Comparison of Two Different Deep Learning Architectures on Breast Cancer

Feyza YILMAZ and Onur KÖSE
Department of Electrical and Electronics Engineering
Eskişehir Osmangazi University
Eskişehir/Turkey
f.yilmaz@ogu.edu.tr, onuur94@hotmail.com

Ahmet DEMİR
Department of Electrical and Electronics Engineering
Bartın University
Bartın/Turkey
ahmetdemir0627@gmail.com

***Abstract*—Breast cancer is one of the diseases becoming widespread gradually nowadays. Diagnosis and treatment of breast cancer are performed by some specialist doctors. Timely and accurate detection of this disease is lifesaving. DenseNet-201 and Xception deep learning architectures are used in this study. The performance of these two different deep learning methods are evaluated on the breast cancer dataset. The dataset consists of some benign and malignant cancer images. There are 20748 images for training and 5913 images for testing. According to the results obtained, DenseNet-201 method reaches an F-1 accuracy score of 92.24%, and the Xception method achieves an F-1 accuracy score of 92.41% when trained on the used dataset.**

Keywords — breast cancer; deep learning; DenseNet-201; Xception

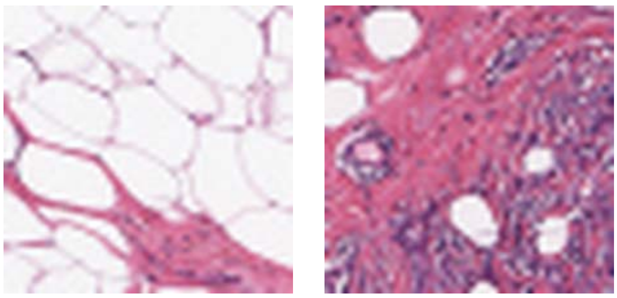
I. INTRODUCTION

Cancer is defined as tumors occurring as a result of division and increase of cells in a tissue or organ without any control. Tumors named as benign do not spread in human body and they are not fatal. Yet, the tumors named as malignant could spread in human body and they can cause death of humans [1].

According to the data saved in 2012, breast cancer is a cancer type seen at about 1.7 million women and known as the most commonly seen cancer type in comparison with the other cancer types [2]. By early diagnosis of breast cancer, 37.3% of patients can be healthy again completely [3]. Mammography is the most important visualization method which is used in diagnosis of breast cancer. Breast image is obtained in mammography with the aid of X-Rays having low energy [4]. Small changes in breast, which cannot be seen with medical examination done by hands, can be diagnosed with the aid of mammography. New research techniques have been needed for early diagnosis of breast cancer in recent years. Classification systems are used for the medical diagnosis by most of expert people in this area. Evaluations of doctors and their medical diagnosis are important. Still, deep learning techniques help us to prevent possible errors done by doctors [5]. Hence, deep learning algorithms are also used in diagnosis of breast cancer [6].

Some methods are developed and some studies are done for diagnosis of breast cancer. In a study, Computer Aided Diagnosis (CAD) system using convolutional neural networks is applied using four different datasets. Accuracy rates obtained for all the datasets are above 95% accuracy rate. According to these results, it can be seen that the proposed method can be used for accurate classification of mass lesions as benign or malignant [7]. In a study using Bayesian Linear Discriminant Analysis method, an accuracy rate of 83.45% is obtained as a result of a classification done by using the dataset obtained from oncology department of India Coimbatore, Sri Kuppuswamy Naidu Hospital [8]. In an another study, Wisconsin Breast Cancer dataset is used for classification of mass lesions as benign or malignant. K-Means algorithm is used for clustering, and C4.5 tree is used for classification. As a result of this study, an accuracy rate of 95.14% is acquired [9]. In an another study again using Wisconsin Breast Cancer dataset, classification is done by using three different decision trees. In this study, classification work is done by using three classifiers, which are J48 decision tree, REP Tree, and Random Tree. Best classification result is acquired as 97.36% by using J48 decision tree [10].

Some example images from our dataset are given in figure.1. Figure 1.a shows benign type image example, and figure 1.b shows malignant type image example. Two different deep learning techniques are used for the classification of breast cancer as benign or malignant in our study. The two different deep learning architectures used are DenseNet-201[11] and Xception [12].



a) Benign b) Malignant

Figure 1. Example test set images

978-1-7281-2420-9/19/$31.00 ©2019 IEEE

II. Material and Method

Our dataset is obtained from Kaggle website which is a public site [13]. Each of the images in our dataset has a pixel size of 50x50x3. Dataset consists of 198738 negative, and 78786 positive images in total. Training, validation and test operations are applied on 31827 images that are selected randomly from this dataset in our study. Training is done by using 20748 images taken from this dataset. Our training set consists of 10285 benign and 10463 malignant breast images. Our test set contains 2973 benign and 2940 malignant breast images so total test image number is determined as 5913 in our study. Lastly, our validation set contains 5166 images selected randomly from our dataset. In our study, Densenet-201 and Xception deep learning architectures which give high accuracy rates in classification tasks are used.

A. *DenseNet*

The employed structure for classification task, DenseNet(2017), stands for Dense Convolutional Network due to its dense connectivity pattern. The DenseNet network architecture used in this study is DenseNet-201. DenseNet architecture is given in Figure 2. This architecture takes input images which have a size of 224x224x3 everytime. Dense blocks used in DenseNet architecture include a batch normalization layer, a ReLU layer, and a 3x3 convolution layer. Unlike previous architectures that are using post-activation, DenseNet uses pre-activation. It is also vital to know that, DenseNet uses concatenating layers instead of summing layers used in the previous network architectures like ResNets. These concatenation layers form the output before sending to the following layers by concatenating the features coming from all the previous layers. Notwithstanding the previous network architectures pass all the feature maps including the ones that are not necessary to the following layers, DenseNet extract the ones that are redundant and doesnot have to re-learn them in the following layers and this is the reason why we have less parameters in DenseNet network. Moreover, we have some transition layers that are doing down-sampling operation and each of which including a batch normalization layer, a 1x1 convolution layer, and an average pooling layer. DenseNet is a stack of these dense blocks and transition layers. The architecture we are working on has 4 dense blocks and 3 transition layers. However, the dense block depths are different as shown in Table 1.

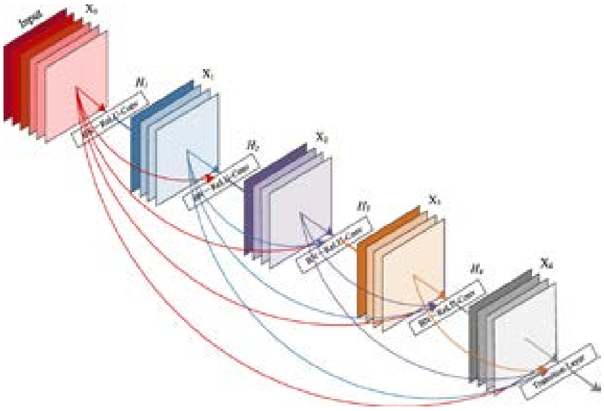


Figure 2. DenseNet network architecture [11].

B. *Xception*

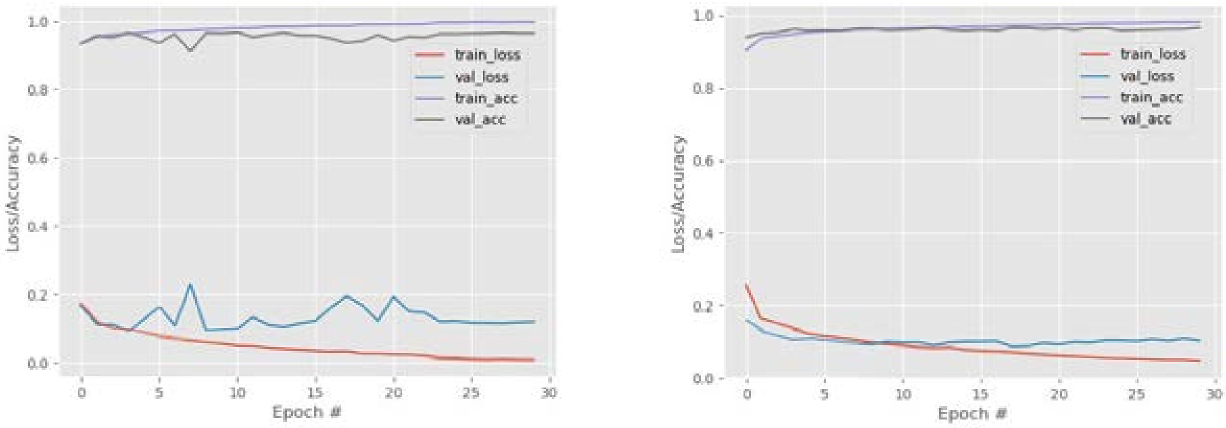
Xception (Chollet,2017), is a convolutional neural network standing for Extreme Inception [12]. It has pointwise(1x1) convolution operations first and then depthwise convolution operations following them. In Xception architecture, there are three main parts basically: input, middle and exit parts. Data enters from input part and flows through the middle part which is repeated eight times and lastly finish the process by flowing through the exit part of the network. The first thing seen in the network architecture is that input image size is determined as 299x299x3 pixels. Thus, all entering images must be in RGB format. Furthermore, all the convolution and separable convolution layers are followed by a batch normalization layer which is exploited to enhance the performance and speed of the network. There are 36 convolution layers used for feature extraction and we have 14 modules, all of which have linear residual connections, except for the first and last modules [12]. Shortcut connections between convolutional blocks as in ResNet [14] network architecture is also seen in Xception network architecture. In network architecture, all depthwise separable convolutional blocks are followed by a maxpooling layer. Convolution filter sizes are selected as 3x3 in all depthwise separable convolutional blocks. Yet, 1x1 filter size is used in all shortcut connections. Along with, ReLU nonlinearity is added to the network by using ReLU layers before all the separable convolution layers. At the end of the network Global Average Pooling is used to cut down on the parameter number and so to diminish overfitting problem.

TABLE I. DENSENET ARCHITECTURE [11]

Layers	Output Size	DenseNet-201(k=32)
Convolution	112×112	7×7 conv, stride 2
Pooling	56×56	3×3 max pool, stride 2
Dense Block (1)	56×56	$\begin{bmatrix}1\times1\ conv\\3\times3\ conv\end{bmatrix}\times 6$
Transition Layer (1)	56×56	1×1 conv
	28×28	2×2 average pool, stride 2
Dense Block (2)	28×28	$\begin{bmatrix}1\times1\ conv\\3\times3\ conv\end{bmatrix}\times 12$
Transition Layer (2)	28×28	1×1 conv
	14×14	2×2 average pool, stride 2
Dense Block (3)	14×14	$\begin{bmatrix}1\times1\ conv\\3\times3\ conv\end{bmatrix}\times 48$
Transition Layer (3)	14×14	1×1 conv
	7×7	2×2 average pool, stride 2
Dense Block (4)	7×7	$\begin{bmatrix}1\times1\ conv\\3\times3\ conv\end{bmatrix}\times 32$
Classification Layer	1×1	7×7 global average pool
		1000D fully- connected layer, softmax

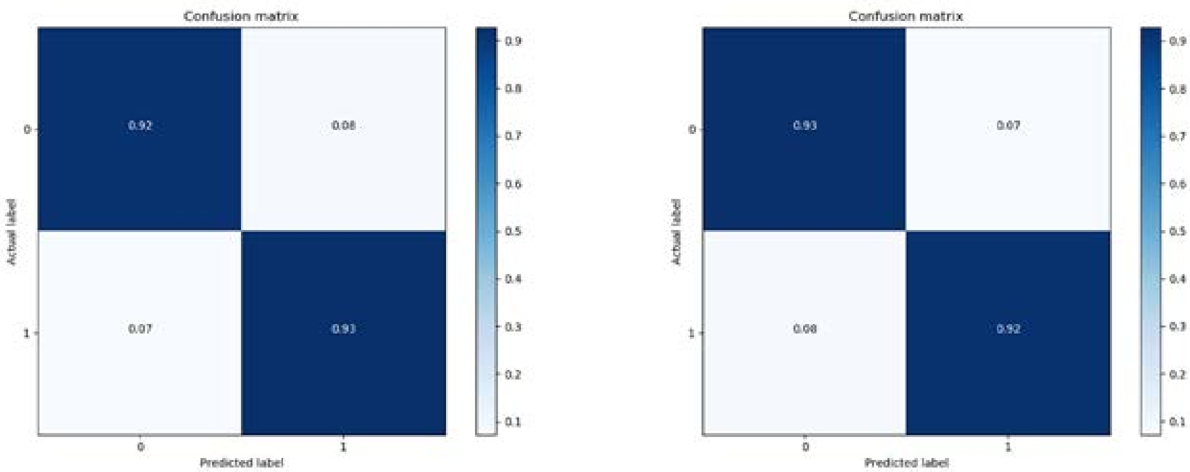
III. Experiments

We demonstrate the effectiveness of DenseNet-201 and Xception on our dataset and compare the acquired classification results obtained by using these two modern neural network architectures.



a) DenseNet b) Xception

Figure 3. Accuracy and loss graph.



a) DenseNet b) Xception

Figure 4. Confusion matrix.

A. Classification Results on DenseNet-201 and Xception

20748 images are used for the training of used architectures. The training process is completed in 30 epochs. The learning rate factor is set to 0.001 at the beginning of training and is decreased at every epoch of training. Model accuracy/loss graph is shown in figure 3. The accuracy and loss graph for DenseNet-201 architecture is given in Figure 3.a. According to Figure 3.a, train accuracy and validation accuracy reach saturation after 25 epochs while train loss and validation loss continue to decrease even after 20 epochs.

The graph of accuracy and loss for the Xception architecture is given in Figure 3.b. According to Figure 3.b, the train accuracy and validation accuracy reach saturation after 20 epochs. After 20 epochs, train loss continues to decrease, while validation loss stops decreasing. The classification results of the two architectures are given in Table 2. These results show that both architectures have classification accuracy rates very close to each other. Confusion matrices are given for Densenet-201 (Figure 4.a) and Xception (Figure 4.b) architectures. In Figure 3, the number of correct and incorrect predictions in the test dataset are given. 0 and 1 in the confusion matrix represent benign cancer and malignant cancer, respectively.

TABLE II. CLASSIFICATION RESULTS

Model	Validation Accuracy	Benign (0) Accuracy	Malign (1) Accuracy	F-1 Score
DenseNet-201	0.9674	0.9177	0.9268	0.9224
Xception	0.9669	0.9280	0.9204	0.9241

IV. CONCLUSION

We use two deep learning algorithms: DenseNet-201 and Xception and use the Breast Cancer dataset on the Kaggle website. Our aim in the classification of breast cancer images is to obtain a definite and reliable result. Comparing the algorithms, DenseNet-201 has an accuracy rate of 92.24%, while Xception has an accuracy rate of 92.41%. The results of both algorithms are very close to each other and have a good accuracy.

REFERENCES

[1] Subashini, T. S., Ramalingam, V. and Palanivel, S. "Breast mass classification based on cytological patterns using RBFNN and SVM", *Expert Systems with Applications*, 36(3): 5284-5290, 2009.

[2] Web address: http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/breast-cancer-statistics, Date of access: 24.06.2019

[3] Salama, G. I., Abdelhalim, M., and Zeid, M. A. "Breast cancer diagnosis on three different datasets using multi-classifiers" *Breast Cancer (WDBC)* 32(569): 2, 2012.

[4] Parthasarathy, P., and Vivekanandan, S. "A typical IoT architecture-based regular monitoring of arthritis disease using time wrapping algorithm" *International Journal of Computers and Applications* 1-11, 2018.

[5] Akay, M. F. "Support vector machines combined with feature selection for breast cancer diagnosis" *Expert systems with applications* 36(2): 3240-3247,2009.

[6] Mohebian, M. R., et al. "A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning" *Computational and structural biotechnology journal* 15: 75-85,2017.

[7] Chougrad, H., Zouaki, H. and Alheyane, O. "Deep convolutional neural networks for breast cancer screening" *Computer methods and programs in biomedicine* 157: 19-30, 2018.

[8] Rajaguru, H., and Prabhakar, S. K. "Bayesian linear discriminant analysis for breast cancer classification" *2017 2nd International Conference on Communication and Electronics Systems (ICCES). IEEE,* 266-269, 2017.

[9] Abdel-Zaher, A. M., and Eldeib, A. M. "Breast cancer classification using deep belief networks" *Expert Systems with Applications* 46: 139-144, 2016.

[10] Douangnoulack, P. and Boonjing, V. "Building Minimal Classification Rules for Breast Cancer Diagnosis" *2018 10th International Conference on Knowledge and Smart Technology (KST). IEEE,* 2018.

[11] Huang, G., et al. "Densely connected convolutional networks" *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2017.

[12] Chollet, F. "Xception: Deep learning with depthwise separable convolutions" *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2017.

[13] Web address: https://www.kaggle.com/paultimothymooney/breast-histopathology-images , Date of access: 28.06.2019

[14] He, K. et al. "Deep residual learning for image recognition" *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2016.