Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)
www.bircu-journal.com/index.php/birci
Budapest Institute

Comparative Analysis of Convolutional Neural Network Methods in Detecting Mask Wear

Handoko[1], Fahrul Rozy[2], Felix Elbert Gani[3], Abdi Dharma[4]
[1,2,3,4] *Universitas Prima Indonesia*
willkip08@gmail.com, rozyfahrul14@gmail.com, felixdevilz@gmail.com, abdidharma@unprimdn.ac.id

Abstract

Covid19 is a disease caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). As a result, the respiratory system becomes disrupted. The spread of this disease is very easy through droplets. The use of masks is one way to prevent being attacked by this virus. The system for detecting the use of masks is very necessary for today in detecting whether someone is wearing a mask. Convolutional Neural Network (CNN) is a method that can be used to detect masks. In this study, the VGG16, Resnet50, and MobileNet models will be used. Before conducting data training, data preprocessing and data augmentation were carried out on the dataset. The test accuracy of the VGG16, Resnet50, and MobileNet models are 96% and 96% and 98%, respectively. From the test results, it is found that the MobileNet model is more appropriate in the case of mask detection. The conclusion obtained is the use of the MobileNet architecture, the resources used for classifying can be reduced compared to other architectures. MobileNet uses the Depth-Wise Separable method in the computing process which reduces the computational process.

Keywords
Classification; convolutional neural network; use of masks; vgg16; resnet50; mobilenet

I. Introduction

On February 11, 2020, WHO announced a disease called Coronavirus Disease (COVID-19). The spread of this disease is very easy through droplets from someone who has contracted the Covid-19 disease. To suppress the spread of disease, it is necessary to use masks, physical distancing, and restriction of mobility. By using a mask, the user can prevent the inhalation of droplet particles in the air [2].

The use of masks has become mandatory nowadays, especially in public places. In public places there is usually a check whether someone is wearing a mask or not. People who do not wear them will be warned to wear masks or are not allowed to be in these public places.

The system for detecting the use of masks is very necessary for today in detecting whether someone is wearing a mask. One method that can be used is Convolutional Neural Network (CNN). CNN has a pretty good ability to learn very abstract features in identifying objects efficiently. This is because CNN uses feed-forward a fairly deepSeveral problems that have been solved using the CNN method are face detection [4], disease [5], writing recognition [6], and others [7].

In implementing CNN, it is necessary to determine the model architecture used [8]. Currently, there are many effective architectures including VGG16 [9], Resnet50 [10], MobileNet [11], and others. These models usually have their own advantages. Models such as VGG16, Resnet50, and MobileNet have been used to detect masks [12] [13], but use different datasets.

DOI: *https://doi.org/10.33258/birci.v5i2.5605*

Based on the above background, the author wants to conduct a study entitled "Comparative Analysis of Convolutional Neural Network in Detecting Mask Usage". This study will compare the implementation of the Convolutional Neural Network in classifying images where the person in the image wears a mask or not with several existing models. The comparison will use the same dataset.

1.1. Problem Formulation

The formulation of the problem in this study is how the performance of several Convolutional Neural Network (CNN) models in detecting the use of masks?

1.2. Purpose and Benefits Problem

a. The Objectives

The purpose of this study was to determine the performance of several CNN models in detecting the use of masks.

b. The Benefits

Benefits of this research are:

1. Can detect images of people wearing masks or not
2. Can choose a better model in detecting masks
3. Can be developed to make a video-based mask detection system

1.3. Problem Limitations

In this study, the problem limits are as follows

1. Used in this research is image.
2. The dataset used is from https://www.kaggle.com/omkargurav/face-mask-dataset.
3. The amount of data from the dataset is 7553 images where there are 3725 images wearing masks and 3828 images
4. The method used in this research is Convolutional Neural Network.
5. The architecture model used is MobileNet VGG-16, and Resnet 50.
6. Pretrained Model will be used in this research.
7. The classification limitation in this study is wearing a mask and not wearing a mask

1.4. Novelty

Convolutional Neural Network is a method that can be used for image classification. CNN uses feed-forward , so it has a fairly good ability to study very abstract features in identifying objects efficiently [3]. CNN is often used to solve problems related to classification such as face detection [4], disease [5], writing recognition [6]

In a study conducted by Farid et al, researchers compared the KNN, SVM, and CNN methods in performing mask detection. . In their research, it was seen that CNN had a fairly good performance compared to other methods [14].

Several architectures that have been used in mask detection using the CNN method are ResNet50 [12] [13] , Sequential CNN, VGG-16 and MobileNetV2 [15]. From the results of using this method, it can be seen that the CNN method has a fairly good accuracy.

In this study, we will compare RestNet50, VGG-16, and MobileNet models in detecting images of wearing masks using the same dataset.

II. Review of Literature

2.1 Types

Research This research is a type of quantitative research where research is conducted by developing mathematical models, theories or hypotheses. This study will detect masks using the *convolutional neural network* (CNN)

2.2 Time and Place

Be carried out from August to November 2021 at the homes of the research team members, the research time is from 17.00 to finish every day except Saturday and Sunday

2.3 Research

The stages of work in research can be described as follows.

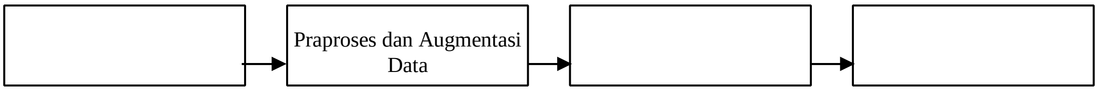

Figure 1. *The Work Phase*

a. Data Collection

Used in this study is an online dataset sourced from Kaggle with the link https://www.kaggle.com/omkargurav/face-mask-dataset.

b. Preprocessing and Data Augmentation

Before carrying out the training process, the image to be trained must go through preprocessing and augmentation. The preprocessing that will be carried out on the image is as follows:

1. Resizing the image
   In this process, the entire image will be converted into one of the same size. The size that is changed will match the size of the input on the *Convolutional Neural Network architecture.*
2. *Rescaling*.
   In this process, each pixel of the image will be divided by 255 to produce a value between 0 to 1. This process aims to speed up convergence during training**.**
3. Split data.
   In this process, the data will be separated into 2 groups, namely training and testing. The ratio used to separate this data is 7:3. Rotation, flipping horizontally and vertically, and cropping the image are augmentation methods performed on the dataset,

c. Training

In this process, data from the training group will be trained using the *Convolutional Neural Network*. This study will use the VGG-16, Resnet50, and MobileNet models with the *Transfer Learning* in dataset training.

VGG-16 is a CNN architecture that won the 2014 ILSVR (ImageNet) competition. This architecture is an example of the model used for the forerunner of the vision model to date. VGG16 uses a convulsion layer with a fairly small filter, namely 3x3. VGG16 has 16 layers consisting of 13 convolution layers, and 3 *fully connected*.

Figure 2. *The VGG-16*

Resnet-50 architecture is a CNN architecture that uses a new concept, namely *shortcut connections*. This concept makes the input from the previous layer into the input to the output of that layer. This concept is used to prevent the loss of important features during the convolution process.

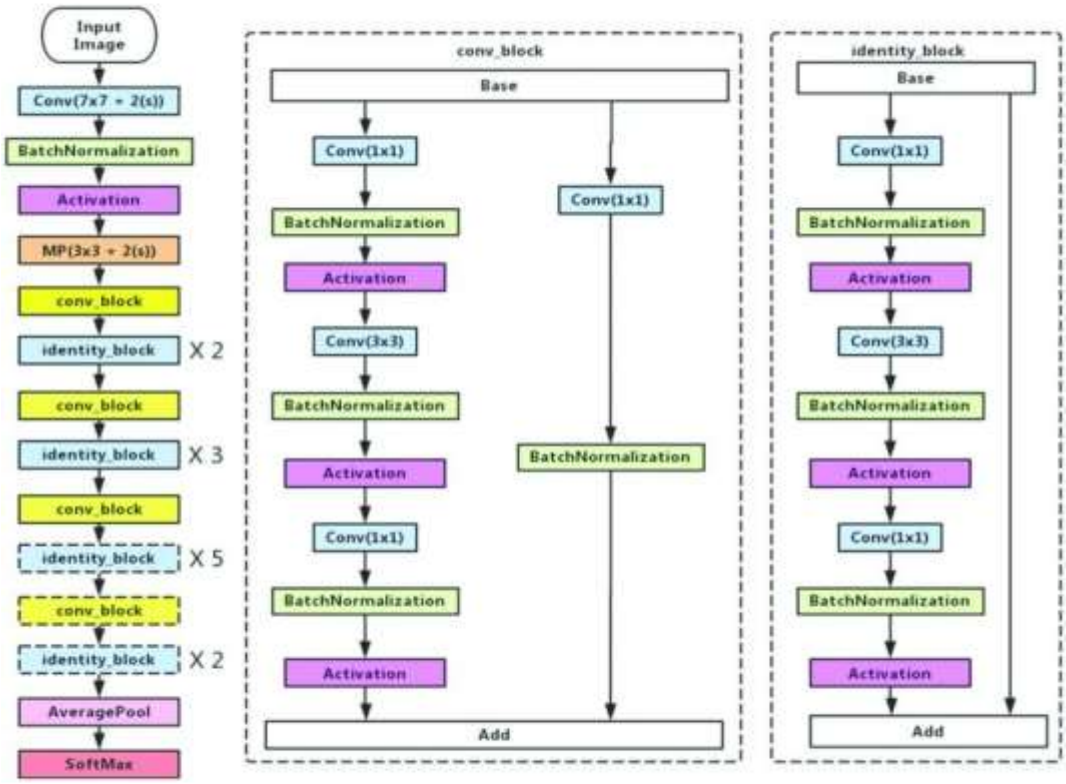

Figure 3. *Architecture Resnet-50*

MobileNets is an architecture that was created to address the problem where resources are lacking. Researchers from Google made this architecture so that it can be used by mobile. The use of layers and filter thickness is what makes MobileNet different from other CNN architectures. The MobileNet convolution layer is divided into *depthwise* and *pointwise convolution* as shown in the following

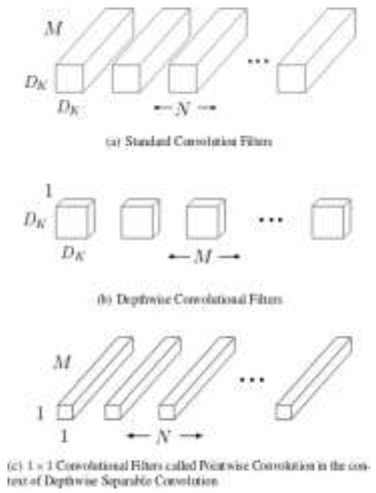

Figure 4. *Standard convolution (a) depthwise convolution (b) pointwise convolution (c) to create depthwise separate filters*

d. Testing

In this process, the model that has been trained will be tested on a group of test data. The test aims to find out whether the performance of the trained model is good enough.

2.4. Tools and Materials

a. Tools

In this study, the tools used have the following specifications:

Windows 10 pro 64 bit, Ryzen 5 3600 6 Core processor (12 CPUs) ~3.6Hz, 16 GB RAM, GPU Nvidia geforce GTX

b.1650Super

Consists of 7553 images, from these images there are 3725 images that wear masks and 3828 images that do not wear masks.

Figure 5. *Image Using Mask*

Figure 6. *Image Not Wearing a Mask*

III. Result and Discussion

3.1 Preprocessing and Augmentation

Preprocessing and augmentation were carried out before training the dataset. Several preprocessing methods were carried out on the dataset as follows:

a. Changing Image Size

   In this process, various image sizes are converted into one size. which is 224 x 224 pixels.

b. Rescaling

   In this process, the entire image is normalized where each pixel of the image is divided by 255. The value of each image pixel is between 0-1.

c. Separating the data

   In this process, the dataset is divided into 2 groups, namely the training and testing groups. With the use of a ratio of 7:3, the training group will use 5,288 data and the testing group will use 2,265 data.

   In training the data will also be divided into training and validation data. The division ratio used is 8:2. With the use of this ratio, the training data becomes 4231 data and validation data is 1,057 data.

a. Training

The number of parameters trained on the VGG-16, Resnet50, and MobileNet models respectively was 134,268,738. 11,511,784, and 4,946,890. In this study, the training was planned to run for 50 iterations. When overfitting has occurred the training will be stopped.

In this case, the parameter that we use to show overfitting is validation accuracy. The time required for training for each VGG-16, Resnet50, and MobileNet models is 21 minutes, 23 minutes, and 16 minutes, respectively. The training process can be visualized through the value of accuracy, and loss as shown below.

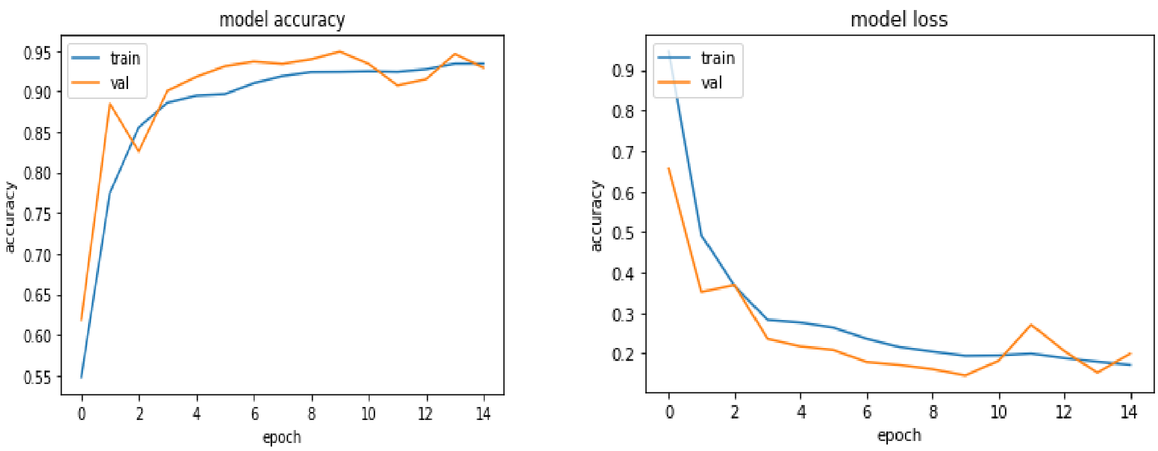

Figure 7. *VGG-16 Training Accuracy and Loss*

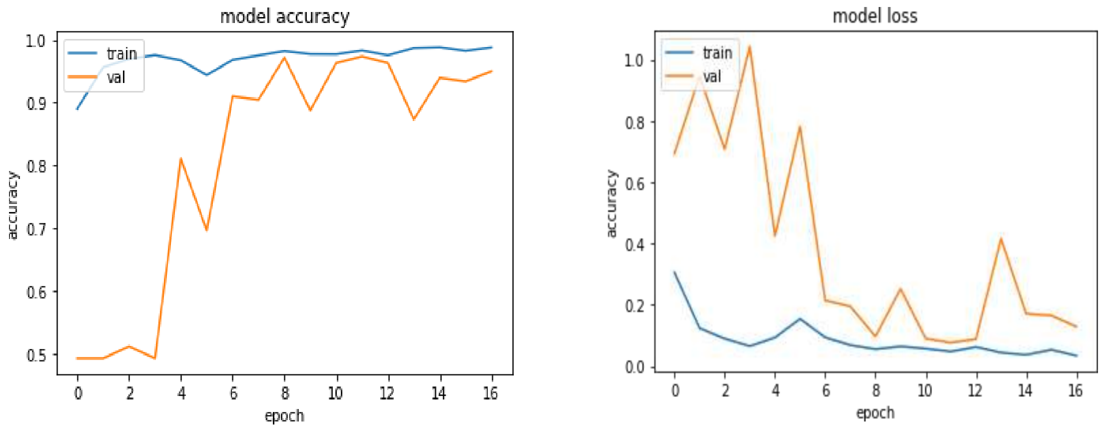

Figure 8. *Resnet 50 Training Accuracy and Loss*

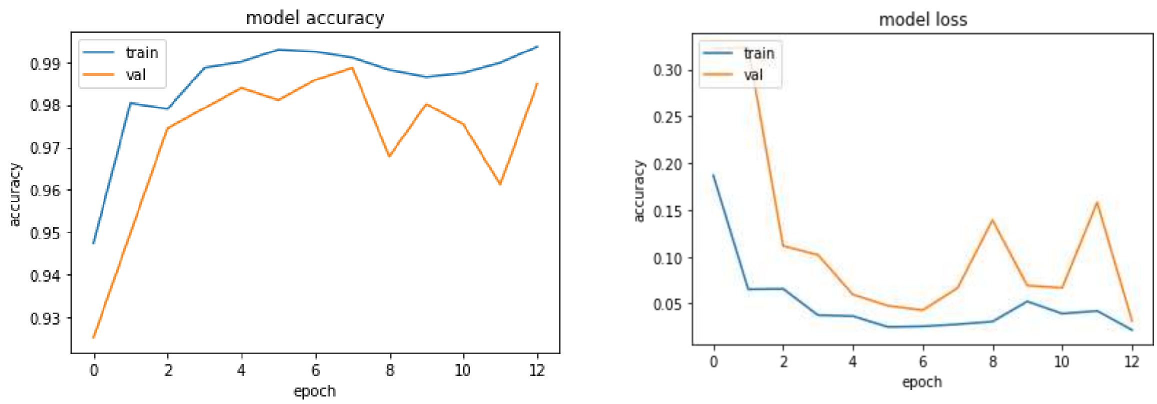

Figure 9. *Mobile Net*

b. Testing

After conducting the training, the next process is to test the model that has been trained. Testing of test data can be seen in the following table.

Table 1. *Confusion Matrix* VGG-16

		Predict	
		Mask	No Mask
Actual	Mask	**1044**	73
	No Mask	17	**1131**

Table 2. *Confusion Matrix* Resnet50

		Predict	
		Mask	No Mask
Actual	Mask	**1069**	48
	No Mask	34	**1144**

Table 3. *Confusion Matrix* MobileNet

		Predict	
		Mask	No Mask
Actual	Mask	**1116**	1
	No Mask	35	**1113**

The following is a comparison of the evaluation values of VGG-16, Resnet50, and Mobile Net which are visualized in the form of tables and graphs.

Table 4. Comparison of Evaluation Value

No	**Name Comparison**	**VGG-16**	**Resnet50**	**MobileNet**
1	Number of trained parameters	134,268,738	11,511,784	4,946,890
2	Number of iterations required	15	17	13
3	Accuracy	96%	96%	98%
4	*Recall*	98.5%	97%	97%
5	Precision	94%	96%	99%

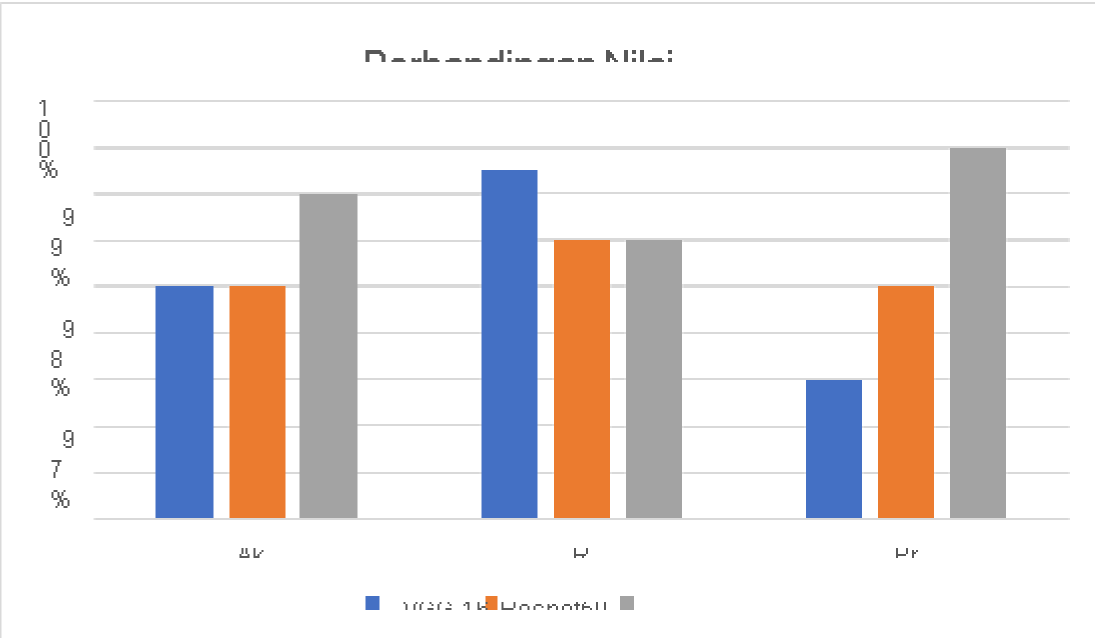

Figure 10. *Comparison Graph of Evaluation Values*

Based on the test results, it can be seen that all models can be used to detect masks where the accuracy value is 90%. However, when compared to the three models, MobileNet has better performance. This can be seen where the accuracy is at 98%.

3.2 Discussion

The average accuracy of the test data from this study is 96.67% with the number of training iterations less than 20. From the research it is proven that MobileNet is an architecture that overcomes problems where resources are lacking [11]. This can be seen where the accuracy of MobileNet is higher than other models. MobileNet is also a model developed after VGG-16, and Resnet so it has quite good performance.

The difference between MobileNet with VGG-16, and Resnet is the use of the Depthwise Separable Convolution in the computational process where VGG-16 and Resnet 50 have not used this method. With the use of Depthwise Separable Convolution , the computational process is reduced. The Depthwise Separable Convolution will factorize the standard convolution for simplification [16]. MobileNet is a solution for using CNN with less resources.

IV. Conclusion

From the experiments carried out, several conclusions were obtained, namely:

1. The VGG-16, Resnet50, and Mobile methods can be used to detect masks.
2. MobileNet is more suitable in detecting masks.
3. MobileNet copes with training that requires a considerable number of resources.

References

AG Howard et al.. (2017). "MobileNets: Efficient Convolutional Neural Networks13for Mobile Vision Applications,".

AM Lad, A. Mishra, and A. Rajagopalan. (2021). "Comparative Analysis of Convolutional Neural Network Architectures for Real Time COVID-19 Facial Mask Detection," J. Phys. conf. Ser., vol. 1969, no. 1.

C. Jagadeeswari and MU Theja. (2020). "Performance Evaluation of Intelligent Face Mask Detection System with various Deep Learning Classifiers Keywords :," Int. J. Adv. science. Technol., vol. 29, no. 11, pp. 3074–3082.

H. Ignatius, R. Chandra, N. Bohdan, and A. Dharma. (2019). "COMPARISON OF CONVOLUTIONAL NEURAL NETWORK MODEL IN CLASSIFICATION OF DIABETIC RETINOPATHY," J. Investigator. Post and Information., vol. 9, no. 2, pp.

K. He, X. Zhang, S. Ren, and J. Sun. (2016). "Deep residual learning for image recognition," Proc. IEEE Computing. soc. conf. Comput. vis. Pattern Recognition., vol. 2016-Decem, pp. 770–778.

K. Madineni and C. (2020). Science, "Handwritten Text Recognition Based On CNN Classification," vol. 29, no. 9, pp. 6870–6880.

K. Nithiyasree and T. Kavitha. (2021). "Face Mask Detection in Classroom using Deep Convolutional Neural Networks," Turkish J. Comput. Math. Educ., vol. 12, no. 10, pp. 1462–1466.

K. Simonyan and A. Zisserman. (2015). "Very deep convolutional networks for large-scale image recognition," 3rd Int. conf. Learn. representative. ICLR 2015 - Conf. Track Proc., pp. 1–14.

L. Alzubaidi et al. (2021).Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, vol. 8, no. 1. Springer International Publishing.

L. Dang, P. Pang, and J. Lee. (2020). "Depth-wise separable convolution neural network with residual connection for hyperspectral image classification," Remote Sens., vol. 12, no. 20, pp. 1–20.

M. Liao et al.(2021). "A technical review of face mask wearing in preventing respiratory COVID-19 transmission," Curr. Opinion. Colloid Interface Sci., vol. 52, p. 101417.

R. Kanotra, Akash, N. Wadhwa, and DN Jeyanthi*. (2021). "Comparative Analysis of Object Detection Algorithms for Face Mask Detection," Int. J. Eng. Adv. Technol., vol. 10, no. 4, pp. 148–151.

S. Indolia, AK Goswami, SP Mishra, and P. Asopa. (2018). "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," Procedia Comput. science., vol. 132, pp. 679–688.

SS Yadav and SM Jadhav. (2019). "Deep convolutional neural network based medical image classification for disease diagnosis," J. Big Data, vol. 6, no. 1.

T. Singhal. (2020).Y. Feng, X. An, and S. Li. (2018). " "A Review of Coronavirus Disease-2019 (COVID-19)," Indian J. Pediatr., vol. 87, no. April, pp. 281–286.

Research on Face Recognition Based on Ensemble Learning," Chinese Control Conf. CCC, vol. 2018-January, pp. 9078–9082.