A deep learning approach for detecting pneumonia in chest X-rays

M. Emin Sahin[1], Hasan Ulutas[1*], Esra Yuce[1]

[1*] Yozgat University, Faculty of Engineering and Architecture, Departmant of Computer Engineering, Yozgat, Turkey, (ORCID: 0000-0001-7729-990X), <u>emin.sahin@bozok.edu.tr</u>

[1*] Yozgat University, Faculty of Engineering and Architecture, Departmant of Computer Engineering, Yozgat, Turkey, (ORCID: 0000-0003-3922-934X), <u>hasan.ulutas@bozok.edu.tr</u>

[1] Yozgat University, Faculty of Engineering and Architecture, Departmant of Computer Engineering, Yozgat, Turkey, (ORCID: 0000-0002-9522-8352), 16008117051@ogr.bozok.edu.tr

Abstract

Pneumonia causes the death of many children every year and constitutes a certain proportion of the world population. Chest X-rays are primarily used to diagnose this disease, but even for a trained radiologist, chest X-rays are not easy to interpret. In this study, a model for pneumonia detection trained on digital chest X-ray images is presented to assist radiologists in their decision-making processes. The study is carried out on the Phyton platform by using deep learning models, which have been widely preferred recently. In this study, a deep learning framework for pneumonia classification with four different CNN models is proposed. Three of them are pre-trained models, MobileNet, ResNet and AlexNet and the other is the recommended CNN Model. These models are evaluated by comparing them with each other according to their performance. The experimental performance of the proposed deep learning framework is evaluated on the basis of precision, recall and F1-score. The models achieved accuracy values of 93%, 97%, 97% and 86%, respectively. It is clear that the proposed ResNet model achieves the highest results compared to the others.

Keywords: CNN, Deep Learning, Python, Pneumonia, X-ray

Göğüs röntgenlerinde pnömoniyi tespit etmek için derin öğrenme yaklaşımı

Öz

Pnömoni her yıl çok sayıda çocuğun ölümüne neden olmakta ve dünya nüfusunun belli bir oranını oluşturmaktadır. Göğüs röntgenleri öncelikle bu hastalığı teşhis etmek için kullanılır, ancak eğitimli bir radyolog için bile göğüs röntgenlerini yorumlamak kolay değildir. Bu çalışmada, radyologlara karar verme süreçlerinde yardımcı olmak için dijital göğüs röntgeni görüntüleri üzerinde eğitilmiş bir pnömoni tespiti modeli sunulmaktadır. Çalışma, Phyton platformunda son zamanlarda yaygın olarak tercih edilen derin öğrenme modelleri kullanılarak gerçekleştirilmiştir. Bu çalışmada, dört farklı CNN modeli ile pnömoni sınıflandırması için bir derin öğrenme çerçevesi önerilmiştir. Bunlardan üçü önceden eğitilmiş modeller, MobileNet, ResNet ve AlexNet, diğeri ise önerilen CNN modelidir. Bu modeller performanslarına göre birbirleriyle karşılaştırılarak değerlendirilmektedir. Önerilen derin öğrenme çerçevesinin deneysel performansı, kesinlik, duyarlılık ve F1-puanı temelinde değerlendirilir. Modeller sırasıyla %93, %97, %97 ve %86 doğruluk değerlerine ulaşmıştır. Önerilen ResNet modelinin diğerlerine kıyasla en yüksek sonuçları elde ettiği açıktır.

Anahtar Kelimeler: ANN, Derin Öğrenme, Python, Zatürre, X-Ray

* Corresponding Author: <u>hasan.ulutas@bozok.edu.tr</u>

1. Introduction

Infectious diseases are one of the most important factors that threaten human health from past to present. Pneumonia comes first among the infectious diseases [1]. Pneumonia is defined as lung inflammation that occurs as a result of the effects of living things such as viruses and bacteria on the microscopic air sacs [2], [3]. Every year, approximately 7% of the world population is affected by pneumonia, and approximately 4 million of these affected patients die [8]. Early diagnosis is very important in such diseases [4]. Typical symptoms include chest pain, shortness of breath, cough etc. takes place. Diagnostic tools include sputum culture and chest X-ray images [5].

Artificial intelligence, which has become popular in recent years, tries to make computers more useful by trying to understand the nature of human intelligence. Machine learning and deep learning, which are sub-branches of artificial intelligence used in many fields, also promise to help experts in the early diagnosis of diseases. CNN-based deep networks have been developed differently by researchers and these networks have been applied to classification, segmentation, object detection and localization, especially in computer vision [6]–[8]. Also, besides computer vision problems, CNNs are used for breast cancer detection [9], brain tumour segmentation [10], Alzheimer's disease diagnosis, skin lesion classification [11], [12] etc. Successful results have been observed by using it in the solution of medical problems.

The following tests can be performed for the diagnosis of pneumonia: Chest X-ray, Lung CT, Chest ultrasound, Lung needle biopsy Chest MRI [5]. Currently, a chest X-ray is one of the best methods for detecting pneumonia [6]. X-ray imaging is preferred over CT imaging, as CT imaging typically takes significantly more time than X-ray imaging, and many undeveloped areas lack adequate high-quality CT scanners. In contrast, X-rays are the most widely available diagnostic imaging technique. When the studies in the literature are examined, there are many studies, especially in the detection of pneumonia.

In the study given in [13], Kieu et al. used multiple CNN constructs to detect abnormal (abnormal) rib cage images. In this method, a rib cage image is divided into three parts as right, left and whole and used as inputs for different ESA models. Then the result of the CNN models is combined with the Fusion method. The applicability of the proposed multiple CNN method is shown with the result of 96% obtained as a result of the study.

In [14], Ivo M. Baltruschat et al., performed the classification of 14 different diseases in 112120 lung X-ray images provided by The National Institutes of Health (NIH) Clinical Center with different hyper parameters of the ResNet50 and Resnet38 models. ResNet38 and ResNet101 also conducted similar studies. Baltruschat has shown that ResNet models get the best results with ResNet38.

In [15], O'Quinn et al. They carried out their studies using 5659 images in DICOM format to detect pneumonia disease. In this study, 70% of the dataset is used for training and the rest for testing. By using the AlexNet deep learning network in learning transfer, 76% recognition success is achieved in this method. In [16], Wang et al. developed a deep CNN method and presented a study for the detection of 14 thoracic cage diseases called Thorax-Net. In this method, it is included in the classification and attention section. A result has been obtained by using the outputs of these two sections with Thorax-Net. They have shown that with the Thorax-net architecture, the area below the average ROC per class reached 0.7876.

In [17], Hammoudi et al. They have worked on deep learning models that could be used to diagnose pneumonia using radiographic images to assist healthcare professionals. ResNet34, ResNet50 and DenseNet169 have used different deep learning methods (Inception ResNetV2-RNN) such as VGG-19 and dual usage model. They attempted to classify X-ray images of the rib cage as viral pneumonia, bacterial pneumonia, and normal, and obtained different results for each model used.

The aim of this study is to provide radiologists and medical professionals with a low-cost software tool to review their interpretation of the disease in the diagnostic process and identify other potential findings that may be overlooked, using deep learning methods that have achieved great success in medical image classification of pneumonia, a disease for which early diagnosis is very important. Also, in many developing countries, people's access to radiologists is limited due to the high cost. With this software, healthcare professionals will be able to help interpret X-ray images. For this purpose, rib cage X-ray images shared on Kaggle, a community of online data scientists and machine learning practitioners, are used. In this dataset, there are a total of 5840 X-ray images, consisting of two classes, Normal and Pneumonia.

In this article, a study about the deep learning methods used in the field of health and the classification of pneumonia, which is the aim of this study in the literature, is presented.

2. Material and Method

2.1. Dataset

X-rays in datasets are labeled by their respective clinicians. The dimensions of the tagged view counts of the data set are specified in the method section. In this study, the publicly available Chest X-ray Images (Pneumonia) dataset is used. The dataset is compiled into two subfolders containing pneumonia and normal chest X-ray images, respectively. The total number of data is 5840, with 1575 normal and 4265 pneumonia cases. In Figure 1, there are images of pneumonia and normal classes from the database.
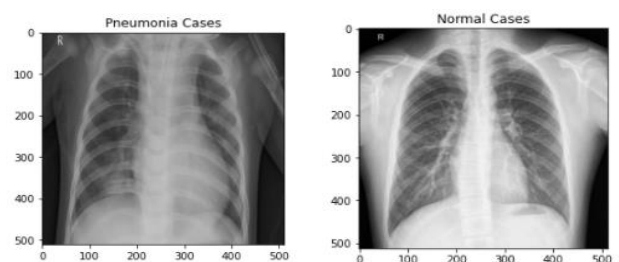

Fig. 1. Sample images from the dataset

For this study, a dataset containing X-ray images of the thorax shared on the Kaggle platform is used. Among these images, low quality or unreadable ones are eliminated, and the remaining images are rated by three experts in order to be used in artificial intelligence studies. Example X-ray images of pneumonia disease are shown in Figure 1.

2.2. Deep Learning

The term deep learning refers to multi-layer artificial neural networks (ANNs). Recently, deep learning has become one of the most powerful tools in the literature because it can process large amounts of data [18]. Convolutional Neural Networks (CNN) are one of the most well-known deep learning networks. Significant successes have been achieved in studies such as voice recognition, image processing and pattern recognition using CNN. CNN is a deep neural network in which feature extraction is performed by applying the convolution operation to at least one of its layers. One of the reasons why CNN is frequently used in studies is that it handles the features in the image independently of the location.

2.3. Convolutional Neural Network

In this section, some of the most well-known CNN models used in similar studies are discussed in order to compare the results of the study.
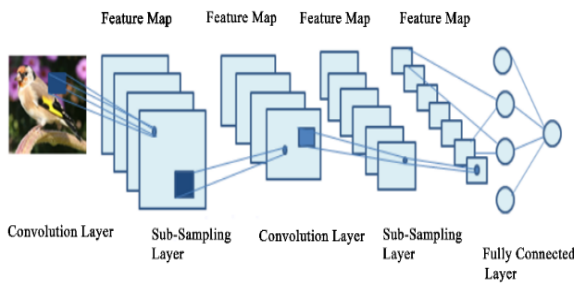

Fig. 2. Convolutional neural network structure

2.3.1. Convolutional Neural Network Layers

Input layer

The input layer forms the first layer of the CNN. In the input layer, the input data is given raw to the network. In order for the computer to understand the input (image), the pixels of the input are converted into numerical data. The size of the image in the input layer in a used network is important. Large input sizes will cause a lot of memory usage and increase the number of parameters, which may slow down the network, but may yield better results as more attributes can be extracted.

In addition, since the number of parameters will decrease with small input sizes, the network may give faster results, but the performance of the network may decrease as the feature extraction is less [19].

Convolution layer

The convolution layer is also known as the feature extraction layer because the features of the image supplied to the network are extracted in this layer. The feature map of the image is created by using various filters (kernel) in the convolution layer [3]. The filters used can be of different sizes such as 2×2, 3×3, 5×5. With these filters, the convolution process is applied to the image. The coefficients of the filters change with each learning iteration in the training set during the training of the CNNs to determine which regions of the data are important in determining the attributes of the network [19].

The pooling layer

The pooling or pooling layer is generally used in CNNs to reduce the number of parameters and the computational load. It reduces the input size for the next convolution layer without affecting the depth. That is, while the height and width values of the input data decrease after pooling, the depth, that is, the number of channels, does not change. In general, two different pooling layers can be mentioned.

Activation layer

The activation layer comes after the convolution layer. Generally, this non-linear layer is used to adjust and limit the output produced [18]. The purpose of use of this layer is to reduce the negative values in the linear network formed after the convolution layer to zero, to put the network into a non-linear form and to enable the network to learn faster. It is also called the activation function. There are many different activation functions such as tanh, sigmoid, ELU, but ReLU is the most widely used.

Fully connected layer

After the successive layers of convolution, ReLU, and pooling in an CNN comes the full link layer. For the full link layer, the feature map is converted to a 1D feature vector, this is called flattening [20]. This layer is called the full link layer because it is connected to all areas of the previous layer and it is shown in Figure 3. The full link layer is used to classify images in different categories.
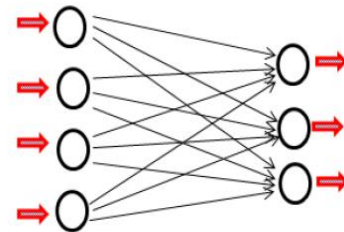

Fig. 3. Fully Connected Layer

2.4 Convolutional Neural Network Models

In this section, some of the most well-known CNN models used in similar studies are discussed in order to compare the results of the studies. These architectures are deep learning models delivered with pre-trained weights on the ImageNet dataset. These models can be used for operations such as prediction, feature extraction and classification. A pre-trained image classification model that has learned to extract powerful and informative features from images can be taken and used as a starting point for learning a new task. Applications are made for three pre-trained models to be used for image classification. In this part, Alexnet, ResNet50 Architecture and MobileNet architecture are used.

Table 1. Pre-Trained Networks and Some Features

Number	Network	Deep	Parameter (million)	Input Image Size
1	MobileNetv2	53	3.5	224 x 224
2	ResNet50	50	25.6	224 x 224
3	AlexNet	8	61.0	224 x 224

Table 2. Parameters

		Predicted	
		Negative	**Positive**
Actual Cases	**Negative**	True Negatives (tn)	True Positives (tp)
	Positive	False Negatives (fn)	False Positives (fp)

The accuracy given in Equation 1 is a measure of how well the learning model is [21];

$$\text{Accuracy, } ACC = \frac{tn+tp}{tn+tp+fn+fp} \quad (1)$$

Equation 2, True Positive Rate (Recall, (TPR)) is a measure that defines how many of the positive predictions are true [21].

$$\text{Recall, } TPR = \frac{tp}{tp+fp} \quad (2)$$

Equation 3, Precision positive predictive value (PPV) is a measure that defines how many of the positive predictions are correct [21].

$$\text{Precision, } PPV = \frac{tp}{tp+fn} \quad (3)$$

3. Results and Discussion

Google Colaboratory workspace is used to diagnose pneumonia from X-ray images. Colab is a free Python development environment based on the Jupyter notebook and cloud environment. Classification for pneumonia diagnosis from X-ray images is performed using the Python programming language. Python is a programming language with both practical and easy documentation used by many researchers working with machine learning and data science. Many libraries have been developed thanks to the work in the field of deep learning. These libraries provide convenience for the easy implementation of algorithms. Libraries used within the scope of the study are NumPy and Matplotlib.
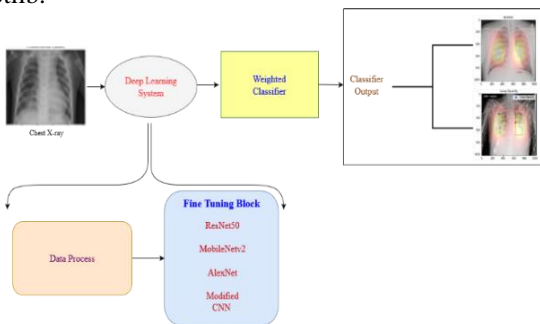

Fig. 4. Proposed system block diagram [22]

3.1. Datasets and their application to the model

CNN models take inputs of fixed size and calculate parameters accordingly. After reading the X-ray images in the dataset, they are resized to 512 x 512 to minimize the loss of details in the images. For training and testing, the data set is randomly divided into 80% (training) and 20% (testing). Of the total of 5840 data sets, 4672 are used for training and 1168 for testing.

3.2. Network Architecture

In this section, the proposed CNN model is classified with the X-ray dataset and its performance is observed. Machine Learning

algorithms create a mathematical model that associates the data in the data set with its labels. Thus, when an image is given to the model, it tries to determine which class the image belongs to. The model to be recommended is Keras' Sequential model. The training of the CNN model consisting of Sequential Convolution (CONV2D) and MaxPooling layers is carried out with the X-ray dataset. The layers of the proposed model are listed below.

Fig. 5. The used Model

The 'Adam' algorithm, which is thought to give good results in image data training, is used as the optimization algorithm, which has a great effect on model performance. The initial learning rate is very important for education. If it is selected large, it may cause it to skip the global minimum, or if it is selected too small, it may get stuck in local minimums and cannot reach the global minimum. Considering these issues, the learning rate is determined as 0.0001. The performance metric accuracy is selected.

In the designed model, the training data is trained 10 times (epoch) in 30 mini-batches (small groups). Then, testing is carried out in 15 mini-batches with the trained model test data. In the test of the model, an accuracy rate of 93% is obtained. The estimated results in classification are given in Figure 6 in the confusion matrix.

Fig. 6. The confusion matrix of the model

The difference in the number of data between the two classes in the training data may affect the model performance. The class with more numbers can suppress the other class. In the training data we used in the study, the normal class consists of 1260 and pneumonia class 3412 samples. In this case, it is not appropriate to only take accuracy as a performance metric. Apart from accuracy, performance metrics such as precision, recall, and F1-

score should also be considered. These metrics are included in the classification report given in Figure 7.

```
                precision    recall   f1-score   support

Pneumonia Cases    0.98       0.92      0.95        853
   Normal Cases    0.82       0.94      0.87        315

       accuracy                         0.93       1168
      macro avg    0.90       0.93      0.91       1168
   weighted avg    0.93       0.93      0.93       1168
```

Fig. 7. Model's classification report

```
                precision    recall   f1-score   support

Pneumonia Cases    0.97       0.99      0.98        853
   Normal Cases    0.97       0.92      0.95        315

       accuracy                         0.97       1168
      macro avg    0.97       0.96      0.96       1168
   weighted avg    0.97       0.97      0.97       1168

-------------------------------------------------
                 Karmaşıklık Matrisi
-------------------------------------------------
                Pneumonia Cases   Normal Cases
Pneumonia Cases       845              8
   Normal Cases        24            291
-------------------------------------------------
```

Fig. 8. Classification Report and Confusion Matrix for MobileNet model

```
                precision    recall   f1-score   support

Pneumonia Cases    0.98       0.99      0.98        853
   Normal Cases    0.96       0.94      0.95        315

       accuracy                         0.97       1168
      macro avg    0.97       0.96      0.97       1168
   weighted avg    0.97       0.97      0.97       1168

-------------------------------------------------
                 Karmaşıklık Matrisi
-------------------------------------------------
                Pneumonia Cases   Normal Cases
Pneumonia Cases       842             11
   Normal Cases        20            295
-------------------------------------------------
```

Fig. 9. Classification Report and Confusion Matrix for the ResNe50t model

```
                precision    recall   f1-score   support

Pneumonia Cases    0.99       0.82      0.90        853
   Normal Cases    0.67       0.97      0.79        315

       accuracy                         0.86       1168
      macro avg    0.83       0.90      0.85       1168
   weighted avg    0.90       0.86      0.87       1168

-------------------------------------------------
                 Karmaşıklık Matrisi
-------------------------------------------------
                Pneumonia Cases   Normal Cases
Pneumonia Cases       703            150
   Normal Cases        10            305
-------------------------------------------------
```

Fig. 10. Classification Report and Confusion Matrix of AlexNet model

The data set is also trained with pre-trained CNN models and their performance is observed. The pre-trained MobileNet, ResNet and AlexNet models are trained with the preprocessed dataset, respectively. According to classification report, 99%, 99% and 96% training accuracy is obtained from the models trained 10 times in 35 mini-batches, respectively. Trained models are tested with test data in 35 mini-batches. In the testing of the models, 97%, 97% and 86% accuracy rates are obtained. The predicted results in classification are given in each model's own confusion matrix and classification report given below Figures 8, 9 and 10.

4. Conclusions and Recommendations

The presence of specialist radiologists is the most important requirement to accurately diagnose pneumonia disease. This study primarily aims to improve medical competence in areas where radiologists are still limited and to facilitate early diagnosis of pneumonia in order to prevent negative consequences (death etc.). Developing algorithms in this area can be highly beneficial for providing better healthcare. At this point, different models such as the proposed CNN model and pre-trained MobileNet, ResNet and AlexNet are used to evaluate chest images of pneumonia patients. The use of these models on X-ray pictures, in particular, has generated promising results. Such a finding is particularly important in that X-ray images are easily accessible and cost-effective. In this paper, four distinct CNN models are proposed as part of a deep learning framework for pneumonia classification. Three of them are pre-trained models, MobileNet, ResNet and AlexNet and the other is the recommended CNN Model. These models are evaluated by comparing them with each other according to their performance. The experimental performance of the proposed deep learning framework is evaluated on the basis of precision, recall and f1-score. The models achieved accuracy values of 93%, 97%, 97% and 86%, respectively. It is clear that the used ResNet model achieves the highest results compared to the others.

5. Acknowledge

References

[1] Akter, S., & Shamsuzzaman, J. F. (2015). Community acquired pneumonia. Int J Respir Pulm Med, 2, 2.

[2] McLuckie, A. (Ed.). (2009). Respiratory disease and its management. Springer Science & Business Media.

[3] Pommerville, J. C. (2012). Alcamo's Fundamentals of Microbiology: Body systems edition. Jones & Bartlett Publishers.

[4] Summah, H., & Qu, J. M. (2009). Biomarkers: a definite plus in pneumonia. Mediators of inflammation, 2009.

[5] D. Berliner, N. Schneider, T. Welte, and J. Bauersachs, "The differential diagnosis of dyspnoea," Dtsch. Arztebl. Int., vol. 113, no. 49, pp. 834–844, 2016.

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

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[8] Ezzy, H., Charter, M., Bonfante, A., & Brook, A. (2021). How the Small Object Detection via Machine Learning and UAS-Based Remote-Sensing Imagery Can Support the

Achievement of SDG2: A Case Study of Vole Burrows. Remote Sensing, 13(16), 3191.

[9] Ragab, D. A., Sharkas, M., Marshall, S., & Ren, J. (2019). Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ, 7, e6201.

[10] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. IEEE transactions on medical imaging, 35(5), 1240-1251.

[11] Öztürk, Ş., & Özkaya, U. (2020). Skin lesion segmentation with improved convolutional neural network. Journal of digital imaging, 33(4), 958-970.

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

[13] Kieu, P. N., Tran, H. S., Le, T. H., Le, T., & Nguyen, T. T. (2018, November). Applying Multi-CNNs model for detecting abnormal problem on chest x-ray images. In 2018 10th International Conference on Knowledge and Systems Engineering (KSE) (pp. 300-305). IEEE.

[14] Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., & Saalbach, A. (2019). Comparison of deep learning approaches for multi-label chest X-ray classification. Scientific reports, 9(1), 1-10.

[15] O'Quinn, W., Haddad, R. J., & Moore, D. L. (2019, January). Pneumonia radiograph diagnosis utilizing deep learning network. In 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology (ICEICT) (pp. 763-767). IEEE.

[16] Wang, H., Jia, H., Lu, L., & Xia, Y. (2019). Thorax-net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography. IEEE journal of biomedical and health informatics, 24(2), 475-485.

[17] Hammoudi, K., Benhabiles, H., Melkemi, M., Dornaika, F., Arganda-Carreras, I., Collard, D., & Scherpereel, A. (2021). Deep learning on chest x-ray images to detect and evaluate pneumonia cases at the era of covid-19. Journal of Medical Systems, 45(7), 1-10.

[18] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET) (pp. 1-6). Ieee.

[19] Özkan, İ. N. İ. K., & Ülker, E. (2017). Derin öğrenme ve görüntü analizinde kullanılan derin öğrenme modelleri. Gaziosmanpaşa Bilimsel Araştırma Dergisi, 6(3), 85-104.

[20] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. Neurocomputing, 187, 27-48.

[21] Burkov, A. (2019). The hundred-page machine learning book (Vol. 1, pp. 3-5). Canada: Andriy Burkov.

[22] Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. Diagnostics, 10(6), 417.