagarwal@abes.ac.in

Abstract

Recently, CNNs have become a preferred deep learning artificial neural network of choice for computer assisted medical image analysis. These models are structured as a series of multiple hierarchical processing layers that can automatically learn feature representations from raw images. Detection and classification of cancer in histopathological images is one of the biggest challenges for oncologists. CNNs have in the past not been in common use, especially in medical imaging field, because of issues such as insufficient image datasets. The revolution in CNN models has been attributed to powerful parallel processing hardware architectures,

2010 Mathematics Subject Classification: 46B80, 92B20, 94A08.
Keywords: Biomedical Imaging, Convolutional Neural Network, Deep Learning, Histopathological Images, Breast Cancer.
Received May 20, 2020; Accepted July 31, 2020

increasing number of image datasets and improved training strategies. Utilizing these deep learning techniques is enabling medical experts such as pathologists to utilize artificial intelligence to transform the world of medicine for more accurate and faster diagnoses. We propose in the paper a convolutional neural network model to distinguish breast cancer histopathological images into two sections i.e. malignant and benign with the training accuracy of 99.25% and 94% testing accuracy.

I. Introduction

Breast cancer is the one of most common cancer in the world kind that is more common in women than in men. As stated by, World Health Organization (WHO), about 15 percent of the cancer incident cases are usually related to breast cancer [1]. Early detection can greatly help to reduce the mortality and morbidity rates resulting not only from these cancer type but also cancer affecting other body parts of the human being.

The most common symptoms indicated by the patients affected by Breast Cancer includes change in size, appearance of Breast and change in skin color over Breast. In lifetime on an average 1 in 8 women is diagnosed to breast cancer.

For early discovery of breast cancer medical imaging techniques are used such as mammography, ultrasound, MRI and histopathological imaging [2]. Histopathological imaging produces some of the most accurate and more reliable results in the detection and staging of breast cancer. This technique is done by obtaining breast tissue biopsies from the patient. The tissue is then stained and placed under a microscope that allows the pathologists to histologically assess the microscopic structure and elements of the tissue [3]. The staining process can be done using hematoxylin and eosin (H&E). H&E staining provides permanency of the specimen and helps the pathologists to differentiate the tissue components [4]. Stained tissue from the glass slides can then be digitized by using high resolution image scanners into whole slide images [5-6]. The histopathological images will therefore be available for long-term storage and further analysis by a computer-aided diagnosis system. Some of the manual techniques such as histological diagnosis, tumor size and axillary lymph node metasis sometimes fail to classify accurately the observed breast tumors [7-8].

The diagnoses of the histopathological images using computer-aided tools

require utilization of machine learning techniques. In the past, classification of such images would require feature engineering techniques to extract features that were supplied to a classical machine learning classifier.

However, advances in machine learning such as those observed in deep learning are providing computers with the ability to automatically extract these features. Consequently, the burden of feature engineering especially in medical image analysis is being taken away from humans by computers for more accurate results [9].

Deep learning models have the ability to extract the features automatically from high dimensional natural raw images for a suitable internal representation. Multiple levels of representation in deep neural networks permit representational learning, allowing hierarchical feature representations from non-linear modules that modify the representation at one level which starting with the raw input into a representation at a higher more abstract level [10], [11]. The automatically extracted features can then be used by a classifier to recognize images that are supplied to the network. Convolution Neural Network is the most widely used deep learning model to extract feature and classify nowadays. In a feature extraction task, the CNN is trained in a fully supervised setting before using the trained network parameters to automatically extract features from images. The extracted features can then be supplied to a different classifier such as logistic regression (LR), k-NN and SVM for a classification task [12], [13].

In this paper, CNN model is offered for breast histopathology image classification. This CNN model is trained using histopathological images of breast cancer to classify these images into two most common types of breast cancer i.e. benign and malignant and then tested on the reserved set of histopathological images for testing.

A. Types of Breast Cancer Tumor

Breast cancer tumors are mainly classified into two broad scenarios.

(i) Benign

Benign cases are non-life-threatening and are considered as noncancerous. But in some rare cases its status may possibly turn into cancerous. Benign tumors are normally segregated from other cells through

an immune system called “sac” and they can be very skillfully taken out from body.

(ii) Malignant

This type of cancer begin from abnormal cell growth and it may be speedily spread or invade its surrounding tissues and rapidly spread around the body through blood or lymph system which can be life threatening as they may recur after been removed.

B. Causes associated with Breast cancer

1. Age: Usually, women age more than 60 years are diagnosed with breast cancer. Around 10-15% chances of breast cancer arise in women age is around or less than 45 years. Whereas in women age around or less than 45 years the chances of breast cancer is around 10-15%. While 81% cases of breast cancer are detected surrounding by women age 50 years [14].

2. Family History: The risk in breast cancer patients can increase with the analysis of past family records on the basis of expected and observed cases of breast cancer in the family [15].

3. Menstrual History: The chances of breast cancer are higher in the women who have menarche or the onset menstruation at the age of more than 12 years [16].

4. Genetics: Most breast cancer appears physical, around 5% are considered as consequences of inherited breast cancer susceptibly [17].

5. Pregnancy and breastfeeding History: Women after the 30 years age who deliver their first child or doesn’t have a full term pregnancy have higher risk of breast cancer than to the women who had their deliveries at the age of less than 30 years [18].

It is essential to breastfeed at least 6 months per child to reduce or minimize the risk of breast cancer. [19].

6. Smoking and drinking alcohol: women smoking tobacco may increase chances of danger of developing breast cancer up to 35% [20].

Women who drink alcohol beverages have 15% higher chances of breast cancer [21].

Other factors related with the causes of breast cancer can be female eating unhealthy food, overweight or obesity, lack of exercises, Radiation exposure, etc.

II. Recent Works

The automatic diagnosis of cancer has been considered as a research topic for more than four decades. The recent researches show that the algorithms based on CNN achieves improved results, which outperforms the finest traditional machine learning methods.

Arevalo et al. [22] proposed a CNN for classification of benign and malignant lesions of breast cancer on dataset BCDR-F03 containing 736 film images of 344 patients. These images were manually segmented into 426 benign and 310 malignant lesions. At first, the image enhancement is done and then these enhanced images are fed to CNN to classify lesions into benign and malignant and reported 82.6% AUC.

Kooi et al. [23] proposed a ConvNet model for detection of breast cancer from mammographic lesions on nearly 45000 images to achieve an accuracy of 85%.

Spanhol et al. [24] trains the existing CNN on the high-resolution images based on the patches' extraction, and when the accuracy is compared with the traditional machine learning methods, show some improvement in accuracy Brook et al. [25] proposed an approach for diagnosis of breast cancer using microscopic biopsy images using machine learning. Generic features and statistical learning algorithms were used to extract features from the images. The extracted features were used to train a SVM for a 3 class classification task. The histopathological images were classified as normal, in situ or invasive carcinoma breast cancer types. The authors recorded an average error rate between 6.6% with 0.8% standard error of the mean.

Zang et al. [26] presented a 3 class classification of the breast cancer histopathological image diagnosis. The author combined a Local Binary Pattern (LBP) feature description with Curvelet Transform (CT) for texture analysis in the images. Training and testing were implemented in 2 ensembles. In the first ensemble, the extracted features were supplied to an

SVM. In the second ensemble, a Multi Layer Perceptron (MLP) was used to focus on the samples that have been rejected from the first ensemble. The author was able to achieved high accuracies of up to 97 percent accuracy with a rejection rate of 0.8 percent.

Zhi, W., et al. [27] proposed a CNN architecture by implementing the transfer learning technique for classifying breast histopathology images. This greatly reduced model creation time instead of building one from scratch. The authors were able to compare their performance using other off-the-shelf deep learning models such as VGGNet.

Araújo, T., et al. [28] proposed a hybrid model for the classification of histological images of breast cancer. The images are preprocessed, using optical density colors conversion, before feeding them to the machine learning models. The authors create 70,000 patches from 250 images with each patch labeled with the same class label. A CNN model is used to features extraction from breast cancer images. The extracted features are trained using an SVM for classification and accuracies of up to 77.8% is achieved.

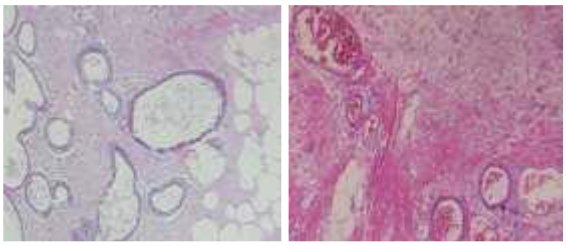
Golatkar et al. [29] proposed a deep learning model to classify the breast cancer histopathological images from the ICIAR BACH image dataset efficiently. Their model is built through transfer learning techniques where the Inception-v3 CNN model is fine tuned for the classification task. The authors record average accuracies of up to 85% in a four class classification (normal tissue, benign lesion, in situ carcinoma and invasive carcinoma).

III. Proposed Methodolgy

A. Dataset Description

Our dataset contains microscopy biopsy images from two types of tumor samples, benign and malignant from BreaKHis dataset. [30]. Our dataset contains 500 images that is 250 pictures of benign and 250 of malignant, each of them of size 460x700 pixels, 40X zoom and with 3 colors or channels.

To us, who have no experience in histology, these two types of image samples look pretty similar.



(A) (B)

Figure 1. Examples of histopathological images from the Brea KH is dataset: (A) benign, (B) malignant.

B. Data Preprocessing

1. Loading the images

The first we have to do is converting the images to a format that will be understood by the CNN. Neural networks speak the language of tensors, so we will convert our dataset to one. For this, we have used Numpy and we have represented real numbers with single precision (i.e. 32 bits) to keep memory usage low.

Our dataset contains 250 pictures of benign and 250 of malignant, each of them of size 460x700 pixels and with 3 colors or channels. Then we have labeled all benign samples as 0 and all malignant as 1.

2. Scaling the dataset

In order to make the most of our data, it is recommended to scale it so that all features (i.e. each pixel in each channel) has roughly the same mean and standard deviation across all samples. This way, it will be easier for the CNN to identify which features are most informative. We have achieved a mean of 0 and standard deviation of 1 in all features.

3. Splitting the dataset into training validating and testing set

We will follow the standard choice of 60% of training and 20% for validation and test each. Note that, in order to avoid unwanted biases, we have to randomly permute the samples before splitting. That way we ensure that the training, validation and test sets are homogeneous.

C. Convolutional Neural Network

A CNN (or ConvNets) is a deep multilayer feed-forward neural network machine learning algorithm that resembles the functioning of a human being's visual cortex. The CNN's architecture is typically made up of one input layer, a feature extraction layer, a classification layer and an output layer. The feature extraction layer is made up of convolution and pooling layers while the classification layer is typically build using an MLP. The convolution, pooling and fully connected (in the MLP) layers found between the input and output layers are non-linear hidden layers. The deeper the layers, the more the complexity of the learned image properties increase [31]. Figure 2. shows the CNN structure used in breast cancer classification.

Deep learning architectures such as CNN have the ability of sparse representations to uncover semantic information from naturally very high dimensional raw image data. CNN combine three architectural ideas that include spatial sub-sampling, shared weights and local receptive fields. The output of each layer, in the feature extraction module, is a feature map. The feature map is created by convolving the filter matrix by a specified number of steps for the whole input image. A convolutional operation can be represented by equation (1).

$$(I * K)_{xy} = \sum_{i=1}^{h} \sum_{j=1}^{g} K_{ij} \cdot I_{x+i-1,\, y+j-1}, \tag{1}$$

Where I, is two-dimensional input image K, is a set of filters in the convolutional layer, h is height of image, g is width of the image. A convolutional operation of an image is represented by $I * K$. If a feature is found, the responsible unit or units generate large activations, whose values can be used later by the classification module [32], [33]. The values in the feature maps are the subjected to an element-wise transfer function, such as the Rectifier Linear Unit (ReLU), hyperbolic tangents and the sigmoid functions, to improve the non-linearity of the decision function.

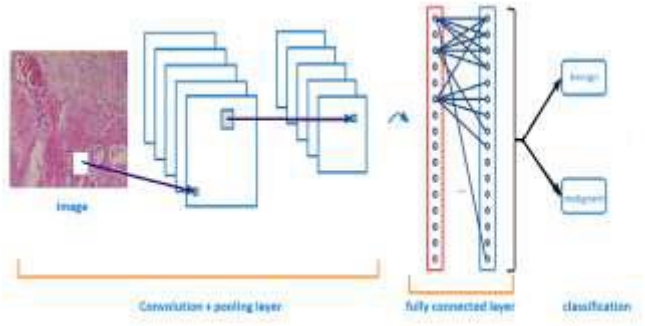


Figure 2. Structure of CNN for breast cancer classification.

For the architecture, our CNN will consist of: 3 convolutional layers (convolution + pooling), 1flatten layer, 2 dense layers (fully connected layer) and the output layer.

The CNN used in this approach involves 3 convolutional layer and 3 pooling layers. The input images are of size 460 × 700 resolution. In layer 1 convolutional2D is of 32- 2 ×2 a filter, in layer 2 convolutional2D is of size 64- 2 × 2 filters. In layer 3 convolutional2D is of size 16- 2×2 filters. After flattening images are sent to dense layers. The architecture of the proposed CNN model with detailed layers along with the filters as a flow chart is shown in Figure 3.

All activations are Rectified Linear Units (ReLU) that is known to lead to fast optimization by back propagation. For the output layer, however, we will use the logistic or sigmoid function. This function ranges in between 0 and 1, so it can be interpreted as probabilities during binary classification.

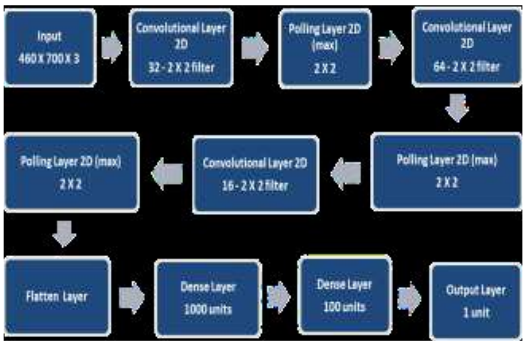


Figure 3. Proposed Convolutional Neural Network Architecture.

Convolution Neural Networks are the suited type of neural networks especially to deal with images because they apply subsequent rounds of convolution and pooling operations that act as Gabor filters to extract progressively more complex features. Pooling operations are loosely inspired on how the human brain processes signals from the eye.

IV. Experimentation and Results

A. Training the model

As the activations for all the layers are ReLU while for the output layer the logistic or sigmoid function is used which ranges in between 0 and 1, so it can be interpreted as probabilities during binary classification.

Finally, we have attempted to prevent overfitting by applying dropout regularization, where a fraction of the units are dropped randomly. This type of regularization has become widespread in deep learning because it is very convenient and also leads to robust models. An intuitive way of understanding such robustness is that the network learns that it cannot completely rely on any individual neuron, but must make predictions which are supported on many neurons collectively.

The choice of architecture is actually a hyperparameter that must be tuned.

(a) Tune the hyperparameters

A machine learning model always has some parameters which cannot be optimized by the optimization algorithm. Instead, they have to be chosen manually. Examples include the learning rate, the architecture, the optimizer itself (whether it is gradient descent, Adam, Adadelta, etc.) or the training time in epochs (an epoch is a single pass through the training set).

The optimal values of these hyperparameters are found by sampling: for each combination of hyperparameters to sample, first we have trained on the training set, and then we have evaluated the accuracy obtained on the validation set (in fact, tuning the hyperparameters is the main purpose of having a validation set). In this case, I've already found that the network above and 30 epochs are a good architecture and training time respectively. We still had to find the appropriate dropout rate.

1. A large dropout rate leads to underfitting

Dropout is a regularization technique that works in the following way: for each forward propagation, randomly select a fraction of neurons in each layer and render them inactive, as though they didn't exist. This way, the network learns to make predictions that are spread out throughout all neurons, rather than very reliant on a small number of highly important neurons. The fraction of neurons to drop is called the dropout rate.

We have used the optimization algorithm Adam (ADAptive Moment estimation), a sophisticated type of gradient descent that has momentum. This means that, in order to estimate how much the parameters will vary in each optimization iteration, both the current gradient and previous gradients are considered. If we make an analogy with physics, a ball descending downhill has a velocity that depends not only on the current slope, but also on the momentum it has accumulated as a result of previous slopes. Hence the "descent with momentum" name.

And first guess at the dropout rate was 0.75; we attained 84.33% as training accuracy and 84% testing accuracy. As we can see, a relatively low accuracy is achieved during both training and validation in both the guesses. This means that the classifier was underfitting, i.e. it doesn't have enough flexibility do fit the data correctly. In machine learning talk, this flexibility is called capacity. In order to increase the classifier's capacity, we need to decrease the regularization by decreasing the dropout rate.

2. A small dropout rate leads to overfitting

Our next guess at the dropout rate was 0.05. We attained 99.67% as training accuracy and 91% testing accuracy. And we had seen there was a huge difference between the train accuracy and the validation accuracy. That means that the classifier was overfitting, i.e. given the permissive regularization it is very flexible and can memorize the training labels instead of learning the overall trends in the dataset. This is undesirable because if it doesn't learn the overall trends, it will do poorly on unseen data, i.e. it will not generalize.

3. A medium dropout rate leads to high validation accuracy

We then tried with a rate between 0.75 and 0.05. At a dropout rate of 0.1, the classifier doesn't overfit (as the difference between the training and validation accuracies is smaller) and doesn't underfit (the training accuracy is

high i.e. 99.33%. We obtain a validation accuracy of 95%, which is much better than previous guesses. We have taken many dropout rate and the accuracy increases as the dropout rate decreases but start overfitting. Figure 6. shows the graph between training accuracy and validation accuracy at different dropout rate taken in training the CNN model.

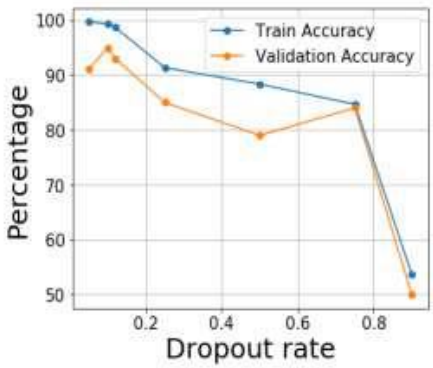


Figure 4. Training accuracy v/s Validation accuracy for different dropout rates.

(b) Retrain on training and validation set

Now that we have known an optimal choice of hyperparameters, we want to train on as many samples as possible before evaluating on the test set. Therefore, we have combined the training and validation sets to train the final model.

We have obtained a model that is able to benign and malignant with an accuracy of 93%. The 1% difference between the test accuracy and the validation accuracy is probably due to the fact that we have overfit slightly to the validation set. Figure 5. displays graph of the accuracy with cross-entropy loss for dropout rate 0.1 after retraining the model.

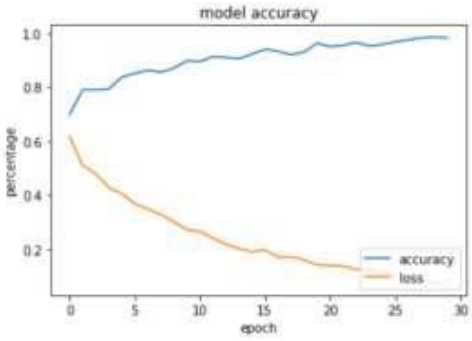


Figure 5. Training accuracy v/s cross-entropy loss after retraining.

B. Testing

After retraining our model on combined training and validating sets, the dataset reserved for testing is used to test the CNN model. Our model shows best performance in classification of images at dropout rate 0.1. It achieves, respectively, 93% of testing accuracy and 99.25% training accuracy.

C. Results

Evaluating performance of models is one of the most essential and final step in building a model. In this approach Precision, Recall, F1-score, support, Accuracy score are considered as Performance metrics for evaluating model. For performance evaluation the dataset is split into training and testing initially, after that values are predicted using test set on trained model.

The accuracy score of model is calculated by considering actual result and predicted value. Mathematical form for calculating accuracy score is as mentioned below.

$$Accuracy = \frac{TP + FP}{Total}. \tag{2}$$

Precision is percentage of relevant results mathematical form is as mentioned below. It is fraction of true positives to Actual Results.

$$Precision = \frac{TP}{TP + FP}. \tag{3}$$

Recall is percentage of total relevant results correctly classified by trained model. It is fraction of True positives to Predicted Results. Mathematical form for calculating Recall is as shown below.

$$\text{Re}\,call = \frac{TP}{TP + FN}. \tag{4}$$

Harmonic mean of Recall and Precision is F1-score. The Mathematical form for calculating F1-score is as shown below.

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{5}$$

The following table represents Accuracy, Precision, Recall, F1-score, Cohens Kappa and ROC AUC on trained and testing set.

Table 1. Performance perimeters for CNN model.

Parameters		Percentage/score
	Training set	Testing set
Accuracy	0.992	0.930
Precision	1.000	0.940
Recall	0.984	0.921
F1 score	0.992	0.930
Cohens Kappa	0.984	0.860
ROC AUC	0.998	0.985

The table below represents Performance metrics Precision, Recall, F1-score and Support values on trained and testing dataset.

Table.2. Classification report of CNN model.

	Training Set				Testing Set			
	P	R	F1	S	P	P	F1	S
Benig n/0	0.99	1.00	0.99	201	0.92	0.94	0.93	49
Malig nant/1	1.00	0.98	0.99	199	0.94	0.92	0.93	51

Micro avg	0.99	0.99	0.99	400	0.93	0.93	0.93	100
macro avg	0.99	0.99	0.99	400	0.93	0.93	0.93	100
weighted avg	0.99	0.99	0.99	400	0.93	0.93	0.93	100

*P=Precision, R=Recall, F1=F1-score, S=Support

We have plot the confusion matrix for this binary classification for both the training set and the test set for the proposed CNN model as shown in Figure 6.

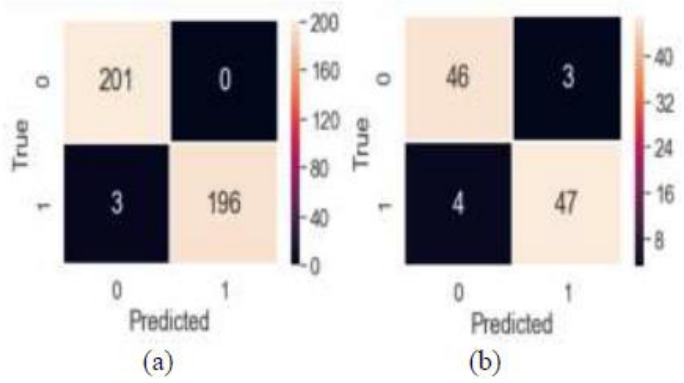


Figure 6. Confusion matrix for (a) training and (b) testing.

V. Conclusion

We use Artificial Neural Networks which are nothing but an imitation of how human brains actually work. The knowledge that models build using these algorithms is later tested on unlabeled observations.

Talking about performance, it has been noticed that convolution neural network models using machine learning tends to give better accuracy than humans in most cases of supervised learning. In this actual task, you see how our algorithm exceptionally outperforms humans. We have obtained a model that is able to classify benign and malignant cases of breast cancer with the training accuracy of 99.25% and the testing accuracy of 93%. Therefore CNN model can be used for diagnosing breast cancer for better performance and less misclassifications.

References

[1] WHO. Cancer: Breast Cancer, 7/12/2017, 2017; http://www.who.int/cancer/prevention/diagnosis-screening/breast-canc er/en/

[2] A. C. Society, Breast Cancer Facts and Figures 2017-2018, 7/12/2017, 2017; https://www.cancer.org/content/dam/cancer-org/research/cancer-factsand-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-fi gures-2017-2018.pdf.

[3] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia and A. Campilho, Classification of breast cancer histology images using Convolutional Neural Networks, PLOS ONE 12(6) (2017), e0177544.

[4] R. Dougall Lillie, Histopathologic technic and practical histochemistry/R. D. Lillie, Harold M. Fullmer, 2017.

[5] D. Hartman, L. Pantanowitz, J. McHugh, A. Piccoli, M. OLeary and G. Lauro, Enterprise Implementation of Digital Pathology: Feasibility, Challenges and Opportunities, Journal of Digital Imaging 30(5) (2017), 555-560.

[6] S. Al-Janabi, A. Huisman, S. M. Willems, and P. J. Van Diest, Digital slide images for primary diagnostics in breast pathology: a feasibility study, Human Pathology 43(12) (2012), 2318-2325, 2012/12/01/, 2012.

[7] L. J. Van't Veer, H. Dai, M. J. Van De Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. Van Der Kooy, M. J. Marton and A. T. Witteveen, Gene expression profiling predicts clinical outcome of breast cancer, nature 415(6871) (2002), 530-536.

[8] W. L. McGuire, Breast cancer prognostic factors: evaluation guidelines, Oxford University Press, 1991.

[9] D. Shen, G. Wu and H.-I. Suk, Deep Learning in Medical Image Analysis, Annual Review of Biomedical Engineering 19(1) (2017), 221-248.

[10] Y. LeCun, Y. Bengio and G. Hinton, Deep Learning, 2015.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research 15 (2014), 1929-1958.

[12] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng and T. Darrell, Decaf: A deep convolutional activation feature for generic visual recognition. pp. 647-655.

[13] A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks 1097-1105.

[14] C. E. DeSantis, J. A. Ma, L. A. Newman and A. Jemal, Breast cancer statistics, 2017, racial disparity in mortality by state, CA: a cancer journal of clinicians 67(6) (2017), 439-448.

[15] H. R. Brewer, M. E. Jones, M. J. Schoemaker, A. Ashworth and A. J. Swerdlow, Family history and risk of breast cancer: an analysis accounting for family structure, Breast cancer research and treatment 165(1) (2017), 193-200.

[16] M. P. Antony, B. Surakutty, T. A. Vasu and M. Chisthi, Risk factors for breast cancer among Indian women: A case control study. Nigerian Journal of clinical practices 21(4),

2018.

[17] K. A. Blood, M. McCullum, C. Wilson and R. Cheifetz, Hereditary breast cancer in British Columbia: Outcomes from BC cancers High-Risk clinic. British Columbia Medical Journal, 60(1) (2018).

[18] V. Chaurasia, S. Pal and B. B. Tiwari, Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms and Computational Techniques 12(2) (2018), 119-126.

[19] K. D. Shield, L. Dossus, A. Fournier, C. M. Micallef, S. Rinaldi, A. Rogel and I. Seorjomataram, The impact of historical breastfeeding practices on the incidence of cancer in France in 2015. Cancer causes and control 29(3) (2018), 325-332.

[20] M. L. Baglia, L. S. Cook, M. T. Tang, C. Wiggins, D. Hills, P. Porter and C. I. Li, Alcohol, smoking and risk of Her2-Overexpressing and Triple Negative Breast Cancer Relative to Estrogen Receptor-Positive Breast Cancer, International Journal of Cancer, 2018.

[21] J. Larson, P. Wallace, F. Sim, J. Chick, S. Jarvis, I. Lidington and Owens, Accuracy of alcohol and breast cancer risk information on Drinkaware's website, Drug and alcohol review 37(3) (2018), 304-306.

[22] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, M. A. G. Lopez, Representation learning for mammography mass lesion classification with convolutional neural networks, Comput. Methods Programs Biomed. 127 (2016), 248-257.

[23] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. den Heeten and N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, Med. Image Anal. 35 (2017), 303-312.

[24] F. A. Spanhol, L. S. Oliveira, C. Petitjean, et al. Breast cancer Histopathological image classification using convolutional neural networks. In: International Joint Conference on Neural Networks (IJCNN), IEEE (2016), 2560-2567.

[25] A. Brook, R. El-Yaniv, E. Isler, R. Kimmel, R. Meir and D. Peleg, Breast Cancer Diagnosis From Biopsy Images Using Generic Features and SVMs, 2007.

[26] B. Zhang, Breast cancer diagnosis from biopsy images by serial fusion of Random Subspace ensembles, pp. 180-186.

[27] W. Zhi, H. W. F. Yueng, Z. Chen, S. M. Zandavi, Z. Lu and Y. Y. Chung, Using Transfer Learning with Convolutional Neural Networks to Diagnose Breast Cancer from Histopathological Images, pp. 669-676.

[28] A. Golatkar, D. Anand and A. Sethi, Classification of Breast Cancer Histology Using Deep Learning, Image Analysis and Recognition (2016), 837-844.

[29] C. f. O. M. I. Computing, ICIAR 2018 Grand Challenge on Breast Cancer Histology Images, Consortium for Open Medical Image Computing, 2017.

[30] https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-da tabase-breakhis/

[31] W. Zhao and S. Du, Spectral-spatial feature extraction for hyperspectral image

classification: A dimension reduction and deep learning approach, IEEE Transactions on Geoscience and Remote Sensing 54(8) (2016), 4544-4554.

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

[33] Y. LeCun and Y. Bengio, Convolutional networks for images, speech, and time series, The handbook of brain theory and neural networks 3361(10) (1995),