Diagnosis of Pneumonia from Chest X-Ray Images using Deep Learning

Enes AYAN*
Department of Computer Engineering
University of Kırıkkale
Kırıkkale, Turkey
enesayan@yandex.com

Halil Murat ÜNVER
Department of Computer Engineering
University of Kırıkkale
Kırıkkale, Turkey
hmunver@hotmail.com

***Abstract*—Pneumonia is a disease which occurs in the lungs caused by a bacterial infection. Early diagnosis is an important factor in terms of the successful treatment process. Generally, the disease can be diagnosed from chest X-ray images by an expert radiologist. The diagnoses can be subjective for some reasons such as the appearance of disease which can be unclear in chest X-ray images or can be confused with other diseases. Therefore, computer-aided diagnosis systems are needed to guide the clinicians. In this study, we used two well-known convolutional neural network models Xception and Vgg16 for diagnosing of pneumonia. We used transfer learning and fine-tuning in our training stage. The test results showed that Vgg16 network exceed Xception network at the accuracy with 0.87%, 0.82% respectively. However, the Xception network achieved a more successful result in detecting pneumonia cases. As a result, we realized that every network has own special capabilities on the same dataset.**

Keywords— Pneumonia; transfer learning; Xception; Vgg16; deep learning.

I. INTRODUCTION

Pneumonia is inflammation of the tissues in one or both lungs that usually caused by a bacterial infection. In the USA annually more than 1 million people are hospitalized with the gripe of pneumonia. Unfortunately, 50.000 of these people die from this illness [1]. Fortunately, pneumonia can be a manageable disease by using drugs like antibiotics and antivirals. However, early diagnosis and treatment of pneumonia is important to prevent some complications that lead to death [2]. Chest X-ray images are the best-known and the common clinical method for diagnosing of pneumonia [3]. However, diagnosing pneumonia from chest X-ray images is a challenging task for even expert radiologists. The appearance of pneumonia in X-ray images is often unclear, can confuse with other diseases and can behave like many other benign abnormalities. These inconsistencies caused considerable subjective decisions and varieties among radiologists in the diagnosis of pneumonia [4-6]. Therefore, there is a need for computerized support systems to help radiologists for diagnosing pneumonia from chest X-ray images. Recent developments in deep learning field, especially convolutional neural networks (CNNs) showed great success in image classification [7]. The main idea behind the CNNs is creating an artificial model like a human brain visual cortex. The main advantage of CNNs, it has the capability to extract more significant features from the entire image rather than hand-crafted features [7, 8]. Researchers developed different CNN based deep networks and these networks achieved state of results in classification, segmentation, object detection and localization in computer vision [9-11]. Besides the natural computer vision problems, CNNs achieved very successful results in solving medical problems such as breast cancer detection [12], brain tumor segmentation [13], alzheimer disease diagnosing, skin lesion classification [14, 15] etc. The detailed reviews presented here about deep learning in medical image analysis [16, 17]. As far as, we are realized that there are a few studies about the detection of pneumonia using deep learning. In 2017, Antin et al. used a DenseNet-121 layer with utilizing transfer leaning method and they achieved 0.60 % area under the curve (AUC) value [18]. In 2017, Rajpurkar, et al. proposed a 121-layer convolutional neural network based on DenseNet [19] and named as CheXNet [20]. They trained their network with 10.000 frontal view chest X-ray images with 14 different diseases. They assessed the performance of their network with four expert radiologists on the f1 score metric which is the harmonic average of the precision and recall metrics. CheXNet achieved a f1 score of 0.435 (95% CI 0.387, 0.481), higher than the radiologist average of 0.387 (95% CI 0.330, 0.442). Based on this information, we modified and trained two well-known networks for classifying pneumonia from chest X-ray images. Our first network is based on the Xception model [21]. The second one is Vgg16 based model [22]. Besides, we utilized transfer learning, fine-tuning and data augmentation methods. For an objective comparison between them, we used same parameters when training both networks. Also, we compared the performance of two networks on the test data with different metrics. The results show that the Xception model outperforms Vgg16 model in diagnosing pneumonia. On the other hand, Vgg16 model showed better performance in diagnosing normal cases.

II. MATERIALS AND METHODS

A. Data

In this study, dataset consisting of 5856 frontal chest X-ray images provided by Kermany et al. [23]. The images in the dataset are varying resolutions such as 712x439 to 2338x2025. There are 1583 normal case, 4273 pneumonia case images in the dataset. Fig. 1 shows some X-ray image samples from the dataset. Table 1 represents the distribution of the data when

978-1-7281-1013-4/19/$31.00 ©2019 IEEE

training, validating and testing phases of the proposed model. In our models 0 represents normal cases, 1 represents pneumonia cases.

TABLE 1. Distribution of dataset

	Train	Validation	Test
Normal	1349	234	234
Pneumonia	3883	390	390
Total	5232	624	624

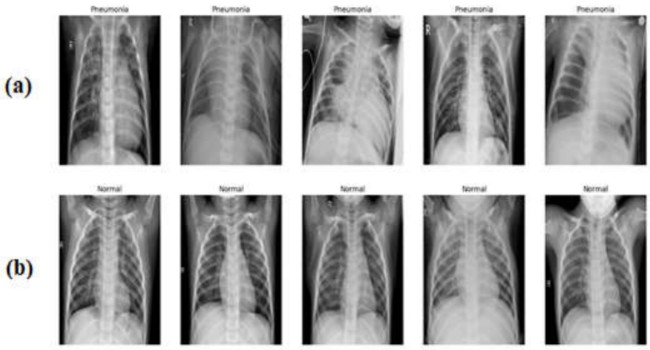

Fig. 1. Data samples from the dataset, (a) shows pneumonia cases and (b) shows normal cases

B. *Data augmentation and transfer learning*

Deep learning needs a huge amount of data to obtain a reliable result. However, there may be not enough data in some problems. Especially on medical problems, to obtain and annotate the data is very expensive and time-consuming process. Fortunately, there are some solutions to solve this problem. One of them is data augmentation which avoids the overfitting and improves the accuracy [15]. In this work, training time data augmentation method was utilized. We used different augmentation methods such as shifting, zooming, flipping, rotating at 40-degree angles. Another performance-enhancing method in deep models especially in CNNs which is named as transfer learning. Transfer learning is the idea of overcoming the isolated learning paradigm and utilizing knowledge acquired for one task to solve related ones [24]. Nowadays, a few people train an entire CNN from scratch. Because it needs a huge amount of data. Instead, using pretrained CNNs on very large dataset such as ImageNet which is contain 1.2 million image and 1000 class [25]. There are three different transfer learning approaches in CNNs. These are feature extractor, fine-tuning and pretrained models [26]. In this study we used fine tuning approach which is motivated by observing in early layers of CNNs have more generic features such as edges, colors, blobs. So, this layer should be useful for many other tasks. But last layers have more data specific features. Therefore, we fixed some early layers of our models and trained our models excluding fixed layers.

C. *CNN architecutres*

In this study, we used two well-known CNN networks Xception [21] and Vgg16 [22]. Xception model stands for the extreme version of Inception model. It uses a modified depth wise separable convolutional layer. In the first step, it uses a 1x1 pointwise convolution and follows by a 3x3 depth wise convolution. Fig. 2 shows used separable convolution in Xception model. The Xception architecture has 36 convolutional layers for feature extraction. After convolutional layer followed by a logistic regression layer. If desired, fully connected layers can be used between convolutional layers and logistic regression layer. The convolutional layers structured as 14 modules and all of them has a linear residual connection between them. The base model Xception has archived better performance than Inception-V3 [27] classification of ImageNet dataset.

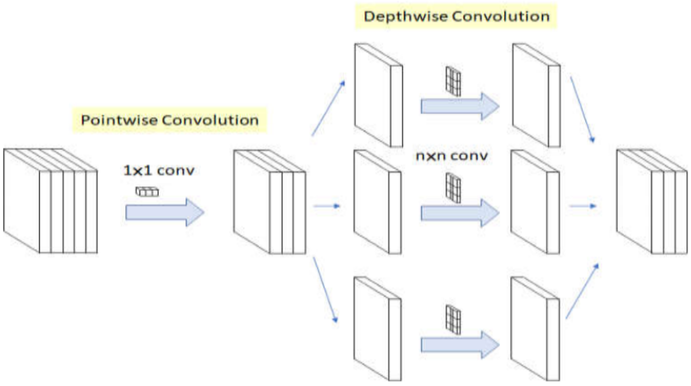

Fig. 2. Separable convolution process used in the Xception model [28]

We utilized transfer learning and fine-tuning on the Xception model, therefore used pretrained ImageNet weights before the start to train and the last 10 early layers were frozen. After the global average polling layer, we added two fully connected layers (1024,512) and a two-way output layer with SoftMax activation function. Fig. 3 shows the proposed Xception model.

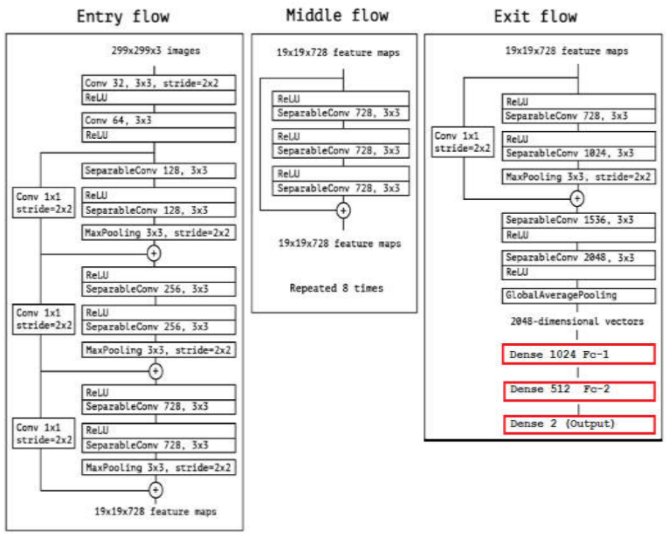

Fig. 3. Fine-tuned Xception network [21]

In 2014, Simonyan et al. proposed a deep model named as VGG-16. The model has 16 convolutional layers with small receptive fields (3x3), 144 million parameters, five max-pooling layers (2x2 size) and three fully-connected layers, with the final layer has a soft-max activation function. We used this model with pre-trained weights on ImageNet and modified fully connected layers of the network. Also, we fixed the last 8 early layers to avoid the train. Fig. 4 shows our modified network.

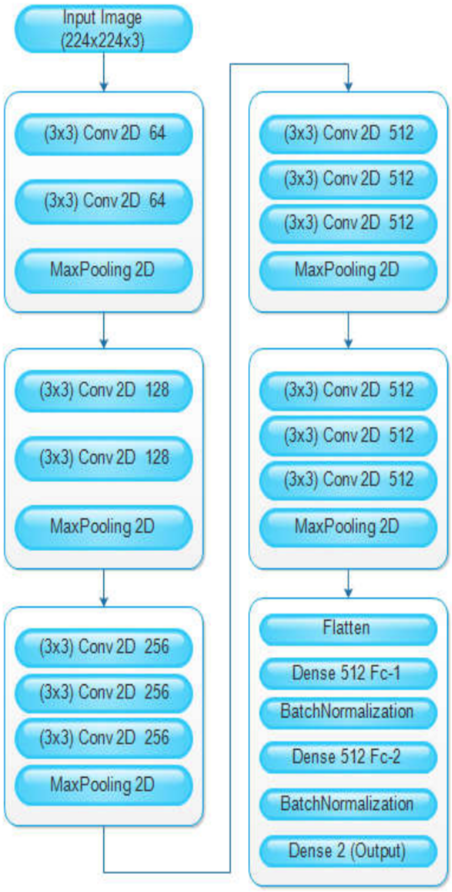

Fig.4. Fine-tuned Vgg16 network

III. EXPERIMENTAL RESULTS

In this section, training strategies and test results were presented. Before the training phase, all images are resized for the target network model. Because, Xception network accepts images at 299x299x3 dimensions whereas Vgg16 network accepts images at 224x224x3 dimensions. Also, all image pixels were normalized range in [0,1]. We trained both networks by using the same parameters. These train parameters are, epoch size is set as 50, categorical cross entropy selected as loss function, RMSprop used as the optimizer, learning rate set as 1e-4, weight decay set as 0.9 and batch size set as 16. Different approaches were used for avoiding overfitting. Firstly, batch normalization was used after every convolutional layer. Secondly, the dropout method was used after fully connected layers with 0.5 rate. Lastly data augmentation was used for avoiding overfitting. Fig.5 and Fig. 6 shows Xception and Vgg16 network's accuracy and loss graphics respectively. The proposed networks are implemented by using Keras deep learning framework [29] using Python programing language on an Ubuntu operating system and Nvidia 1080 TI GPU. Training and test calculation run times of two model are given in Table 2. Given test time in Table 2 represents evaluation time of 624 test images. The estimated time per image is 0.016 and 0.020 seconds Vgg16 and Xception models respectively. These results show that the models are suitable for real-time predictions.

TABLE 2. Network's train and test calculation time

	Train Time	Test Time
Vgg16	83 minutes	10 seconds
Xception	100 minutes	13 seconds

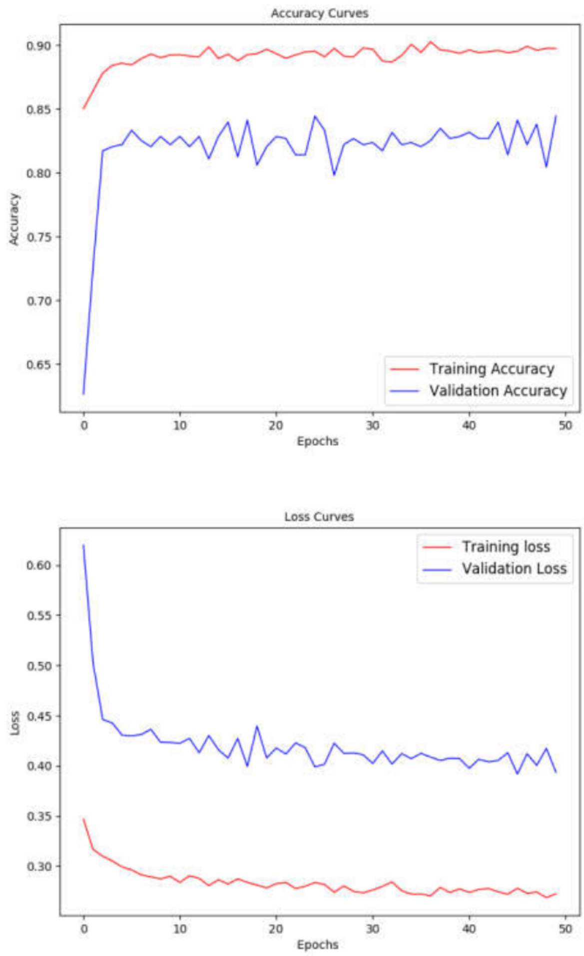

Fig. 5. Xception model's accuracy and loss graphics

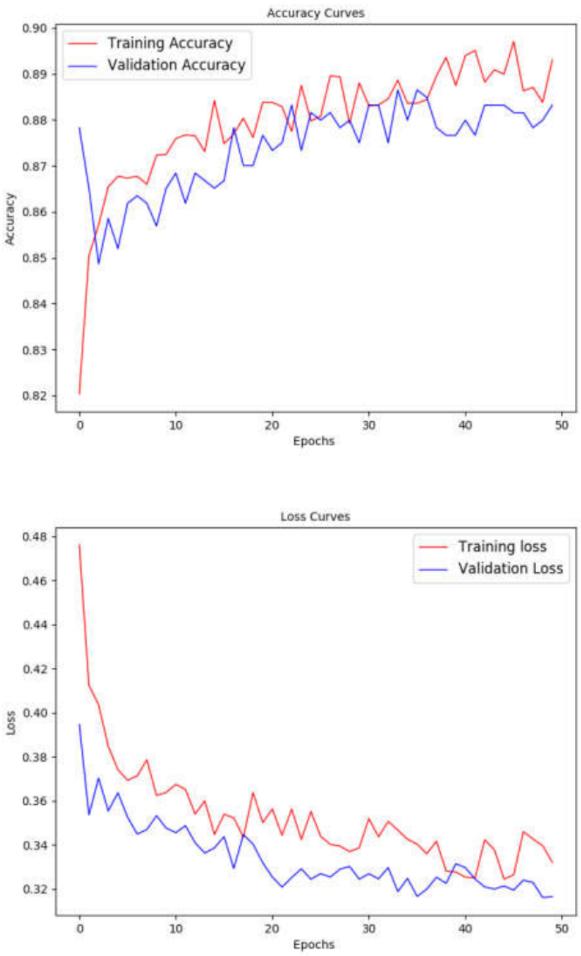

Fig. 6. Vgg16 model's accuracy and loss graphics

A. *Evaluation metrics*

The proposed models were evaluated by using performance metrics such as accuracy, sensitivity, specificity, recall, precision, and f1 score. The formulas of the metrics are given below:

$$\text{Accuracy} = (TP+TN)/(TP+FN+TN+FP) \quad (1)$$

$$\text{Sensitivity} = TP/(TP+FN) \quad (2)$$

$$\text{Specificity} = TN/(TN+FP) \quad (3)$$

$$\text{Precision} = TP/(TP+FP) \quad (4)$$

$$\text{Recall} = TP/(TP+FN) \quad (5)$$

$$\text{F1 score} = 2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}) \quad (6)$$

TP, TN, FN, FP represents the number of true positive, true negative, false negative, false positive respectively.

B. *Results*

We evaluated two models by using 624 frontal chest X-ray images. The test set contains 234 normal and 390 pneumonia cases. Table 3 shows a comparison of two networks in terms of accuracy (acc), sensitivity (sen), specificity (spec) metrics. Also, Table 4 and Table 5 presents case base precision, recall and f1 scores of two networks. In addition, Fig. 7 shows both network's confusion matrices.

TABLE 3. Results of two networks by accuracy, sensitivity, and specificity metrics

	accuracy	sensitivity	specificity
Xception	0.82	**0.85**	0.76
VGG16	**0.87**	0.82	**0.91**

TABLE 4. Results of Xception network by precision, recall, f1 score metrics

Xception	precision	recall	f1 score
Normal	**0.86**	0.65	0.74
Pneumonia	0.82	**0.94**	0.87

TABLE 5. Results of Vgg16 network by precision, recall, f1 score metrics

Vgg16	precision	recall	f1 score
Normal	0.83	**0.86**	**0.84**
Pneumonia	**0.91**	0.89	**0.90**

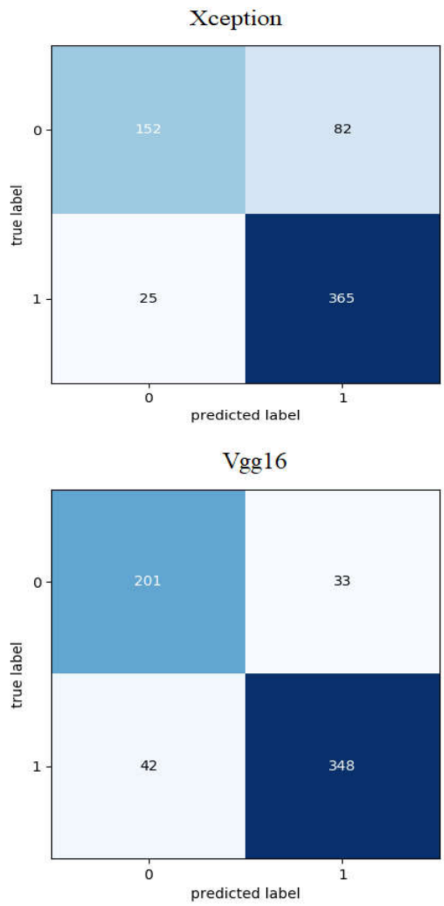

Fig.7. Xception and Vgg16's confusion matrices

IV. CONCLUSION

In this study, we compared two CNN network's performance on the diagnosis of pneumonia disease. While training our model we used from transfer learning and fine-tuning. After the training phase, we compared two network test results. The test results showed that Vgg16 network outperforms Xception network by accuracy 0.87 %, specificity 0.91%, pneumonia precision 0.91% and pneumonia f1 score 0.90%. Whereas Xception network outperforms Vgg16 network by sensitivity 0.85%, normal precision 0.86% and pneumonia recall 0.94%. According to the experimental results and confusion matrices in Fig. 7 every network has own detection capability on the dataset. Xception network is more successful for detecting pneumonia cases than Vgg16 network. At the same time Vgg16 network is more successful at detecting normal cases. In the future work we will ensemble of two networks. In this way we will combine strengths of two networks and will achieve more successful results on diagnosing of pneumonia from chest X-ray images.

ACKNOWLEDGMENT

This work was supported by Research Fund (Scientific Research Projects Coordination Unit) of the Kırıkkale University. Project Number:2018/40.

REFERENCES

[1] CDC. (2019). Pneumonia Statistics. Available: https://www.cdc.gov/pneumonia/prevention.html

[2] M. Aydogdu, E. Ozyilmaz, H. Aksoy, G. Gursel, and N. Ekim, "Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores," Tuberk Toraks, vol. 58, no. 1, pp. 25-34, 2010.

[3] W. H. Organization, "Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children," 2001.

[4] M. I. Neuman et al., "Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children," Journal of hospital medicine, vol. 7, no. 4, pp. 294-298, 2012.

[5] H. D. Davies, E. E.-l. Wang, D. Manson, P. Babyn, and B. Shuckett, "Reliability of the chest radiograph in the diagnosis of lower respiratory infections in young children," The Pediatric infectious disease journal, vol. 15, no. 7, pp. 600-604, 1996.

[6] R. Hopstaken, T. Witbraad, J. Van Engelshoven, and G. Dinant, "Inter-observer variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections," Clinical radiology, vol. 59, no. 8, pp. 743-752, 2004.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, p. 436, 2015.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234-241: Springer.

[11] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[12] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," PeerJ, vol. 7, p. e6201, 2019.

[13] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1240-1251, 2016.

[14] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, p. 115, 2017.

[15] E. Ayan and H. M. Ünver, "Data augmentation importance for classification of skin lesions via deep learning," in 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2018, pp. 1-4: IEEE.

[16] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," IEEE Access, vol. 6, pp. 9375-9389, 2018.

[17] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical image analysis, vol. 42, pp. 60-88, 2017.

[18] B. Antin, J. Kravitz, and E. Martayan, "Detecting Pneumonia in Chest X-Rays with Supervised Learning."

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700-4708.

[20] P. Rajpurkar et al., "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.

[21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," arXiv preprint, p. 1610.02357, 2017.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[23] Z. Kermany Daniel, Kang, Goldbaum Michael "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," ed, 2018.

[24] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big Data, vol. 3, no. 1, p. 9, 2016.

[25] O. Russakovsky et al., "Imagenet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211-252, 2015.

[26] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in Advances in neural information processing systems, 2014, pp. 3320-3328.

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818-2826.

[28] S.-H. Tsang. (2018). Review: Xception—With Depthwise Separable Convolution, Better Than Inception-v3. Available: https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568

[29] F. Chollet, "Keras," ed, 2015.