# Turkish Journal of Engineering

*Turkish Journal of Engineering (TUJE)*
*Vol. 4, Issue 3, pp. 129-141, July 2020*
*ISSN 2587-1366, Turkey*
*DOI: 10.31127/tuje.652358*
*Research Article*


## CLASSIFICATION PERFORMANCE COMPARISONS OF DEEP LEARNING MODELS IN PNEUMONIA DIAGNOSIS USING CHEST X-RAY IMAGES


Osman Doğuş Gülgün *[1] and Hamza Erol [2]

[1] Mersin University, Graduate School of Natural and Applied Sciences, Department of Computer Engineering, Mersin, Turkey
ORCID ID 0000 – 0003 – 1824 – 4401
dogusgulgun@gmail.com

[2] Mersin University, Engineering Faculty, Department of Computer Engineering, Mersin, Turkey
ORCID ID 0000 – 0001 – 8983 – 4797
herol@mersin.edu.tr

* Corresponding Author



Received: 28/11/2019 Accepted: 23/12/2019



**ABSTRACT**

In recent years, the analysis of medical images using deep learning techniques has become an area of increasing popularity. Advances in this area have been particularly evident after the discovery of deep artificial neural network models and achieving more successful performance results than other traditional models. In this study, the performance comparison of different deep learning models used to efficiently diagnose pneumonia on chest x-ray images was performed. The data set used in the study consists of a total of 5840 chest x-ray images of individuals. In order to classify these data, three different deep learning models are used: Convolutional Neural Network, Convolutional Neural Network with Data Augmentation and Transfer Learning. The images in the data set were classified into two categories as pneumonia and healthy people using these three deep learning models. The performances of these three deep learning models used in classification were compared in terms of loss and accuracy. In the comparison of three different deep learning models with two different performance values, 5216 chest x-ray images in the data set were used to train the deep learning model and the remaining 624 were used to test the model. At the end of the study, the most successful performance result was obtained by convolutional neural network model applied with data augmentation technique. According to the best results of this study, this model was able to accurately predict the class of 93.4% of the test data.



***Keywords:*** *Deep Learning, Medical Diagnosis, Data Augmentation, Convolutional Neural Network, Transfer Learning*

## 1. INTRODUCTION

In this section, previous studies about medical image classification and deep learning in the literature are discussed and the techniques used in these studies are explained.

In recent years, there have been significant developments in medical image analysis and machine learning. The most important developments in this field have been experienced especially after the emergence of deep artificial neural network models and performing better than other models. Following the intense use of artificial neural networks, the number of studies on medical imaging, medical data analysis and disease diagnosis has increased and many of them have been realized with significant potential in this field. In their study, Lundervold and colleague examined the deep learning and machine learning studies in the field of medical imaging performed on MRI images and gave information about current studies in this field (Lundervold and Lundervold, 2019). In their studies, they examined a wide range of studies such as data generation using generative adversarial network model, image classification with convolutional neural networks. They discussed current reference works in this field. Especially convoluted neural networks have shown very good results in the classification of images in the field of medical image analysis, the detection of abnormalities in medical images, and the identification of the most important features on medical images. In addition to data analysis, features obtained from convolutional neural networks can also be used for different purposes, such as the generation of new data, very similar to the original. Therefore, because of the successful performance of deep learning models observed in previous studies in this field, it was considered appropriate to use deep learning techniques to diagnose pneumonia on x-ray images in this study.

Convolutional neural networks are one of the most powerful computerized vision techniques in terms of usability for different tasks. Recent studies in the field of computer vision have shown that the features obtained using convolutional neural can be used in a classifier other than the original network structure after the completion of neural network training. Van Ginneken *et al*, the study of the detection of nodules in the lung on computed tomography images, can be given as an example of studies on this subject (Van Ginneken *et al*., 2015). In this study, 865 computed tomography scanning images of publicly available LIDC data set were used. The images were evaluated by 4 qualified radiologists and the classes of the images were determined. Using 2 dimensional sagittal, coronal and axial parts in all images, 4096 features were extracted for each image. These features were classified using a linear support vector machine.

Recent studies have shown that deep learning techniques have significant advantages over traditional methods based on handmade attributes. However, deep learning techniques also have some limitations due to the similarities and differences of the data in different classes caused by class diversity in various medical scenarios. In order to reduce these limitations, Zhang *et al*. proposed a synergic deep learning model in their study (Zhang *et al*., 2019). In their synergic deep learning model, more than one convolutional neural network model was used and they enabled each other to learn from each other. Both convolutional neural networks were designed according to the ResNet50 architecture. If one of the convolutional neural network performs the correct classification while the other performs the wrong classification, this false creates an extra synergic force for updating the weights of the faulty model. Therefore, in this synergic model, networks are mutually learned through classification errors. The model used in the study was evaluated using 4 different data sets. According to the results, the synergic deep learning model obtained the best results for each data set.

Recent research in the field of deep learning reveals that deep neural networks are highly sensitive to small irregularities and differences in images. Although this provides an advantage in some special studies, it is a significant disadvantage that may adversely affect the classification performance. Li *et al*. conducted studies with 3D brain MRI images to examine the effect of such adverse conditions on the medical image processing field (Li *et al*., 2019). In their study, they were interested in designing deep learning models that could predict the age of subjects on 3D brain MR images. Their data set consists of 3D brain MRI images of 3921 subjects obtained from 7 different data sets. The subjects were between 4 and 94 years old and the mean age was 25.5 years. They used two different models for estimating images: a conventional deep neural network and a hybrid deep learning model using attributes reported by the anatomical context. In addition, they succeeded in creating incorrect results in age estimation by adding noise to the images. They found that their hybrid model obtained more successful estimation results on noisy images. As a result of the noise they added to the images, they realized that the age of a 19-year-old person could be estimated as 80 due to the noise in the image. This reveals that differences and noise in images significantly affect results in deep learning models.

Due to the limitations of data sources and the unbalanced number of samples of different classes in the data set, it is difficult to perform computer-assisted segmentation of 3D medical images at high levels of success using deep learning methods. Indraswari *et al*. proposed an advanced deep learning model for segmentation of 3D images (Indraswari *et al*., 2019). They used three different data sets in their studies. The first and second data sets were used for brain tumor segmentation and the third data set was used for dental segmentation of the jaw images of individuals. The first two sets of data were MRI images of low and high-grade glioma patients, while the third data set consisted of human jaw scanning images taken with cone beam computed tomography. They propose a convolutional autoencoder architecture consisting of encoder and decoder structures to perform segmentation of the image data they use. Since the input images used in the proposed architecture are 3D, these images are first decomposed into axial, coronal and sagittal components to obtain a 2D image component for each plane. The images are passed in three sets of convolution and deconvolution layers to obtain the most important information separately for each plane. The information obtained from axial, coronal and sagittal slices are combined after the final pooling process to obtain the actual input vector of the model. Then, the combined input information is convolution to obtain the most important features. They also proposed a special cost function for their models. The proposed cost function adds a weight to the network model to take into account the probability of each class in the data classes. Despite the imbalance of the class data in 3 different data sets, they achieved good results with the proposed cost function and model approach.

In another important study in this field, Şengür *et al*. proposed a hybrid classification model approach to perform optical disc detection on eye retinal images (Şengür *et al*., 2018). They used convolutional neural network architecture together with the k nearest neighbors classifier. The features in the fully connected layer of the convolutional neural network model, which were previously trained on retinal images, were extracted and these features were used to classify optic disc and non-optic disc regions with k nearest neighbor classifier. Using the AlexNet architecture as a convolutional neural network, they extracted 4096 dimensional feature vectors of each image. They used 3 different retinal image data sets to create the training and test data of the model they created. Patches of 280 x 280, 500 optical discs and 1565 non-optical discs were collected from the datasets, and then the patches were scaled to 227 x 227 to extract feature from these patches. The k nearest neighbor classifier was trained and tested on the generated features, and 165 retinal images were used in the test phase. The accuracy, sensitivity and specificity criteria of the hybrid model were examined in order to test the hybrid model. According to their results, the models reached 95.74% accuracy, 84.46% sensitivity and 99.08% specificity.

Determining the laterality of speech before the operations performed is very important in terms of predicting the possible risks accurately. Toroman *et al*. have studied the use of non-invasive machine learning techniques to determine speech laterality over EEG signals (Toroman *et al*., 2019). In their study, they used data obtained from 67 subjects diagnosed as healthy according to EEG examination in the 18-65 age range. In the data used, 35 of the subjects were individuals of right dominant type and the speech center was located in the left hemisphere of the brain. The remaining 32 subjects are left dominant individuals and the speech center is located in the right hemisphere of the brain. In this study, a spectrogram image was created for each of the 18 EEG channels using various convolutional neural network architecture including VGG16, VGG19, ResNet, MobileNet, NasNet and DenseNet. Then, attributes were extracted from the images using these network architectures. Extracted attributes were classified using support vector machines. According to the results obtained from the study, it was observed that VGG16 network architecture is more successful than other architectures.

The most dangerous type of sleep disorder is obstructive sleep apnea syndrome (OSAS) that occurs during sleep and can cause sudden death of patients. This disease depends on many parameters and the diagnosis of the disease is laborious and time consuming for the experts. Tuncer *et al*. have studied the design of a deep learning-based decision support system to diagnose OSAS with PPT signals (Tuncer *et al*., 2019). In their study, they used only Pulse Transition Time (PTT) signal parameter in classification of patients and healthy individuals instead of classical diagnostic parameters used in the literature. In the study, feature extraction from PPT signals was performed using AlexNet and VGG16 convolutional neural network models. Then, healthy and patient individuals were classified by using this extracted features with k nearest neighbors and support vector machines classifiers. At the end of the study, the best results were obtained by classifying the features extracted using VGG16 network architecture with support vector machines classifier. This hybrid structure was able to accurately predict the classes of 92.78% of the test data.

In another study on the detection of abnormalities from medical images, Setio *et al*. proposed a computer-assisted detection system using a multi-view convolutional artificial neural network (ConvNets) for pulmonary nodule detection (Setio *et al*., 2016). The artificial neural network used in their studies was formed by using the features of three different convolutional neural networks previously used in the detection of solid, subsolid and large nodules. They used data augmentation and dropout techniques to prevent overfitting problems during the training of their models. Tested their model on 888 lung tomography scans containing 1186 lung nodules. The model succeeded in detecting 1016 of 1186 nodules and reached 85.7% detection accuracy rate.

In another similar study, Roth *et al*. conducted a study to improve the performance of the three computer-aided detection systems previously established (Roth *et al*., 2015). They were interested in detecting colon polyps from computed tomographic colonography images. In addition, they studied for the detection of enlarged lymph nodes and the detection of spinal metastases on body computed tomography images. In this study, they benefited from the CNN architecture and they applied data augmentation technique to their existing model images by randomly rotating them up to 100 degrees on three orthagonal axes. The models developed as a result of the methods used were evaluated over 3 different data sets consisting of 1186 for detecting colon polyps, 176 for detecting enlarged lymph nodes and 59 for detecting spinal metastases. According to the results, the sensitivity of lesion detection was increased in the range of 13-34% for 3 models developed independently for each data set. Sensitivity values increased from 57% to 70%, 43% to 77% and 58% to 75% for spine metastasis detection, enlarged lymph node detection and colon polyp detection models, respectively.

Gastroscopy is a diagnostic method in which the gastrointestinal system structure can be examined directly. This method is widely used in gastrointestinal examinations. Diagnosis of gastrointestinal diseases through endoscopy images requires considerable experience and can only be diagnosed by experienced physicians. It is difficult for clinicians other than specialist physicians to correctly diagnose such diseases. In order to overcome this problem, Zhu *et al*. proposed a computer-assisted lesion detection system that works on endoscopy images (Zhu *et al*., 2015). They also used a convolutional neural network model to extract the most important features from images. They used image features from the fully connected layer of convolutional neural network to train and test a support vector machine classifier. At the end of the study, they observed that their models performed better than other endoscopic image classification models in the literature.

Histopathology images are obtained from tissues suspected of disease. These images are examined under a light microscope and used in the definitive diagnosis of many diseases, especially cancer. In order to benefit from the definitive diagnosis of histopathology images, Wang *et al*. proposed a classification model that provides diagnosis by performing classification directly on these images (Wang *et al*., 2017). In their models, they used a different network architecture called Bilinear CNN (BCNN) by making use of convolutional neural network architecture. In the first step, histopathology images are divided into two different components. Afterwards, these image components are evaluated with different convolutional neural networks and feature extraction is performed separately from the components. The feature vectors obtained from two separate convolutional neural networks for two different components are combined to obtain a high dimensional feature vector. A support vector machine (SVM) classifier is trained using this high dimensional feature vector, and then the performance of the trained model is tested. At the end of the study, the proposed BCNN based classification model accurately predicted the class of most of the test data and reached a classification accuracy rate of 92.6%. In addition, the proposed model achieved 92.8% sensitivity and 98.9% specificity. This study has achieved the most successful results in terms of correct classification of images when compared with other studies in this field in the literature.

The studies described in this section show that deep learning techniques can be used for different purposes such as classification, object detection and data generation in the medical field. The results obtained from previous studies with deep learning techniques show that deep learning techniques can be used efficiently in this field. In our study, we propose different model architectures that can be used to diagnose pneumonia with chest x-ray images using deep learning techniques. After creating the proposed models, we compare the performance of these models.

The dataset, used in this study, containing chest x-ray images consists of image data belonging to two classes (Kermany and Goldbaum, 2018). The two classes in the data set refer to pneumonia and healthy people. The data set was obtained from x-ray images of children aged one and five years from Guangzhou Women's and Children's Medical Center. All chest x-ray images of the patients were taken as part of their routine clinical care. In order to perform accurate analysis of chest x-ray images, low quality or incomprehensible images were identified from all lung graphs and extracted from the data set and quality control of the images was ensured.

The diagnosis of the dataset images for the training of the artificial neural network system used in the deep learning model was first made by two specialist doctors. The data set was also checked by a third specialist to take into account any possible misconceptions of specialist physicians evaluating the data. Chest x-ray images of 5840 individuals are included in the data set. Of these images, 5216 were used to train the deep learning model and 624 were used to test the model (Kermany and Goldbaum, 2018). The deep learning models used will be trained with 89% of the image data set and thus will be able to diagnose pneumonia on the x-ray image. The trained models will then be tested on the remaining 11% of the same data set. In this way, the data will be evaluated by observing the success performance of the models in the classification and the success comparison of the models will be made. An exemplary chest x-ray image from the data set is given in Fig. 1.

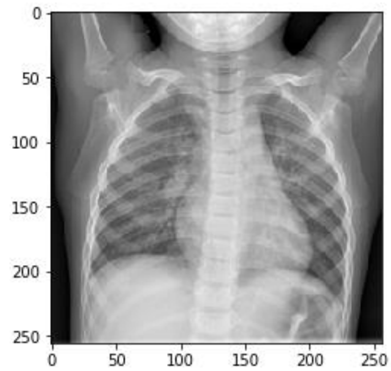


Fig. 1. A sample chest x-ray image in the data set ("Chest X-Ray Images", 2019).

## 2. DEEP LEARNING METHODS APPLIED

In this section, the dataset consisting of chest x-ray images of 5840 individuals into two categories as pneumonia and healthy individuals. Convolutional neural network, convolutional neural network with data augmentation and transfer learning deep learning models will be explained for classification of dataset of chest x-ray images.

### 2.1. Classification Using Convolutional Neural Network Model

In order to diagnose pneumonia from the image data, the convolutional neural network deep learning model was used first. In the design of the convolutional neural network model, keras library was used in the software. By applying deep learning methods to image data, it is necessary to introduce the images to computer systems first in the diagnosis of diseases with computer systems. Therefore, the images need to be converted to digital format that computer systems can detect.

All images are made up of digital units called pixels. Pixel is the name given to the smallest unit of the image in a digital environment. Many pixels come together to form images. Each pixel value is between 0 and 255 according to the image color tone. Images must be converted to digital matrix format in order to be perceived by computer systems. Each element of this matrix corresponds to one pixel. Digitization and pixel values of the images are shown in Fig. 2.

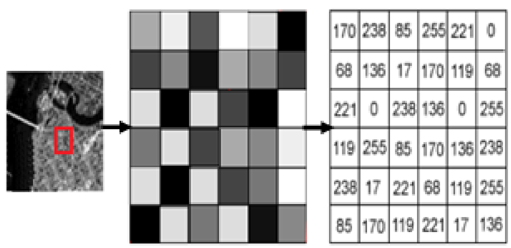


Fig. 2. Digitization and pixel values of an image ("Image Pixels", 2019).

In order to process image data in the applied model, image matrices (64x64) of the data set are obtained first. When classifying images in convoluted neural networks, it is necessary to extract the feature map which affects the class of an image via image matrices. These features may represent a simple shape such as a line or corner in the image, as well as more complex convex shapes. In the applied model, it determines the class of the test images in the test phase and determines the class of the related image depending on whether these features are detected in the image or not. Therefore, the resulting image matrices are first convolution in the convolution layer with filters of a certain size to extract the feature map from the images. In this way, it is provided to detect the distinctive parts in the images and to find the features that affect the classification.

While convolution is performed, image matrices are screened with filters. In the convolution process, the numerical values of the filter matrix to be applied in certain dimensions are multiplied by the numerical values of the portion of the image matrix up to the filter size starting from the beginning of the image matrix and these multiplication results are added. The new total value obtained; it becomes the new singular value on the image matrix, representing the corresponding area in the filter matrix size. This process is repeated by shifting the filter matrix over the image matrix with a certain step size. Thus, convolution is performed to filter the image matrix by filter. As a result of this process, the classifier properties of the image matrices are determined. The convolution layer and the convolution process are shown in Fig. 3.

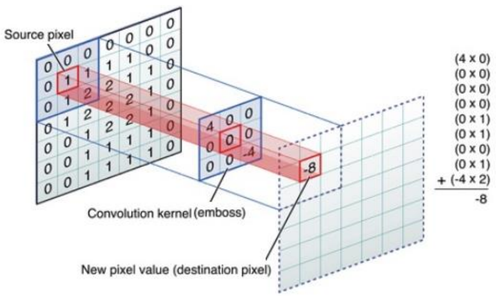


Fig. 3. Convolution layer and convolution process ("Convolution Operation", 2019).

In the convolutional neural network model used in the study, the obtained image matrices are transformed with 3x3 size filters in order to obtain classifying properties. In this way, it is provided to determine the desired parts in the images and find the desired features (Aghdam and Heravi, 2017). As a result of convolution, the size of the original image matrix decreases. Decreasing the image matrix size increases the processing speed of the models used and is advantageous in time. However, this reduction in size also leads to a certain level of information loss. As a result of the convolution process, the size of the image matrix after the process and the possible loss of information can be calculated with the following formulas: Eq. 1 and Eq. 2 (Krizhevsky *et al.*, 2012). Width of output matrix is defined by G.

$$G = \frac{W-F+2P}{S} + 1 \tag{1}$$

Where W: means the width of the image matrix, F: the width of the filter matrix used, P: the padding element depending on the padding technique used, and S indicates how many units the filter advances on the image matrix at each step in the convolution process. Height of the output matrix is defined by Y.

$$Y = \frac{H-F+2P}{S} + 1 \tag{2}$$

Where H: means the height of the image matrix, F: the height of the filter matrix used, P: the padding element depending on the padding technique used, and S: how many units the filter advances on the image matrix in each step in the convolution process.

Using these formulas, the amount of reduction in the size of the image matrix in each convolution layer can be calculated. In order to prevent such loss of information due to the decrease in the size of the image matrix, padding technique can be utilized in such models. When this technique is applied, the size of the image matrices presented as input to the convolution process does not change and a matrix of the same size is obtained as the result of the process. This prevents the loss of information from the image data as a result of convolution. In the study, the same padding technique, which is one of the mentioned padding methods, was used. In the same padding technique, before the convolution is applied to the image matrix, a frame consisting of 0 elements is added around this matrix and after this addition, convolution is performed. As a result of the convolution process applied, both the size of the image matrix is preserved and the loss of information especially in the pixels close to the corners of this matrix is prevented. The same padding technique is shown in Fig. 4.

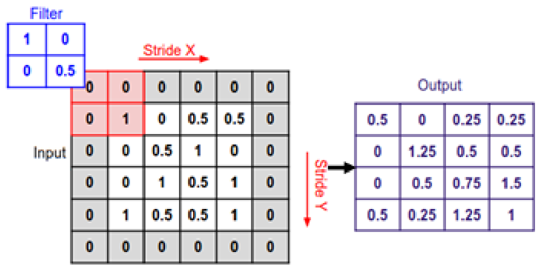


Fig. 4. The same padding technique ("Same Padding", 2019).

The number of filters used in the convolution layer can vary depending on the designer and the images. After convolution, the image matrices in the convolution layer are included in an activation function. In this study, relu activation function was used at this stage. The main reason for using the relu activation function in the model is its complexity-increasing structure. With the use of this activation function, convoluted neural networks perform better in detecting complex nonlinear shapes during the determination of classifying properties in images. When negative values are given as input to Relu activation function, the function gives 0 as the output and if positive values are given, the function gives the same value as the output. Therefore, these positive outputs produced with the same magnitude versus positive inputs enable the function to show better results in complex and non-linear situations. The relu activation function is shown in Fig. 5.

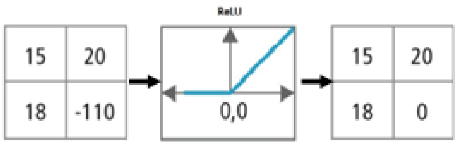


Fig. 5. The relu activation function ("Relu Activation Function", 2019).

The convolution layer is then transferred to the image matrices by the model to the pooling layer. In this layer, the number of pixels is reduced by reducing the image size (Krizhevsky *et al*., 2012). In this way, the number of parameters and operations used in the network is reduced, the performance of the network increases, and consequently the performance and speed of the model are increased and the processing time is reduced. Different pooling techniques such as max, min and avarage can be applied in the pooling layer. In the max pooling technique, the image matrix is divided into areas of a certain size. Then, the pixels with the highest numerical value are selected from the pixels in these fields, and this field is now expressed with this single pixel, thus reducing the number of pixels as the other pixels are disabled. The same process applies to all areas separated in the image matrix. In another pooling technique, min pooling, the image matrix is similarly divided into areas of the same size, but the pixel with the lowest numerical value is selected to represent the areas. In the Avarage pooling technique, after the image matrix is divided into areas of the same size, the pixel values in each area are averaged and the area is represented by this single average numerical pixel value. Max pooling technique is shown in Fig. 6.

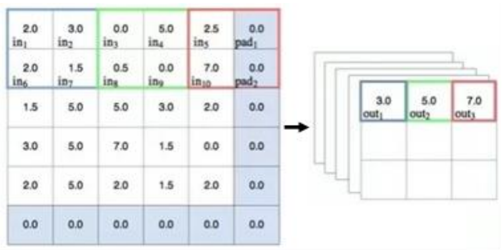

Fig. 6. Max pooling technique ("Max Pooling", 2019).

In this study, max pooling technique was used in the model among these techniques. In this way, it is provided to increase the performance and speed of the model, while avoiding the problem of producing incorrect results based on memorization due to the over-learning of the model called overfitting (Goodfellow *et al*., 2016). After this step in the applied model, the image matrices are passed through a different convolution layer for the second time. In this way, the possibility of detecting the features of classifiers in the images that are effective for the determination of classes increases. This enables the image classes of the model to be detected more accurately. Therefore, the second convolution layer applied in the model is included in the model architecture in order to improve the classification performance of the model.

After this step, the image matrices are transferred back to the pooling layer for the second time. In this way, the performance of the model is increased and especially the "overfitting" problem is prevented. In this pooling layer, max pooling technique was used similar to the first pooling layer, and the highest pixel values of the areas allocated in the pool size were selected to represent the relevant areas in the image. In the model, the image matrix data is transferred to a different convolution layer after the 2nd pooling layer for the third and last time. With the first two convolution layers applied, low and medium level classifying properties were determined which were effective in determining the classes in the image data. However, in order to increase the classification performance and obtain a higher percentage of accurate classification, it is necessary to determine the more complex high-level classifier properties. Consequently, the third convolution layer applied at this stage is included in the model to determine the more complex high-level classifier properties.

In this layer, as in the first two convolution layers, the same padding technique was used to preserve the size of the image matrix and prevent loss of information. Also, similar to the first two convolution layers, relu activation function is used in this layer and thus the probability of detecting nonlinear features in the images is increased. After this stage, the image data is transferred to a different pooling layer for the third and last time. In this layer, max pooling technique was applied similar to the first two pooling layers, thus reducing the processing time and further reducing the possibility of "overfitting" problem. Therefore, since the image matrix data is transferred to the model as input, a total of 3 convolutions and 3 times are included in the pooling layer.

32 filters were used in the first convolution layer and 64 filters were used in the second and third convolution layers. After all these steps, the processed image matrix data with classifier properties are transferred to the fully connected layer by the model used. Fully coupled layer is the section where image classification is performed depending on the presence or absence of classifier properties in image matrices. Therefore, the diagnosis of the test images in the data set will be performed in this section. The model performs the classification process and therefore the diagnosis by means of an artificial neural network in this layer. The general structure of the artificial neural network in the fully connected layer is shown in Fig. 7.

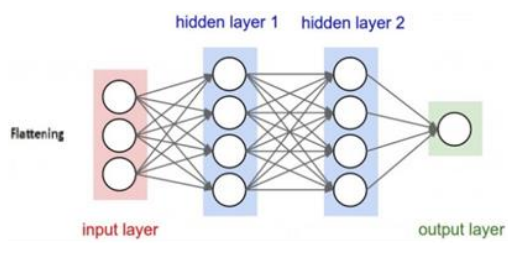


Fig. 7. The general structure of the artificial neural network in the fully ("Fully Connected Layer", 2019).

However, the artificial neural network in the layer cannot process the image data in matrix format. Therefore, said image matrix data must be supplied to the artificial neural network in vector format. Therefore, in this layer, the image matrices are first converted to a vector by flattening operation so that the data of the images can be input to the artificial neural network in this layer. The flattening process is shown in Fig. 8.

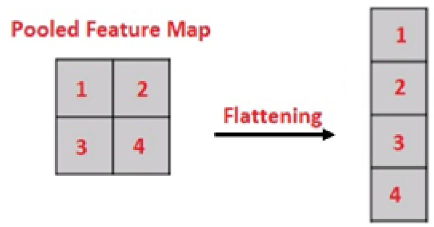


Fig. 8. The flattening process ("Flattening", 2019).

The vectors carrying the generated image data are transferred to an artificial neural network with 512 neurons. Depending on the input training image data, the weights of the artificial neural network are rearranged in an iterative manner and as a result, it gains the ability to classify the artificial neural network image data in this layer and thus diagnose the disease. The trained model is then tested over the image data reserved for the test stage in the data set and the classification accuracy percentage and performance of the model is tested. In this stage, artificial "sigmoid" function was used as the activation function in the artificial neural network. Sigmoid function is an activation function that can be used in neural networks in two class classification problems.

In this study, sigmoid function, which is more suitable for 2-class classification problems, was preferred at this stage, since there were a total of 2 classes, sick and healthy, in the classification and diagnosis of image data. The sigmoid function is shown in Fig. 9.

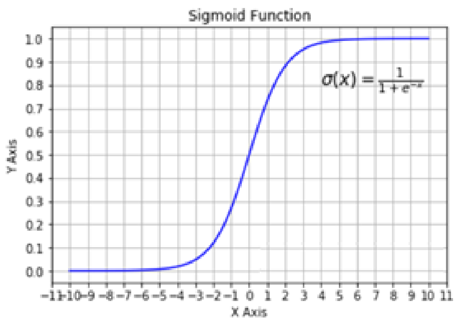


Fig. 9. The sigmoid function ("Sigmoid Function", 2019).

In addition, overfitting problem can be seen in this type of neural network as described previously. In this problem, depending on the excessive learning of the model by the model, the model outputs the results of incorrect classification based on the memorization. Therefore, as an additional step to prevent this problem, dropout technique is also utilized in this layer. Dropout technique is a regulation method used to train artificial neural network in this layer (Srivastava *et al.*, 2014). With this technique, half of the neurons are disabled in different iterations with different possibilities during data processing in the artificial neural network. Thus, in each iteration, a certain number of weights are changed instead of all the weights of the artificial neural network in this layer. Thus, the effect of any image input to the model on the weights of the neural network is limited, which prevents the network from performing misclassifications due to over-learning by heart. Dropout technique is shown in Fig. 10.

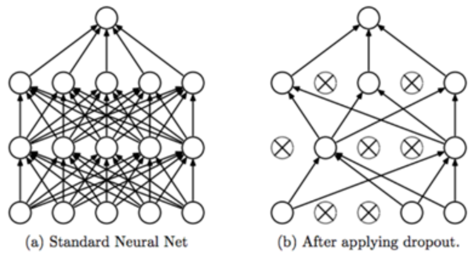


Fig. 10. Dropout technique ("Dropout", 2019).

The first model used is shown in Fig. 11.

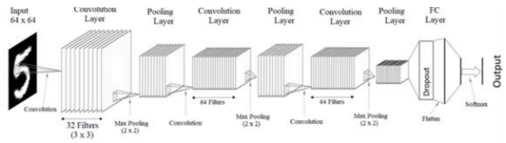


Fig. 11. The first model used ("First Model", 2019).

Therefore, the model described in this stage is the first of the three different deep learning models used in the study, and the results of the diagnostic performance obtained from the model and the correct classification percentage of the test data of the model are included in the results section.

## 2.2. Classification Using Convolutional Neural Network with Data Augmentation Model

The second deep learning model used to diagnose pneumonia from image data is a convective neural network model similar to the first deep learning model used. However, in this second model, data augmentation technique was applied unlike the first model. In the data augmentation technique applied to the model, the images in the image data set are given to the model both in their original format and through some changes. The changes are made by displaying images at different angles and at different distances (zooms) to the model (Wong *et al.*, 2016). As the model re-evaluates the same images from different angles and different distances, the possibility of identifying the classifying features in the image in question is increased. Therefore, with the higher probability of detecting these features, the model can learn and analyze the problem better via image data, and consequently, the classification performance of the model increases.

This technique is applied for training data to be transferred as input to the model. The testing phase of the model is performed only on the test data as in the first model used. Therefore, with this technique, it is aimed to train the model more successfully by transferring the training data to the model in different formats. Data augmentation is shown in Fig. 12.

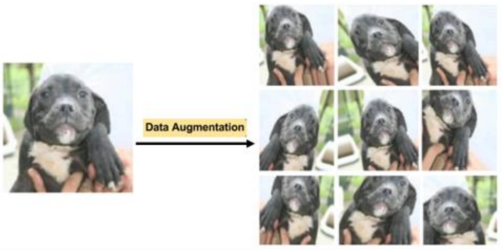


Fig. 12. Data augmentation technique ("Data Augmentation", 2019).

In the data set used in the study, 5216 images are used to train the model. This image data is transferred to the model in 32 size packages while training the model according to the software created. Therefore, with the transfer of images to 163 models performed in 1 cycle, all 5216 training images can be transferred to the model. However, in the software prepared for analysis, the model receives 200 of these 32 packets as input in 1 cycle. Therefore, the model takes a total of 6400 images as input in 1 cycle. Of the 6400 images taken as input of the model, 5216 are the original data set images, while the remaining 1184 images are the images created by the software using the data augmentation technique. The said 1184 images are created by replacing the original images with the software used.

Changes to images can be described as: First, images are rotated randomly to the right at a value between 0-30 degrees. Second, images are rotated randomly to the left at a value between 0-30 degrees. Third, the images are enlarged by 30% (Zoom in). Fourth, images are reduced by 30% (Zoom out). Fifth, images are rotated 90 degrees to the right. Sixth, images are rotated 90 degrees to the left.

The images that have been modified above are also used during the training of the model. The changes that may be made in the data are not limited to this and may vary according to the preferences of the software developer. Therefore, the rotation angle, direction and magnification ratios of the images can also be arranged in different ways. However, these parameters were adjusted as specified, considering the operating performance and processing time of the model. In addition, normalization pretreatment is applied to all of the image data in question. With this pre-treatment, the value range of all pixels of the images with pixels in the 0-255 value range is changed and this range is set to the 0-1 value range. Thus, the classification performance of the model is improved by filtering the defective pixels in the images. Apart from this technique, the model used in this stage is the same as the convolutional neural network model described in the first stage. Therefore, this model consists of a total of 7 layers, 3 convolution, 3 pooling and 1 fully connected layer. Also, as in the first stage model, 32 filters were used in the first convolution layer and 64 filters were used in the second and third convolution layers. The pooling, padding and dropout techniques used in the first model are also applied in this model. The percentage of classification accuracy and performance evaluation obtained from the model is given in the results section. The second model used is shown in Fig. 13.

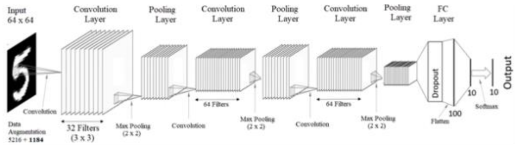


Fig. 13. The second model used ("Second Model", 2019).

## 2.3. Classification Using Transfer Learning Model

Transfer learning technique was used in the formation of the third model to be used for the diagnosis of pneumonia from the image data. The basis of this technique is based on the idea of storing the information previously obtained while solving a problem and then applying that information to a different but similar problem. Therefore, in this technique, the weights of an artificial neural network model which has been previously trained over a large data set for solving a problem and which is known to be successful in solving this problem are transferred to a different model in order to solve a similar problem. The output layer of the new model created using this technique is rearranged in accordance with the new problem to be solved, and after these arrangements, the model can now be used to solve the new problem. The transfer learning method is shown in Fig. 14.

In the present study, weights of the VGG16 neural network, previously trained on a 1000-class data set, will be used to classify lung x-ray image data (Shin *et al*., 2016). The output layer of the VGG16 model has been rearranged and this new model has been adapted to this 2 class problem. The VGG16 model is basically a simple network model and the difference is that the convolution layers are used in 2s and 3s. There are 4096 neurons (artificial neural network cells) in the full coupling layer of this model. This network has been developed for the solution of a classification problem of 1000 classes as previously mentioned, and the model calculates approximately 138 million parameters.

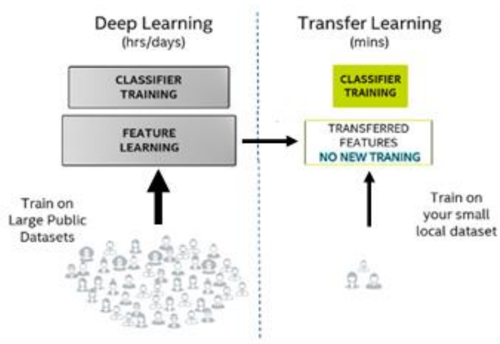


Fig. 14. The transfer learning method ("Transfer Learning", 2019).

As the depth of the model increases, the size of the image matrices decreases from the input layer to the output layer, while the depth value, ie the number of channels, increases. The filters with different weights are calculated at each convolution layer output of the model, and as the number of layers increases, the classifying properties in the filters represent the depths of the image.

The VGG16 model architecture is shown in Fig. 15.

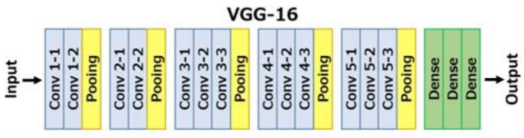


Fig. 15. The VGG16 model architecture ("VGG16 Model", 2019).

The most important advantage of this technique and therefore of the model used is that this model has already been optimally trained by a multidimensional data set. Therefore, the feature map of the model is highly developed. In this way, the model can detect the differences on the images and according to the findings can be divided into classes. The classifier features of the VGG16 model are shown in Fig. 16. The classification results obtained from the model and evaluations of the model's performance in diagnosis are given in the results section.

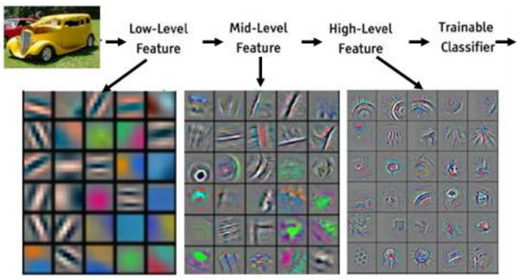


Fig. 16. The classifier features of the VGG16 model ("Features", 2019).

## 3. PERFORMANCE CALCULATIONS OF DEEP LEARNING METHODS APPLIED

In comparison of three different deep learning models with two different performance values, 5216 of chest x-ray images in the data set were used to train the deep learning model and the remaining 624 were used to test the model. Three deep learning models used in the classification; loss and classification accuracy values were calculated with the software prepared in python environment. As a result of the calculation, graphs of performance values were created for deep learning models.

### 3.1. Calculation of Classification Performance Values of Convolutional Neural Network Model

The classification performance values of the convolutional neural network model: loss and classification accuracy values were calculated with the software prepared in python environment. Graphs of the values obtained as a result of the calculation are given.

#### *3.1.1. Loss Performance Values*

The graph of loss performance values obtained as a result of the application of convolutional neural network model is shown in Fig. 17.

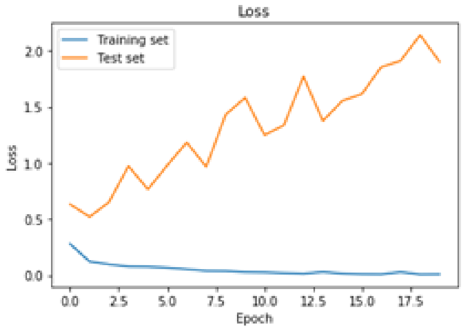


Fig. 17. The graph of loss performance values obtained as a result of the application of convolutional neural network model.

#### *3.1.2. Classification Accuracy Values*

The graph of the classification accuracy values obtained from the application of convolutional neural network model is shown in Fig. 18.

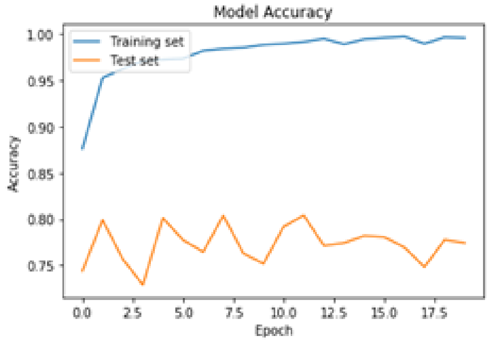


Fig. 18. The graph of the classification accuracy values obtained from the application of convolutional neural network model.

#### *3.1.3. Performance Evaluation of Model*

It is seen that the percentage of correct classification of test data in the test stage of the model trained with this data set is 75-80%. The best classification performance of the model is 80.4% accuracy at 12th iteration. It can be said that the CNN model used in this stage shows a successful classification performance and is suitable for diagnosis purposes, but it is seen that the same model gives better results when a "data augmentation" technique is used in the next stage. The confusion matrix results obtained from the classification task performed are given in Fig. 19.

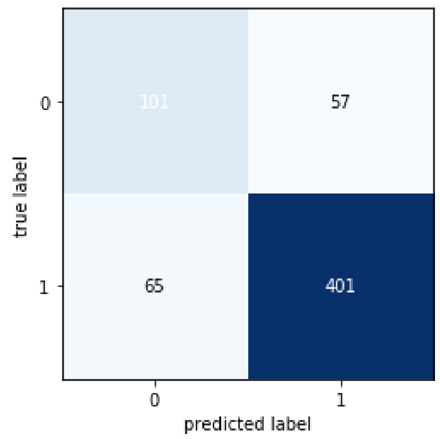


Fig. 19. Confusion matrix results of the first model.

When the results are examined, it is seen that the model misclassifies 122 of 624 test samples and the remaining samples correctly classify.

## 3.2. Calculation of Classification Performance Values of Convolutional Neural Network with Data Augmentation Technique

The classification performance values of the convolutional neural network model applied by data augmentation technique: loss and classification accuracy values were calculated with the software prepared in python environment. Graphs of the values obtained as a result of the calculation are given.

### *3.2.1. Loss Performance Values*

The graph of the loss performance values obtained as a result of the application of the convolutional neural network model applied by data augmentation technique is shown in Fig. 20.

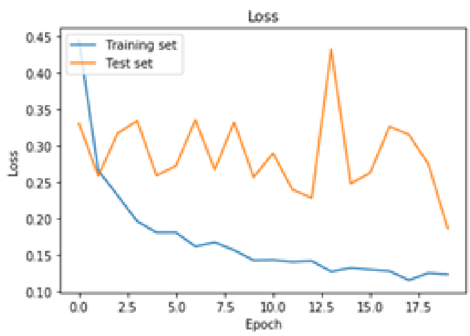


Fig. 20. The graph of the loss performance values obtained as a result of the application of the convolutional neural network model.

### *3.2.2. Classification Accuracy Values*

The graph of classification accuracy performance values obtained as a result of the application of convolutional neural network model applied by data augmentation technique is shown in Fig. 21.

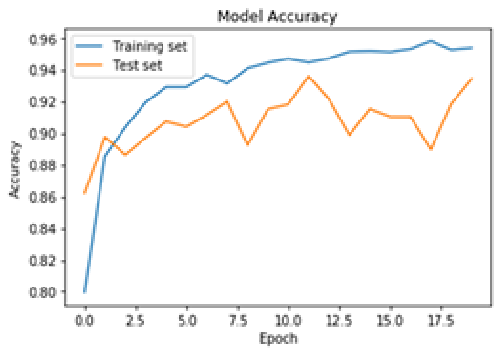


Fig. 21. The graph of classification accuracy performance values obtained as a result of the application of convolutional neural network model.

### *3.2.3. Performance Evaluation of Model*

When the results obtained from the model are examined, it is seen that the model has reached the lowest loss (total error) value in the 20th iteration and this value is 0.1864. It is seen that the percentage of correct classification of test data in the test stage of the model trained with this data set is 89-93%. The best classification performance of the model is 93.4% accuracy at the 20thiteration. Therefore, the model was able to accurately predict the class of 93.4% of the test data. At this stage, it can be seen that the data augmentation technique used in convolutional neural network model increases the model success and classification performance and when these results are compared with the results obtained in the previous stage.

The data augmentation technique is better, the model was able to extract the classifying features of the disease more successfully from the changing images (in terms of size, angle and proximity). Thus, the model showed a higher classification performance compared to the previous stage. It can be seen in the results that the model created here is suitable for use at the point of diagnosis on images. The confusion matrix results obtained from the classification task performed are given in Fig. 22**.**

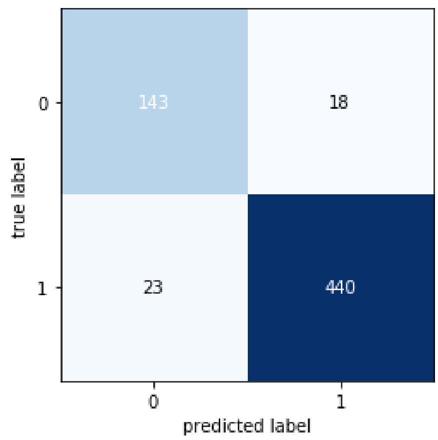


Fig. 22. Confusion matrix results of the second model.

When the results are examined, it is seen that the model misclassifies 41 of 624 test samples and the remaining samples are classified correctly.

### 3.3. Calculation of Classification Performance Values of Transfer Learning Model

The classification performance values of loss and classification accuracy values of the classification model created by transfer learning technique were calculated with the software prepared in python environment. Graphs of the values obtained as a result of the calculation are given.

#### *3.3.1. Loss Performance Values*

The graph of the loss performance values obtained as a result of the application of the transfer learning model is shown in Fig. 23.

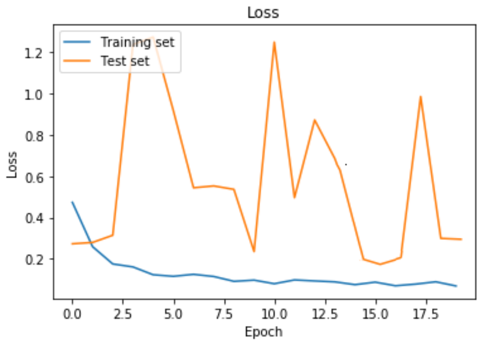


Fig. 23. The graph of the loss performance values obtained as a result of the application of the transfer learning model.

#### *3.3.2. Classification Accuracy Values of Model*

The graph of the classification accuracy values obtained from the application of transfer learning model is shown in Fig. 24.

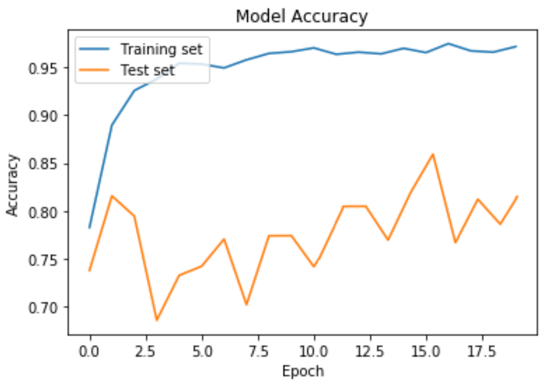


Fig. 24. The graph of classification accuracy performance values obtained as a result of the application of transfer learning model.

#### *3.3.3. Performance Evaluation of Model*

When the results obtained from the model are examined, it is seen that the lowest loss value reached by the model is 0.17. It is seen that the correct classification percentage of the test data is in the range of 70-82% at the test stage of the model trained with this data set. The best classification performance of the model was calculated as 85.6% accuracy. The confusion matrix results obtained from the classification task performed are given in Fig. 25.

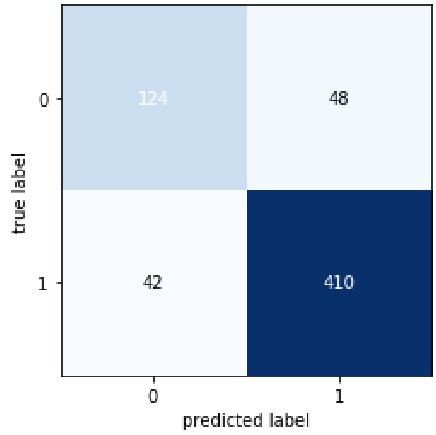


Fig. 25. Confusion matrix results of the third model.

### 3.4. Samples Pneumonia Detection in Chest X-ray Images

In order to examine the classification estimation performances of the models used in this study, experiments were performed with sample images. The described models take a chest x-ray image as input and give the class possibilities of this image as output. From the outputs of the models, information of the likelihood of a sample being pneumonia or healthy can be obtained. Sample experiments were carried out with the created models. Two sample experiments using convolutional neural network with data augmentation model are given. The output result obtained from the model for the chest x-ray image of a person known to have pneumonia is shown in Fig. 26.

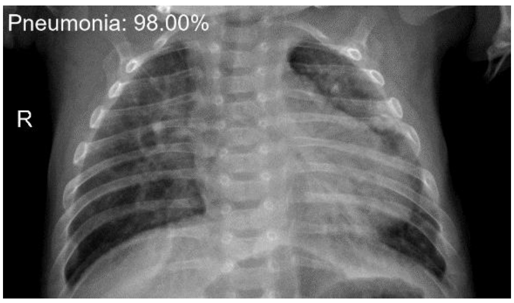


Fig. 26. The output result obtained from the model for the chest x-ray image of a person known to have pneumonia.

According to the output result, pneumonia was detected in this person's image with a probability of 98% and the estimate was correct. In another example, the output of the model for a chest x-ray image of a person known to be healthy is shown in Fig. 27.

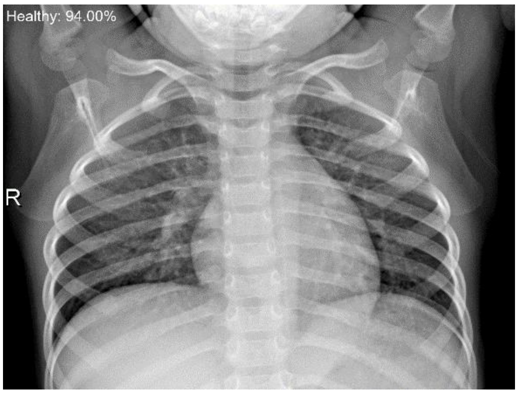


Fig. 27. The output of the model for a chest x-ray image of a person known to be healthy.

According to the output, this person was estimated to be 94% likely to be healthy, and this estimate is also true. In the examples, the model was able to accurately predict the classes of two different image samples. The model adds the forecast class label and probability information to the images as output.

## 4. CONCLUSIONS AND SUGGESTIONS

When the results of the models applied in the study were examined, it was observed that the correct classification rate of the convolutional neural network model, which is the first model, reached 80.4%. On the other hand, the second neural network model, which was created by using data augmentation technique on this model, showed that the correct classification rate of the test data reached up to 93.4%. Therefore, it is understood that applied data enhancement technique increases the classification performance of the model. The last model applied in the study was obtained by using the structure of VGG16 model, which has been obtained by training with a data set of 1000 classes, by using feature learning transfer technique. The output layer of this model is arranged in accordance with the two-class classification problem in the study. According to the results, the model succeeded in classifying the test data with an accuracy rate of 85.6%. All performance value results obtained from the models are shown in Table 1.

Table 1. All performance value results obtained from the models

| Models | Accuracy | Loss |
|---|---|---|
| CNN | 80.4 | 0.7645 |
| CNN with Data Augmentation | 93.4 | 0.1864 |
| Transfer Learning | 85.6 | 0.3984 |

Considering the performance of three different deep learning models at the point of classification of the image data in the data set, the most successful classification results were obtained with convolutional neural network model with data increase. Therefore, it is understood from the results that data augmentation and dropout techniques used in convective neural networks have a positive effect on the classification diagnostic performance of the model. In addition, these techniques prevent the problem of overfitting in the deep learning models, which causes the model to produce incorrect results by heart.

## ACKNOWLEDGEMENTS

We would like to thank anonymous refrees for their revision ideas that add depth to the article.

## REFERENCES

Aghdam, H. H. and Heravi, E. J. (2017). Guide to Convolutional Neural Networks, NY: Springer, New York, USA

Chest X-Ray Images, https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia [Accessed 20 July 2019].

Convolution Operation, https://medium.com/@bdhuma/6-basic-things-to-know-about-convolution-daef5e1bc411 [Accessed 21 July 2019].

Data Augmentation, https://developers.google.com/machine-learning/practica/image-classification/preventing-overfitting [Accessed 26 July 2019].

Dropout, https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5 [Accessed 26 July 2019].

Features, https://medium.com/abraia/getting-started-with-image-recognition-and-convolutional-neural-networks-in-5-minutes-28c1dfdd401 [Accessed 29 July 2019].

First Model, http://fourier.eng.hmc.edu/e176/lectures/ch10/node8.html [Accessed 27 July 2019].

Flattening, https://www.kaggle.com/kanncaa1/convolutional-neural-network-cnn-tutorial/notebook [Accessed 25 July 2019].

Fully Connected Layer, http://cs231n.github.io/convolutional-networks/ [Accessed 22 July 2019].

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*, MIT press, Massachusetts, USA.

Image Pixels, https://ai.stanford.edu/~syyeung/cvweb/tutorial1.html [Accessed 20 July 2019].

Indraswari, R., Kurita, T., Arifin, A. Z., Suciati, N. and Astuti, E. R. (2019). "Multi-projection deep learning network for segmentation of 3D medical images." *Pattern Recognition Letters*, Vol. 125, pp. 791-797.

Kermany, D. K., & Goldbaum, M. (2018). Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification. Mendeley Data, 2.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks." *Neural Information Processing Systems 2012*, NIPS, Lake Tahoe, Nevada, USA, pp. 1097-1105.

Li, Y., Zhang, H., Bermudez, C., Chen, Y., Landman, B. A. and Vorobeychik, Y. (2019). "Anatomical context protects deep learning from adversarial perturbations in medical imaging." *Neurocomputing*.

Lundervold, A. S. and Lundervold, A. (2019). "An overview of deep learning in medical imaging focusing on MRI." *Zeitschrift für Medizinische Physik*, Vol. 29, No. 2, pp. 102-127.

Max Pooling, http://adventuresinmachinelearning.com/convolutional-neural-networks-tutorial-tensorflow/[Accessed 22 July 2019].

Relu Activation Function, https://www.kaggle.com/kanncaa1/convolutional-neural-network-cnn-tutorial/notebook [Accessed 25 July 2019].

Roth, H. R., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., Kim, L. And Summers, R. M. (2015). "Improving computer-aided detection using convolutional neural networks and random view aggregation." *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, pp. 1170-1181.

Same Padding, https://medium.com/@ayeshmanthaperera/what-is-padding-in-cnns-71b21fb0dd7 [Accessed 21 July 2019].

Second Model, http://fourier.eng.hmc.edu/e176/lectures/ch10/node8.html [Accessed 27 July 2019].

Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., Van Riel, S. J., Wille, M. M. W., Naqibullah, M., Sanchez, C. I. and van Ginneken, B. (2016). "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks." *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, pp. 1160-1169.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I. and Summers, R. M. (2016). "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." *IEEE transactions on medical imaging*, Vol. 35, No. 5, pp. 1285-1298.

Sigmoid Function, http://buyukveri.firat.edu.tr/2018/04/16/derin-ogrenme-yapay-sinir-aglari-3/ [Accessed 26 July 2019].

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research*, Vol. 15, No 1, pp. 1929-1958.

Şengür, A., Akılotu, B. N., Tuncer, S. A., Kadiroğlu, Z., Yavuzkılıç, S., Budak, Ü. and Deniz, E. (2018). "Optic disc determination in retinal images with deep features." *Proc., 2018 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, Turkey, pp. 1-4.

Toraman, S., Tuncer, S. A. and Balgetir, F. (2019). "Is it possible to detect cerebral dominance via EEG signals by using deep learning?" *Medical hypotheses*, Vol. 131, 109315.

Transfer Learning, https://mlconf.com/blog/use-transfer-learning-for-efficient-deep-learning-training/ [Accessed 28 July 2019].

Tuncer, S. A., Akılotu, B. and Toraman, S. (2019). "A deep learning-based decision support system for diagnosis of OSAS using PTT signals." *Medical hypotheses*, Vol. 127, pp. 15-22.

Van Ginneken, B., Setio, A. A., Jacobs, C. and Ciompi, F. (2015). "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans." *Proc., 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, New York, USA, pp. 286-289.

VGG16 Model, https://neurohive.io/en/popular-networks/vgg16/ [Accessed 28 July 2019].

Wang, C., Shi, J., Zhang, Q. and Ying, S. (2017). "Histopathological image classification with bilinear convolutional neural networks." *Proc., 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, South Korea, pp. 4050-4053.

Wong, S. C., Gatt, A., Stamatescu, V., and McDonnell, M. D. (2016). "Understanding data augmentation for classification: when to warp?." *Proc., 2016 international conference on digital image computing: techniques and applications (DICTA)*, IEEE, Canberra, Australia, pp.1-6.

Zhang, J., Xie, Y., Wu, Q. and Xia, Y. (2019). "Medical image classification using synergic deep learning." *Medical Image Analysis*, Vol. 54, pp. 10-19.

Zhu, R., Zhang, R. and Xue, D. (2015). "Lesion detection of endoscopy images based on convolutional neural network features." *Proc., 2015 8th International Congress on Image and Signal Processing (CISP)*, Shenyang, China, pp. 372-376.