Mathematical Biosciences and Engineering

MBE, Volume (Issue): Page.
DOI:
Received:
Revised:
Accepted:
Published:

http://www.aimspress.com/journal/MBE

Type of article

Performance Analysis of Deep Feature Extraction Method on COVID-19 Classificatioin Using Chest X-ray Image

First name Last name[1], First name Last name[2] and First name Last name[1,*]

[1] Department of Mathematics, Missouri State University, 901 S. National, Springfield, MO 65897, USA
[2] Affiliation

* **Correspondence:** Email address of corresponding author; Tel: +1-111-111-1111; Fax: +1-111-111-1111.

Abstract: Coronavirus disease 2019 or COVID-19 is an outbreak that was discovered at the end of 2019 in the province of Wuhan, China, which then spread throughout the world. Reverse transcription-polymerase chain reaction (RT-PCR) was then used as a method of diagnosing COVID-19. However, the RT-PCR method requires a long time in the diagnostic process, so the American College of Radiography (ACR) recommends the use of radiographic tools such as computed tomography scan (CT-Scan) and X-ray as additional methods in diagnosing COVID-19. X-ray was then chosen as an additional method in diagnosing COVID-19 because the tool used is more flexible and is already widespread in various health clinics. In this study, the author used a neural network approach, namely the Convolutional Neural Network (CNN) for the Deep Feature Extraction method and the machine learning approach for the classification method, in making a model that can classify normal lungs, lungs infected with COVID-19, and pneumonia based on X-ray images. The CNN architecture used in this study was ResNet-50 and the classifier used was Support Vector Machine (SVM), Random forest, K-Nearest Neighbor (KNN), and Extreme Gradient Boosting (XGBoost). The datasets used in this study were the COVID-19 Image Data Collection by J. P. Cohen, the ChestX-Ray8 Dataset by the National Institute of Health, and the Chest X-ray Dataset by Mendeley Data. The model was then trained using the CNN method with the ResNet-50 architecture. Furthermore, the fully connected layer in the ResNet-50 architecture was replaced using the SVM, Random forest, KNN, and XGBoost classifiers. Based on the simulation results, the best accuracy was obtained by a combination of ResNet-50 and SVM with 94.22%. The best recall was obtained by a combination of ResNet-50 and KNN with 94%. The best precision was obtained by ResNet-50 with 94.36%. The best running time was obtained by ResNet-50 with 0.0006 s.

Keywords: Covid-19, Pneumonia, machine learning, Deep Learning

1. Introduction

Coronavirus Disease 2019 or COVID-19 is an infectious outbreak caused by a virus from the Coronaviridae family, which was first discovered at the end of 2019 in Wuhan, China, which then spread throughout the world, including Indonesia. The symptoms caused by this virus are similar to those caused by the flu, severe acute respiratory syndrome (SARS), and the Middle East respiratory syndrome (MERS). In its development, COVID-19 infection has been confirmed by various countries through the reverse transcription polymerase chain reaction (RT-PCR) method, which only has a sensitivity of 30%-70% [1]. The American College of Radiography (ACR) then recommends using a computed tomography scan (CT-scan) and X-ray as an additional method in diagnosing COVID-19 because the diagnosis results are faster than that of RT-PCR. Although it has a higher sensitivity than RT-PCR [2], the use of CT scans in the detection of COVID-19 raises concerns about viral contamination in the CT scan equipment, which allows the risk of transmission of the COVID-19 virus to health workers. X-ray images were then used as an additional method because X-ray radiography equipment has low operating costs and is widespread in various health clinics. However, X-ray images in detecting COVID-19 have poor performance [3]. The poor sensitivity of the diagnosis of COVID-19 on X-ray images may be due to the lack of experience of radiologists in diagnosing COVID-19 coupled with the similarity of an X-ray image representation of COVID-19 patients with pneumonia patients, causing more significant challenges for radiologists [4]. Therefore, we need a system that can assist radiologists in classifying lungs infected with pneumonia and COVID-19 using X-ray image data.

Rapid developments in the world of technology directly affect the medical world, as evidenced by the many institutions that open medical data to the public, such as the Stanford University School of Medicine, which released the CheXpert Dataset, which contains more than 200,000 X-ray image data. This certainly opens opportunities for researchers to develop their models so that medical diagnosis can be faster and more accurate. Several previous studies have applied machine learning and deep learning methods for pneumonia and COVID-19 detection systems using X-ray image data. Some of the methods that have been used include the Deep Neural Network method on X-ray image data in detecting pneumonia [5], the Convolutional Neural Network (CNN) method on X-ray image data in detecting COVID-19 [6], the combination method of CNN and Support Vector Machine (SVM) on X-ray image data in detecting COVID-19 [7], and Convolutional Neural Network (CNN) method on X-ray image data and CT scan in detecting COVID-19 [8].

In this study, the authors propose to create a model that can classify normal lungs, lungs infected with pneumonia, and lungs infected with COVID-19 using X-ray image data and Deep Feature Extraction approach from the CNN model that will be classified by classic machine learning classifier. The classic machine learning classifier is one of the machine learning algorithms that is used to solve classification problems particularly on numerical data. Previous studies have succeeded in solving classification problems using methods such as Support Vector Machine (SVM) and Random Forest (RF) in detecting chronic kidney disease using gene expression data [9] and XGBoost in predicting diabetes using physical examination data from the National Institute of Diabetes and Digestive and Kidney Diseases [10]. CNN is one of the developments of the Neural Network, which is commonly used in image, text, and voice data processing [11]. Previous research has shown that the combination of CNN as Deep Feature Extraction and machine learning as a classifier can produce a model that outperforms the CNN model [12,13]. Based on this research, the author used the CNN model with the ResNet-50 architecture as a Deep Feature Extraction method and several classical classification methods such as SVM, RF, K-Nearest Neighbor (KNN), and Extreme Gradient Boosting (XGBoost) in classifying X-ray image data.

2. Materials and methods

2.1. Materials

The data used in this study were chest X-ray image data obtained from three different online databases, namely, the COVID-19 Image Data Collection by JP Cohen, the ChestX-Ray8 Dataset by the National Institute of Health, and the Chest X-ray Dataset by Mendeley. A radiologist annotated each lung X-ray image data from the three databases. Each image data has a description of the disease. Data with disease descriptions of COVID-19, Pneumonia, and normal were then selected to produce a new dataset used in this study. A total of 900 images of data were used in this study, with each class consisting of 300 pieces of data in the form of chest X-ray images.

2.2. Preprocessing

The data used in this study had various sizes and aspect ratios. Meanwhile, the CNN model used in this research only had 224 × 224 as the input size. To overcome this problem, preprocessing was be carried out on the dataset. Resize was one of the transformation methods used to prepare the data, wherein the size of the image data was changed to 224 × 224. However, it should be noted that an image of size 224 × 224 has an aspect ratio of 1:1. In contrast, not all image data used in this study has an aspect ratio of 1:1. A center crop transformation was done before resizing the image to solve the different aspect ratio problem. Center crop was a cropping process performed on the image data so that the remaining part of the image data was the middle part of the image data. Furthermore, the preprocessing stages carried out in this study is in Figure 1.

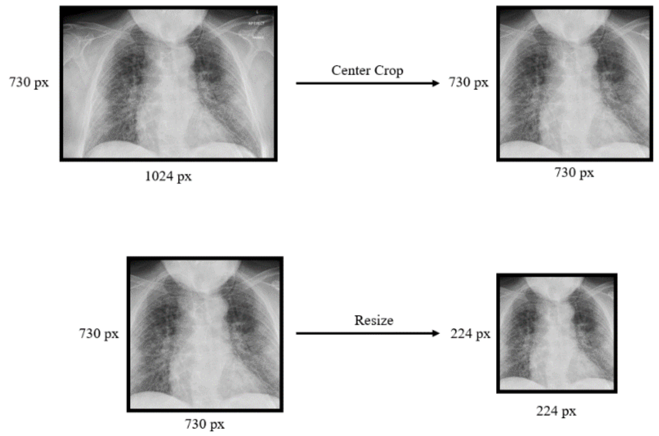

Figure 1. Illustration of the preprocessing used in this research.

2.3. Deep Feature Extraction

In this study, the authors used the CNN method to extract and classify COVID-19 using X-ray image data. The CNN architecture used in this study was ResNet-50, developed by He et al. in 2016 [17]. Due to the limited dataset used in this study, a pre-trained ResNet-50 model was used to speed up the CNN training process. The architecture of ResNet-50 is shown in figure 2.

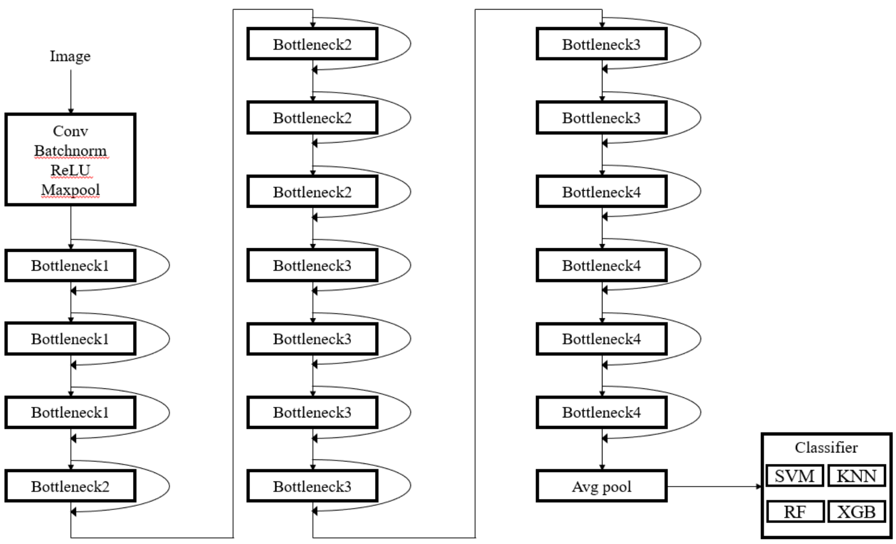

Figure 2. Deep Feature Extraction using ResNet-50 Architecture.

2.3.1. Deep Learning

Deep learning is part of machine learning consisting of an input layer, a hidden layer, and an output layer. In general, the hidden layer is multiple layers of nonlinear processing units that extract features with high complexity from the input data. The hidden layer extracts feature with different complexity levels in each layer, wherein the layer closest to the input layer extracts the most superficial features.

In contrast, the layer closest to the output layer extracts more complex features [14]. The use of deep learning methods such as deep neural networks, recurrent neural networks, and convolutional neural networks have been widely applied to solve problems of classifying images, sounds, videos, and others which are known to be challenging to do in ordinary machine learning [15].

2.3.2. Convolutional Neural Network

CNN is a deep learning model often used to analyze visual images [16]. CNN is one of the neural network models that has solved commercial problems using deep learning and is still at the forefront of deep learning applications in the commercial field [11]. CNN uses a mathematical operation called convolution, which is a particular type of linear operation. In terms of structure, CNN has similarities with a multilayer perceptron, but the operation used in the hidden layer CNN is a convolution operation. One of the advantages of CNN compared to other classification algorithms is feature extraction, which is the model's ability to design features independently. The feature extraction on CNN usually consists of a convolution layer and a pooling layer, which outputs from the feature extraction layer to the fully connected layer as a classification layer on CNN.

2.3.3. Deep Residual Network

The Vanishing Gradient Problem is a common problem when training deep neural networks. To solve this problem, He, Zhang, Ren, and Sun proposed a new Deep Residual Network (ResNet) architecture. This architecture uses a skip connection or identity shortcut connection that passes through one or more layers of the forward propagation. Skip connection performs identity mapping at a layer of which output is summed to the output of the next layer [18]. In ResNet, a residual block is defined as follows:

$$y = F(x, \{W_i\}) + x \tag{1}$$

where W_i represents the weight of the i-th convolution layer, x represents the input, y represents the output, and the function F represents the residual mapping. The ResNet variant that is used in this research was ResNet-50 which consist of 50 convolution layers. These convolution layers are grouped into a three-stacked convolution layer called bottleneck layers.

2.3.4. Machine Learning Classification

Machine learning is the ability possessed by computer algorithms to solve problems such as classification, regression, or grouping without being programmed explicitly (Mitchell, 1997). Machine learning is generally divided into two types based on the presence or absence of target variables in the data, namely supervised machine learning and unsupervised machine learning. In supervised machine learning, the data provided consists of features and target variables. The computer then learns the relationship between the feature and the target variable to give recommendations or predictions when given new feature data. On the other hand, data shown in unsupervised machine learning has no target variables, and the computer is expected to find hidden patterns in the given data. In this study, supervised machine learning was used for classification problems. The classification algorithms used in this research are SVM, Random Forest, KNN, and

XGBoost.

2.4. Evaluation Metrics

Several evaluation metrics would be used in this study, such as accuracy, recall, precision, and running time. Accuracy describes the accuracy of the model in predicting the given data. Recall measures how often the model correctly classifies data into a class. Precision measures a model's ability to produce the same performance over multiple trials, while running time is the time it takes the model to make predictions from the data. In this study, the precision and recall used were from the COVID-19 class. The following were the formulas used to calculate accuracy, precision, and recall:

$$Accuracy = \sum_{i=1}^{n} \frac{tp_i}{tp_i + fp_i + fn_i + tn_i} \tag{2}$$

$$Recall_i = \frac{tp_i}{tp_i + fn_i} \tag{3}$$

$$Precision_i = \frac{tp_i}{tp_i + fp_i} \tag{4}$$

2.5. Training and Testing

2.5.1. Data Separation

Data separation is essential to measure the performance of machine learning models and detect overfitting [15]. Overfitting is a case wherein the model has a small error value in one subset of data and has a significant error value in another data subset [18]. In this study, data separation was carried out after the data passed the preprocessing stage, wherein the data was divided into training and testing data. Then 10% of the training data was taken randomly as validation data. Training data is used to train the model. Validation data is used to validate the model's performance to avoid overfitting and is a reference for fine-tuning the deep learning model. In contrast, a data test is a data that has never been seen by the model and is used to analyze model performance. In this study, several data distributions were carried out and is shown in table 1:

Table 1. Data Separation

Case	Data Training	Data Validation	Data Testing
Case 1 (70:30)	534 images	96 images	270 images
Case 2 (80:20)	612 images	108 images	180 images
Case 3 (90:10)	687 images	123 images	90 images

2.5.2. Model Parameter

Fine-tuning the model will be done by training the model using a minimal learning rate. Because CNN has a hierarchical learning nature, the learning rate used in each layer will be different, wherein a small learning rate is used in the initial layer, and the learning rate will be higher in the later layers. The amount of learning rate used in each layer is shown in Table 2.

Table 2. Learning Rate

Layer	**Learning rate**
Initial Layer (Convolution)	5×10^{-5}
Initial Layer (Batch Normalization)	10^{-4}
Layer 1	10^{-4}
Layer 2	10^{-4}
Layer 3	10^{-3}
Layer 4	10^{-3}
Fully Connected Layer	10^{-2}

The feature vectors that have been extracted from the ResNet-50 model are then given as an input to the predefined classical classification methods, namely, SVM, RF, KNN, and XGBoost. The parameters of each classical classification method in this study were obtained with the help of GridSearchCV by Scikit-Learn. GridSearchCV performs an optimal parameter search for the specified classifier based on the training data provided. Table 3 shows the parameters used in the classical classification method.

Table 3. Model Parameter

Model	**Parameter**
Support Vector Machine	*Kernel:* RBF*, C:*1*, Gamma:* 0,006
Random forest	*N estimators:*800*, Min sample leaf:*3
K-Nearest Neighbor	*N neighbors:* 4
XGBoost	*Learning Rate:*0,3*, Max Depth:*4, M*in Child Weight*: 1

3. Results

In this study, the training and testing phase of the model was done repeatedly five times on each data split case. This repetitive trial is used to evaluate the stability of the model, given that there are randomly initiated weights in the final layer of ResNet-50 architecture. The loss and accuracy values of the train and validation data for each epoch are shown in Figures 3 and 4.

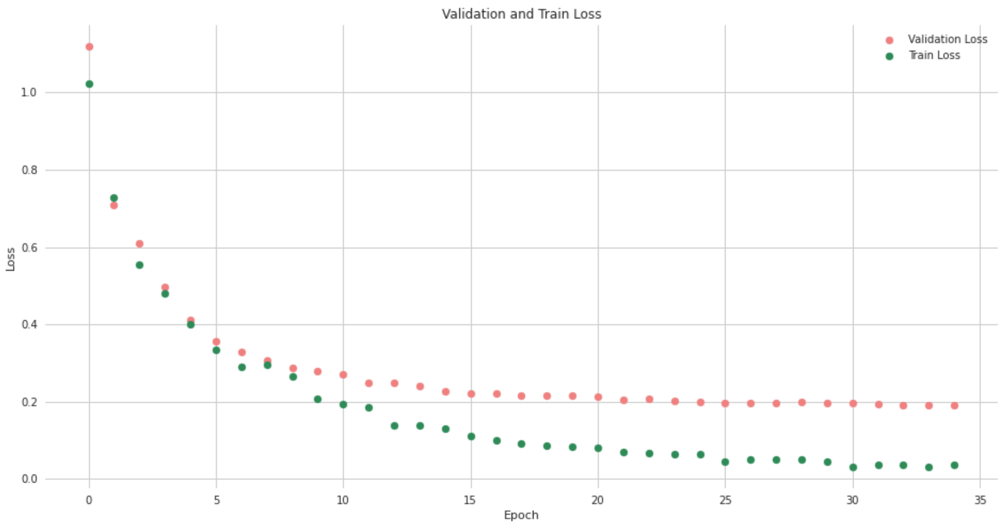

Figure 3. Loss on training and validation data.

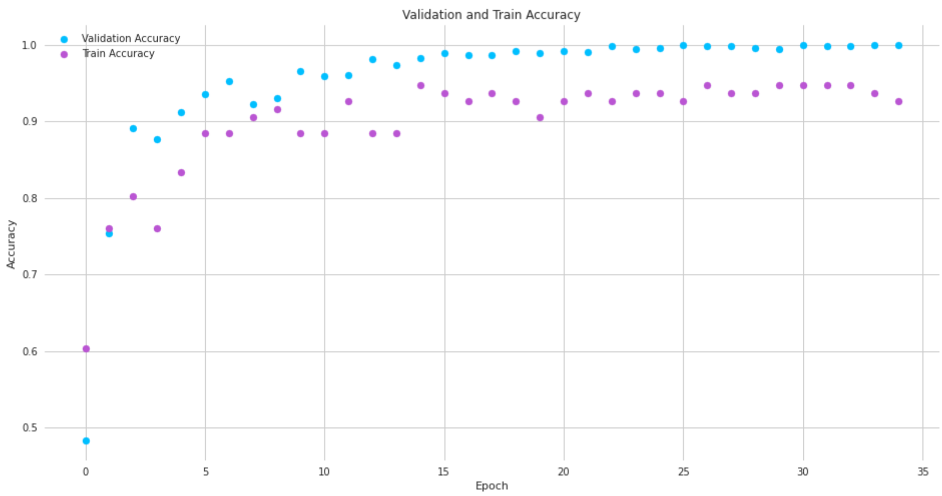

Figure 4. Accuracy on training and validation data.

Based on Figure 3 and 4, the loss value on both training and validation data converges to a certain number, indicating that the model is not underfitting. In addition, the value of training loss and validation loss decreases and converges to a value which indicates that the model is not overfitting.

After the training stage, the fully connected layer in the ResNet-50 model was replaced with an identity layer. This was done to extract the features generated by the ResNet-50 model. A feedforward operation was applied to both the training and testing data resulting in numerical data with 2,048 features resulting from Deep Feature Extraction. The feature extraction results from the train data were then used to train the machine learning classifier. At the same time, the feature extraction results from the test data were used to evaluate the machine learning classifier model. The running time calculations in Table 4 are the time required by the trained machine learning classifier to predict the test data.

Table 4. Deep Feature Extraction Result

Case	Model	Accuracy	Recall	Precision	Running Time
Case 1 (70:30)	ResNet-50	92.15%	86.00%	92.92%	0.0010
	DFE + SVM	92.44%	86.89%	93.35%	0.2004
	DFE + RF	91.56%	85.78%	92.85%	0.0758
	DFE + KNN	92.67%	91.11%	90.75%	0.5871
	DFE + XGB	90.89%	85.78%	90.90%	0.0109
Case 2 (80:20)	ResNet-50	92.22%	83.67%	94.36%	0.0014
	DFE + SVM	93.00%	86.33%	93.86%	0.1438
	DFE + RF	92.67%	86.33%	94.23%	0.0691
	DFE + KNN	93.00%	91.00%	91.92%	0.4424
	DFE + XGB	92.11%	86.33%	92.50%	0.0086
Case 3 (90:10)	ResNet-50	93.11%	91.33%	90.76%	0.0006
	DFE + SVM	94.22%	93.33%	92.08%	0.0755
	DFE + RF	92.67%	92.00%	89.66%	0.0668
	DFE + KNN	92.00%	94.00%	87.06%	0.2380
	DFE + XGB	91.33%	89.33%	88.74%	0.0053

Based on the accuracy value from Table 4, it is known that the ResNet-50 and DFE+SVM models achieve the best performance in case 3, while DFE+KNN and DFE+XGB models achieve the best performance in case 2. At the same time, the DFE+RF model achieves the best performance in both case 2 and case 3. Based on the accuracy value, the DFE+SVM model in case 3 has the best performance with an average accuracy value of 94.22%. Based on the recall value, the DFE+KNN model in case 3 has the best performance, with an average recall value of 94%. Based on the precision value, the ResNet-50 model in case 2 has the best performance, with an average precision value of 94.36%. Meanwhile, the best running time in this study was achieved by the ResNet-50 model in case 3, with a running time of 0.0006 s. The classical classification method model with the fastest running time was achieved by the XGB model in case 3, with a running time of 0.0053 s.

4. Conclusions

In this study, as many as 900 chest X-ray image data taken from three different online databases were used to classify COVID-19 with a Deep Feature Extraction (DFE) approach using ResNet-50 and classical classification methods such as SVM, RF, KNN, and XGBoost. By using categorical class entropy as loss function, Stochastic Gradient Descent as an optimizer, batch size of 50, and 40 epochs in ResNet-50 architecture, we got an optimal DFE model. All data were then extracted into feature vectors by ResNet-50 with 2,048 features. The data was then classified using the classical classification method.

The simulations carried out in this study were carried out five times. This was done to evaluate the stability of the model because there are weights that are randomly initiated in the last layer of the ResNet-50 architecture. The results of the comparison of the evaluations were the average of the five trials. Based on the simulation results, it is known that the best accuracy was obtained by a combination of ResNet-50 and SVM from data sharing in case 3 with 94.22%. A DFE and KNN received the best recall from case 3 with 94%. ResNet-50 got the best precision from case 2 with 94.36%. ResNet-50 brought the best running time from case 3 with 0.0006 s. It could be concluded that the combination of DFE using ResNet-50 with SVM could provide the best performance in classifying COVID-19 using X-ray image data.

Acknowledgments (All sources of funding of the study must be disclosed)

We would like to thank you for following the instructions above very closely in advance.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., . . . Xia, L. (2020). Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology, 296*(2), E32-E40.
2. Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., & Ji, W. (2020). Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology, 296*(2), E115-E117.
3. Yoon, S. H., Lee, K. H., Kim, J. Y., Lee, Y. K., Ko, H., Kim, K. H., . . . Kim, Y.-H. (2020). Chest radiographic and CT findings of the 2019 novel coronavirus disease (COVID-19): analysis of nine patients treated in Korea. *Korean journal of radiology, 21*(4), 494.
4. Cleverley, J., Piper, J., & Jones, M. M. (2020). The role of chest radiography in confirming covid-19 pneumonia. *BMJ, 370*, m2426.
5. Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement, 145*, 511-518.
6. Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in biology and medicine, 121*, 103792.
7. Ismael, A. M., & Şengür, A. (2021). Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Systems with Applications, 164*, 114054.

8. Hilmizen, N., Bustamam, A., & Sarwinda, D. (2020). *The Multimodal Deep Learning for Diagnosing COVID-19 Pneumonia from Chest CT-Scan and X-Ray Images.* Paper presented at the 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI).
9. Rustam, Z., Sudarsono, E., & Sarwinda, D. (2019). Random-Forest (RF) and Support Vector Machine (SVM) Implementation for Analysis of Gene Expression Data in Chronic Kidney Disease (CKD). Paper presented at the IOP Conference Series: Materials Science and Engineering.
10. Li, M., Fu, X., & Li, D. (2020). *Diabetes prediction based on xgboost algorithm.* Paper presented at the IOP Conference Series: Materials Science and Engineering.
11. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*: MIT Press.
12. Wagner, S. (2014). *Combination of convolutional feature extraction and support vector machines for radar ATR.* Paper presented at the 17th International Conference on Information Fusion (FUSION).
13. Lu, X., Duan, X., Mao, X., Li, Y., & Zhang, X. (2017). Feature extraction and fusion using deep convolutional neural networks for face detection. *Mathematical Problems in Engineering, 2017*,1-9.
14. Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing, 7*, 197-387.
15. Chollet, F. (2018). *Deep learning with Python* (Vol. 361): Manning New York.
16. Valueva, M. V., Nagornov, N., Lyakhov, P. A., Valuev, G. V., & Chervyakov, N. I. (2020). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Mathematics and Computers in Simulation, 177*, 232-243.
17. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
18. Patterson, J., & Gibson, A. (2017). *Deep learning: A practitioner's approach*: “O'Reilly Media, Inc..”

Supplementary (if necessary)