

A Computer-Aided Feasibility Implementation to Detect Monkeypox from Digital Skin Images with Using Deep Artificial Intelligence Methods



Ali Berkan Ural 

Department of Electrical Electronics Engineering, Circuit and Systems/Biomedical, Kafkas University, Kars 36000, Turkey

Corresponding Author Email: berkan.ural@kafkas.edu.tr

Special Issue: The 3rd ICAENS Conference

<https://doi.org/10.18280/ts.400139>

ABSTRACT

Received: 6 October 2022

Accepted: 10 January 2023

Keywords:

computer-aided diagnosis, skin lesion diagnosis, Monkeypox, image processing, deep learning

A sudden outbreak of Monkeypox disease has been reported recently in up to 70 countries so far and the spreading rate may be seen significantly around the world. The clinical aspects of Monkeypox have been reported that this disease looks like similar in many attributes when comparing some specific skin lesions as Chickenpox, Measles etc. These similarities make Monkeypox diagnosing and detecting difficult for doctors, clinicians or professionals by examining the visual appearance of the lesion on the skin. In addition, there has been a problem with the lack of detailed information about ultimate diagnosing of novel Monkeypox disease. It is also important that by the success of the studies about AI, Machine Learning and Deep Learning models in COVID-19 detection, the community has begun to give importance to detect Monkeypox via comprehensive AI methods from digital skin images. Moreover, in this paper, we develop a larger dataset to study and analyze the feasibility of common Artificial Intelligence based Deep Learning methods on skin images for Monkeypox detection. Our study has shown that Deep Learning models have a great and important success for detecting this disease from digital skin images via modifying/ updating some layers in the Transfer Learning. The other important information can be explained as because of being quite similar in some aspects to the other skin lesions and the lack of the detailed attributes/features of Monkeypox, detecting via specific AI models with Feature extraction process have become a bit difficult, unknown and time consuming in contrast to the Deep Learning models (AlexNet and VGG16 models in MATLAB software). The future aim is to develop a prototype web application and it is important that to improve the accuracy of Monkeypox detection, a larger demographically diverse dataset is required.

1. INTRODUCTION

While COVID-19 has still continued at the recovery stage, Monkeypox virus has revealed and spread throughout the world at a rapid rate. The World Health Organization (WHO) has declared that this outbreak creates a moderate risk according to the global public health [1].

Monkeypox is a common disease which is defined as the zoonotic class from the genus Orthopoxvirus. It generally looks like Chickenpox, Measles and Smallpox according to the clinical features [2, 3]. Although the fatality rate of the case has been reported to be between 3 to 6 percent for the recent outbreak, early detection of Monkeypox is important for decreasing the low limited ratio. In this scenario, Artificial Intelligence (AI) related systems/algorithms may ultimately limit the global spread.

Generally, the main clinical features of Monkeypox are similar to that of Smallpox, but less severe [4]. On the contrary, according to the skin rashes, Monkeypox disease resembles to Chickenpox or Cowpox [5]. The similarity attribute makes challenging for diagnosing Monkeypox at an earlier time for the healthcare professionals. For the diagnosis of Monkeypox, the Polymerase Chain Reaction (PCR) test has been commonly considered the most accurate tool in order to visual observation of skin lesions [6].

Different AI tools, especially Deep Learning (DL) analysis, have been commonly chosen and used in the medical image analysis for years [7]. Moreover, some different AI methods have played an important role in COVID-19 detection and diagnosis phases from radiological images. The successful results give a chance to the health professionals in using AI approaches for Monkeypox analysis from digitized skin images of patients. However, there is only a limited number of Monkeypox image database which are not publicly and less reliable. So, there are impediments that considering privacy and validity concerns [8]. For this study, the first database has been inspired and collected from "Skin Image Dataset 2022" from Kaggle which is public and freely accessible. Indeed, we present a DL-based preliminary feasibility study which consisted from Transfer Learning models of AlexNet and ResNet-50 in MATLAB software for Monkeypox Detection from digitized skin images.

According to the literature, only two AI based Monkeypox detection studies have been obtained as preprints [9, 10]. However, these studies have some important limitations. First, these studies commonly used only three cases of disease (Monkeypox, Chickenpox and Measles). Second, these studies were implemented on a small image database and the results were obtained for only these imageset. Third, these studies tested/analyzed one of a few AI and DL models in the

classification task.

In this paper, the feasibility of utilizing common AI methods are tested to classify different types of pox from digital skin images of lesions and rashes. Some novelties of this study can be given below.

- 1) We develop a collective database consisted from skin lesions/rashes images of totally five important diseases and healthy skin images.
- 2) Our database were collected from more data from various sources (for example news, media, websites, public database websites, also some of them were chosen and used from Kaggle database) for these diaseases used in this study and some healthy images were scrapped on the web.
- 3) We analyzed and tested the disease classification metrics of two popular and important deep models of MATLAB (customized AlexNet and VGG16) from digital skin images.
- 4) Indeed, we performed 5 fold cross validation tests for each of the AI based DL models to more analyze our results, that is not achieved in previous studies [9, 10], however their dataset volume is smaller than ours.

2. MATERIALS AND METHODOLOGY

In this section, the whole method, data collection and experimental phase are described in detail.

2.1 Data collection

Data collection phase have been achieved and consisted from Web-scraping, Expert screening, Data pre-processing and Data augmentation. As there has been no shared dataset which has been available by the authorized and public by health departments and health professionals, thus, to establish a preliminary various dataset, the Monkeypox image dataset has been collected from different sources such as websites, news and public online portals (Kaggle etc.) using the Google search engine via searching “Monkeypox skin image data”. According to Figure 1, the searching process was achieved according to the “Common licences”. However, we found hardly some of the pox images, we have also collected. images according to the “Commercial and other licences” part. In Figure 2, some of the pox images we were obtained are given in detail. Moreover, to increase the data sample size, data augmentation progress was applied to all images.

In expert screening phase, two expert doctors who were expert in infection diseases from Kafkas University and Harakani Hospitals, screened all images from our database and labeled and checked all of the imageset.

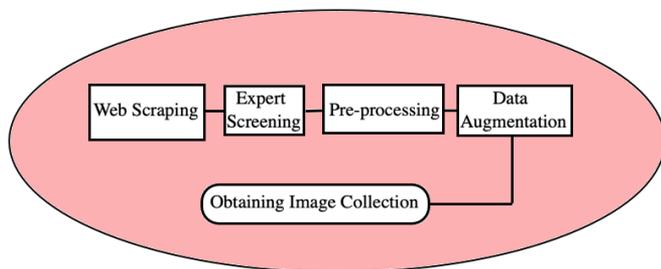


Figure 1. The chart of improving the image database

In pre-processing phase, the images were cropped and unwanted regions were eliminated from the digitized skin images. Since AI deep models take inputs such as square shaped images related to the pixel count of 224x224x3, all images were arranged according to this criteria and if needed zero padding was applied to the outside of the images to obtain 224x224x3 pixel criteria.



Figure 2. Examples of skin images of Monkeypox, Chickenpox, Measles and healthy images from our database (after obtaining from search engine and expert screening parts)

In data augmentation phase, Keras image processing library such as “Imagedatagenerator” was mainly used to increase the number and variability of images for collected image database [11]. For this reason, totally 12 augmentation operations were performed on the web-scraped images. The augmentation operations consisted from (1) color modification with a randomly developed factor value ([0.4, 1.8]), (2) brightness modification with a randomly chosen value ([0.3 1]), (3) sharpness modification with a random value ([0.6 2]), (4) rotate images with the randomly chosen angle ([-45 45]), (5) adding Gaussian noise with a randomly chosen variance value ([0.002 0.25]), (6) adding salt&pepper noise to randomly generated image pixels (range %3 to %7), (7) adding speckle noise to randomly generated image pixels, (8) contrast modification with the value ([0.4 0.85]), (9) image translation with a random distance between 15 and 25 pixels, (10) zooming in an image with the value of %10, (11) zooming in an image with the value of %15, (12) flipping images with the height and width value ([-90 90]). In Table 1, the sample size of each collected dataset was given in detail. Indeed, in Figure 3, each collected dataset after augmentation were given in detail.

Table 1. The sample size of each collected dataset in this study

Collected dataset	Total sample
Monkeypox	100
Chickenpox	100
Measles	100
Normal/Healthy	100
Monkeypox augmented	3000
Chickenpox augmented	3000
Measles augmented	3000
Normal augmented	3000
Total sample	16000

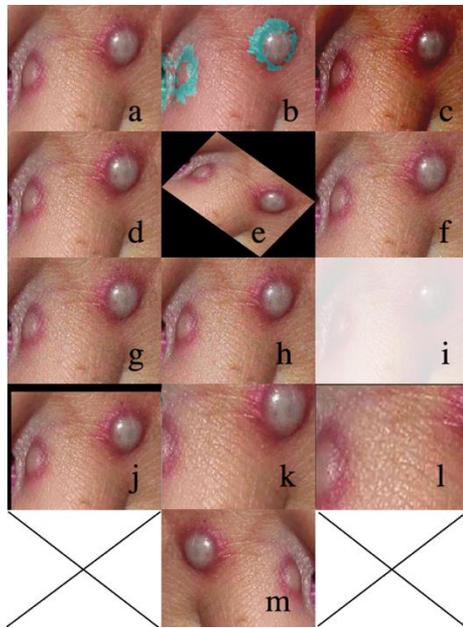


Figure 3. Illustration of (a) original sample image and related augmented images by (b) color modification with a randomly developed factor value $[0.4, 1.8]$, (c) brightness modification with a randomly chosen value $[0.3, 1]$, (d) sharpness modification with a random value $[0.6, 2]$, (e) rotate images with the randomly chosen angle $[-45, 45]$, (f) adding Gaussian noise with a randomly chosen variance value $[0.002, 0.25]$, (g) adding salt&pepper noise to randomly generated image pixels (range %3 to %7), (h) adding speckle noise to randomly generated image pixels, (i) contrast modification with the value $[0.4, 0.85]$, (j) image translation with a random distance between 15 and 25 pixels, (k) zooming in an image with the value of %10, (l) zooming in an image with the value of %15, (m) flipping images with the height and width value $[-90, 90]$

2.2 Transfer learning model approaches

To evaluate the performance of the AI deep models on developed dataset transfer learning approach was performed in the preliminary experimentation. In this phase, common AlexNet and a modified version of VGG16 models were used and results were compared in detail. The main model consisted from three main parts; these were pre-trained phase, an up-to-date layer and an estimation class. Indeed, according to MATLAB, AlexNet has mainly 8 constant layers and this model can be defined and used in transfer learning approaches [12, 13]. In Figure 4a, the proposed AlexNet model was shown in detail. According to this figure, first, images were used as input of the system. Then, convolution phases were performed and possible redundancy were eliminated. Then, fully connected layers were performed. In this phase, these layers contained Artificial Neural Network (ANN) part and learning has been achieved via ANNs. Finally, the classification has been performed successfully. Indeed, mainly, the second and fifth layers of AlexNet were customized and modified according to the fixing of the weights, because the other layers were unchangeable layers and only the changeable layers were chosen and used for training phase for AlexNet.

The pre-trained architecture was used to obtain the high dimensional features and in future this was added to update/improve modified layer. In Figure 4b, the proposed VGG16 updated model was shown in detail. The model

consisted from totally 17 Convolutional Neural Network (CNN) layers with different filter sizes and stride values [14]. As shown in Figure 4, after the initial input layer (224x224 images) two convolutional layer with 3x3 filters was added. Then, this layer was followed by a Max Pooling layer, followed by another two convolutional and one Max Pooling layer until this reached the modified layer parts. The modified layer of Flattened layer followed by the three dense and two dropout layers.

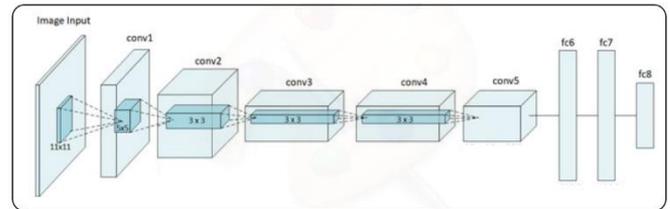


Figure 4a. Implemented AlexNet model in this study

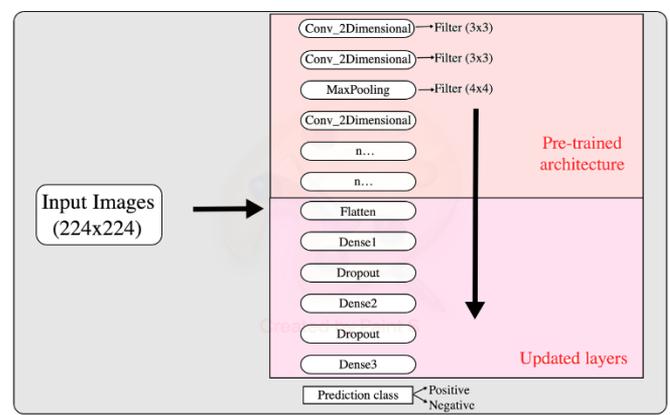


Figure 4b. Implemented modified VGG16 model using transfer learning in this study

The detailed parameters of deep models such as number of epochs, batch size and learning rate were examined during the experimental phase of the study to maximize the performance of the proposed models. For the first AI deep model of AlexNet, the parameters were inspired by and used as [15];

Number of epochs=[30, 35, 40, 45];
Batch size=[5, 10, 15, 20];
Learning rate=[0.1, 0.01, 0.001];

With using the Grid search method, the most optimal parameters were identified and used as;

Number of epochs=40;
Batch size=15;
Learning rate=0.001;

For the second AI deep model of modified VGG16, the parameters were used as;

Number of epochs=[40, 60, 90, 150];
Batch size=[20, 40, 60, 90];
Learning rate=[0.1, 0.01, 0.001];

With using the Grid search method, mainly, this took five different hyper parameters such as hidden node size of layer one and two, optimizer type, maximum epoch and transfer

function parameters. The hyper parameters were added for trying the function with 5 fold cross validation of each combination and then, finally, the best hyper parameter combination returned as cell and this was used as parameters of the AI model. Moreover, after more trials, the most optimal parameters were identified and used as;

Number of epochs=90;
 Batch size=20;
 Learning rate= 0.001;

2.3 Experimental setup

The experimental part was evaluated using a Macbook Pro, 64 GB RAM and Intel Core i7 microprocessor. At the end of the experiment, 5 fold cross validation was applied on the finalized train/test data and outcomes were evaluated and compared in detail.

In the deep neural networks, specifically, %80 of the sample data were allocated to train and %20 of the sample data were allocated to test the models. Since our digital image database is small, we used augmentation and 5 fold cross validation to improve and emphasize the models' outcomes. We also split our preprocessed original skin images in 5 equal folds per class [16]. Because of the possible imbalance among our original images, we use equally augmented images per class during training to make the data more balanced [17, 18]. In Table 2, the distribution of the images used in validation per class and the number of augmented images used in training per class in each fold.

Table 2. Number of images per class in 5 fold cross validation phase

Class	Training images	Validation images	Augmented images
Monkeypox	80	20	2400
Chickenpox	80	20	2400
Measles	80	20	2400
Normal	80	20	2400

3. Results And Discussion

In this part, the comparative classification performance of the AI deep models were summarized and given in this section. Indeed, in Figures 5 and 6, the confusion matrices for 5 fold cross validation estimation for AlexNet and VGG16 were given in detail. Moreover, we saw some misclassifications according to the disease classes for all deep models, except for the healthy group. The least number of misclassifications for 5 folds was obtained for VGG16 AI deep model.

In this part, we also present the comparison of precision,

recall, F1 score and mean accuracy via 5 fold cross validation and these were given in Table 3.

According to Table 3, we clearly saw that the best accuracy was achieved via VGG16 deep model (%80). From the overall observation on the classification performance by two different deep models, as given in Table 3, the main hypothesize was that a larger number of trainable parameters might be underfitted according to the small training sample size of the imageset. In contrast, although AlexNet has fewer number of trainable attributes, this was overfitted on the small number of training samples and this occasion was resulted in worse prediction performance on the validation data.

We also present classification performance via using majority voting to make the estimation phase. In Figures 5 and 6, we show the confusion matrices of two models for 5 fold cross validation predictions by AlexNet and VGG16 deep models.

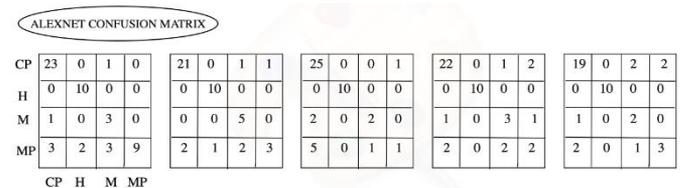


Figure 5. Confusion matrices of prediction via AlexNet model in 5 fold cross validation (CP: Chickenpox, H: Healthy, M: Measles, MP: Monkeypox)

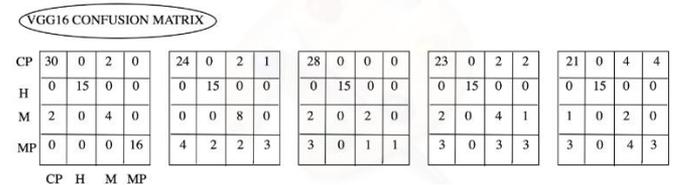


Figure 6. Confusion matrices of prediction via VGG16 model in 5 fold cross validation (CP: Chickenpox, H: Healthy, M: Measles, MP: Monkeypox)

According to the confusion matrices given in Figures 5 and 6, the last model (VGG16) approach reduced the misclassifications importantly. Thus, we present the quantitative comparison of mean precision, recall, F1 score and accuracy, estimated over 5 fold cross validation. According to the models, the best accuracy was obtained from VGG16 deep model compared to AlexNet model.

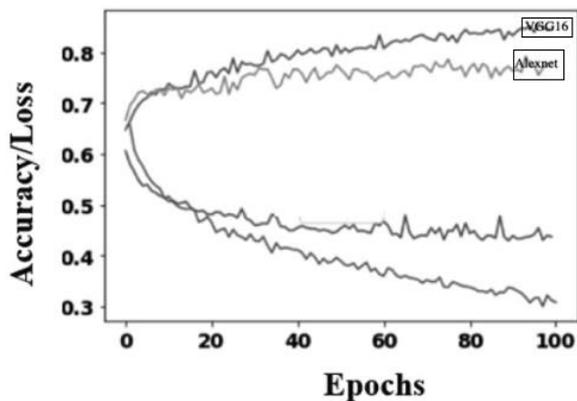
In addition, in Table 4, the proposed models' performances on the developed dataset used in this study, along the confidence interval as 0.05. Also, in Figure 7, accuracy/loss distribution for each epoch of two AI deep models was given in detail.

Table 3. Quantitative comparison of mean precision, recall, F1 score and accuracy via 5 fold cross validation

Methods	Class	Precision	Recall	F1 Score	Accuracy
AlexNet	Monkeypox	0.70	0.52	0.60	0.71
	Chickenpox	0.74	0.69	0.70	
	Measles	0.44	0.50	0.47	
	Normal	0.95	1.00	0.95	
VGG16	Monkeypox	0.80	0.59	0.66	0.80
	Chickenpox	0.80	0.68	0.71	
	Measles	0.55	0.73	0.61	
	Normal	1.00	0.98	0.99	

Table 4. Proposed models' detailed performances on the developed dataset used in this study

Models	Dataset	Accuracy	Sensitivity	Specificity
AlexNet	Train set	0.71±.009	0.82±0.010	0.90±0.008
	Test set	0.63±0.021	0.650±.022	0.85±0.13
VGG16	Train set	0.80±.018	0.975±0.015	0.98±0.018
	Test set	0.75±0.071	1	0.70±0.14

**Figure 7.** Deep models accuracy and loss during each epoch

In this study, we developed a new digitized skin image dataset that could be used to train the AI based deep models to classify Monkeypox disease. Indeed, we present preliminary results of Monkeypox disease detection from digitized skin images using custom and modified AI deep models. For this reason, custom AlexNet and modified VGG16 models were chosen and used. Our proposed models achieved successful results on the small image database. First model AlexNet achieved $0.71\pm.009$ accuracy value and the second model VGG16 achieved $0.80\pm.018$ accuracy value, respectively. According to the ROC analysis parameters such as Precision, Recall, F1 score and Accuracy metrics, higher and successful results were obtained and this performance test was an acceptable and a wide-range-used test in literature. According to the main advantage of this method, the AI models used in this study were common but customized/modified models and as a result, after fixing steps, more higher and successful results were obtained in order to using the models in raw/common versions. Recently, a report has been announced from WHO that better and appropriate ML models could have been developed according to the period of the diseases. Indeed, our data collection part and performance part of the models were analyzed and checked according to the doctors who were expert in infection diseases and this could be said that our system models' performance metrics were satisfactory. Since there has not been any detailed Monkeypox image dataset, therefore, we did not find any studies which had important outcomes that can be used to compare our model's performance.

Although the classification performance metrics of two models were quite important and satisfactory, some constraints limit the feasibility of the results. Firstly, although we developed a new various dataset via several methods, the current number of images in the dataset was limited. This occasion decreased the general capability performance of the deep models. If more and more image samples could be added to the image database with a better demographical, geographical, gender distribution etc., this could be possible to achieve a higher and consistent performance values due to the possible variabilities. Secondly, the accuracy and other performance metrics of the system could be further improved

by using a multiple source based thermoscopic images in the pretraining phase. Also, a more concerted effort and different collaborations have been needed to improve the image database and obtain more higher important results from the Monkeypox detection studies.

4. CONCLUSION

In this paper, we tested the feasibility of using two popular important MATLAB based AI deep models of AlexNet and VGG16 in classifying different pox types to mainly investigate and distinguish the Monkeypox disease among the other skin lesions and rashes. Also, we used and tested the dataset on balanced and developed dataset, so important outcomes were obtained from the experimental part of the study. Moreover, our 5 fold cross validation experiment phases showed that AI based deep models had the capability to disseminate and discriminate among different pox types using skin images of pox/measles images and rashes. Although there have been some constraints to limit the Monkeypox detection study, these could be overcome by developing/updating the dataset by continuously collecting new infected patients images, performing the proposed AlexNet and VGG16 deep models' performance metrics on highly imbalanced data, evaluating our proposed models in developing mobile based diagnosis tools for future development and comparing all findings continuously.

REFERENCES

- [1] Thornhill, J.P., Barkati, S., Walmsley, S., et al. (2022). Monkeypox virus infection in humans across 16 countries—April–June 2022. *New England Journal of Medicine*, 387(8): 679-691. <https://doi.org/10.1056/NEJMoa2207323>
- [2] Sklenovská, N., Van Ranst, M. (2018). Emergence of Monkeypox as the most important Orthopoxvirus infection in humans. *Frontiers in Public Health*, 6: 241. <https://doi.org/10.3389/fpubh.2018.00241>
- [3] Rizk, J.G., Lippi, G., Henry, B.M., Forthal, D.N., Rizk, Y. (2022). Prevention and treatment of Monkeypox. *Drugs*, 82(9): 957-963. <https://doi.org/10.1007/s40265-022-01742-y>
- [4] Gong, Q., Wang, C., Chuai, X., Chiu, S. (2022). Monkeypox virus: A re-emergent threat to humans. *Virologica Sinica*, 37(4): 477-482. <https://doi.org/10.1016/j.virs.2022.07.006>
- [5] Hussain, M.A., Hamarneh, G., Garbi, R. (2021). Cascaded regression neural nets for kidney localization and segmentation-free volume estimation. *IEEE Transactions on Medical Imaging*, 40(6): 1555-1567. <https://doi.org/10.1109/TMI.2021.3060465>
- [6] WHO. (2022). Monkeypox Fact Sheet. <https://www.who.int/news-room/fact-sheets/detail/monkeypox>, accessed on May 2022.

- [7] Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., Yang, G.Z. (2016). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1): 4-21. <https://dx.doi.org/10.1109/JBHI.2016.2636665>
- [8] Wang, F., Casalino, L.P., Khullar, D. (2019). Deep learning in medicine—promise, progress, and challenges. *JAMA Internal Medicine*, 179(3): 293-294. <https://doi.org/10.1001/jamainternmed.2018.7117>
- [9] Hosny, K.M., Kassem, M.A., Foad, M.M. (2019). Classification of skin lesions using transfer learning and augmentation with Alex-net. *PloS One*, 14(5): e0217293. <https://doi.org/10.1371/journal.pone.0217293>
- [10] Pan, S.J., Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10): 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- [11] Shorten, C., Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1): 1-48. <https://doi.org/10.1186/s40537-019-0197-0>
- [12] 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. <https://doi.org/10.1109/CVPR.2016.90>
- [13] Ahsan, M.M., Alam, T.E., Trafalis, T., Huebner, P. (2020). Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and Non-COVID-19 patients. *Symmetry*, 12(9): 1526. <https://doi.org/10.3390/sym12091526>
- [14] Ahsan, M.M., Ahad, M.T., Soma, F.A., Paul, S., Chowdhury, A., Luna, S.A., Yazdan, M.M.S., Rahman, A., Siddique, Z., Huebner, P. (2021). Detecting SARS-CoV-2 from chest X-Ray using artificial intelligence. *IEEE Access*, 9: 35501-35513. <https://dx.doi.org/10.1109/ACCESS.2021.3061621>
- [15] Miranda, G.H.B., Felipe, J.C. (2015). Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization. *Computers in Biology and Medicine*, 64: 334-346. <https://doi.org/10.1016/j.compbimed.2014.10.006>
- [16] Wang, L., Lin, Z.Q., Wong, A. (2020). Covid-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. *Scientific Reports*, 10(1): 1-12. <https://doi.org/10.1038/s41598-020-76550-z>
- [17] Ahmed, W.S. (2020). The impact of filter size and number of filters on classification accuracy in CNN. In *2020 International Conference on Computer Science and Software Engineering (CSASE)*, pp. 88-93. <https://doi.org/10.1109/CSASE48920.2020.9142089>
- [18] Menzies, T., Greenwald, J., Frank, A. (2006). Data mining static code attributes to learn defect predictors. *IEEE Transactions on Software Engineering*, 33(1): 2-13. <https://doi.org/10.1109/TSE.2007.256941>

NOMENCLATURE

AI	Artificial Intelligence
CAD	Computer Aided Diagnosis
CNN	Convolutional Neural Network
DL	Deep Learning
MATLAB	Matrix Laboratory Software
VGG16	Visual Geometry Group 16