ÇUKUROVA UNIVERSITY

ENGINEERING AND ARCHITECTURE FACULTY

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

**GRADUATION THESIS**

# DETECTING COVID-19 FROM X-RAY IMAGES USING TRANSFER LEARNING ALGORITHM

**By**

2016513019 - Kamil DİNLEYİCİ

**Advisor**

Prof. Dr. Turgay İBRİKÇİ

June 2021

**ADANA**

## ABSTRACT

The coronavirus 2019 (COVID-19), which emerged in China in 2019, has turned into an epidemic and its spread continues. Various methods have been developed and continue to be developed to prevent the spread of this epidemic. These methods are also developed by working together in different work areas. By using X-ray or CT images, it is possible to make an early diagnosis of COVID-19 disease through artificial intelligence (AI). In this study, using X-ray images of 1200 COVID-19 positive and 1341 healthy people, a total of 2541 X-ray images were trained on models with high success rates called state-of-the-arts. These models were trained by using fine-tuning methods and training the model from the scratch. Among these models, MobileNetV2 achieved an accuracy of 99.01% and a specificity of 99.58% for binary classification. This study is not sufficient for the early diagnosis of COVID-19 for alone, the results must be examined in the presence of a radiologist.

## TABLE OF CONTENT

ABBREVIATIONS

**COVID-19**: Coronavirus disease 2019

**SARS**: Severe Acute Respiratory Syndrome

**MERS**: Middle East Respiratory Syndrome

**CT**: Computed Tomography

**PCR**: Polymerase Chain Reaction

**TP**: True Positive

**TN**: True Negative

**FP**: False Positive

**FN**: False Negative

**CM**: Confusion Matrix

LIST OF TABLES

LIST OF FIGURES

# 1. INTRODUCTION

The first 2019 novel coronavirus cases emerged in early December of 2019 and the number of confirmed coronavirus disease (COVID-19) cases has reached 155,506 million people as of May 7, 2021[1]. The COVID-19 has a fatality rate of 2.09%, this fatality rate is not high as Severe Acute Respiratory Syndrome(SARS) and Middle East Respiratory Syndrome(MERS). However, it has spread to more areas and it has led to more total deaths[2]. The COVID-19 patients show more slight symptoms around 99% of cases when the rest of the patients show severe or critical[4].

The mRNA vaccines for COVID-19 were produced with high impacts(94-95%) to prevent the spread of COVID-19 and its multiple variants(SARS-CoV-2 variants). Despite the vaccination has started the spreading speed has been increased with time and early detection of COVID-19 is still keep its importance. The PCR(Polymerase Chain Reaction) test is used as the golden standard method to detect COVID-19 but the PCR test is time-consuming and has a high false-negative rate and rare[3]. Radiologic screening methods like CT(Computed Tomography) or CXR(Chest X-ray) are useful in the early detection of COVID-19 and decided to utilize with PCR test[5]. Computed Tomography(CT) gives more accurate results rather than other chest imaging techniques in the early diagnose of COVID-19. However, chest X-ray is cheaper and more accessible compared to CT [10]. Also, it is shown that CXR is less harmful than CT through its lower radiation effect. Although CXR looks more advantageous, to finalize the diagnosis successfully chest X-rays require expert radiologist knowledge.

Medical images are raw materials of machine-learning based deep learning applications to diagnose health cases. One of the important problems that will be encountered when working with health data is that the data is imbalanced. It occurs when there is no enough healthy data and less diseased data(or vice versa)[6]. The other problem is the small dataset for medical imaging with computer vision and there is a special case that is the sudden outbreak of COVID-19. Due to the personal nature of the x-ray images, the sharing of these data with the public has progressed very slowly, and since the first months of the virus emergence, it has been necessary to work with a small number of data.

Deep learning algorithms generally give a better accuracy ratio when the dataset is large[7]. Despite small dataset is used Transfer Learning shows up, it provides to work on insufficient datasets in other words small datasets. Thanks to state-of-art models that are trained on the ImageNet dataset(includes 1M images) we can utilize the robust filters learned by pre-trained models and their weights. There are two types of transfer learning concepts: feature extraction and fine-tuning, these concepts are genuinely solid and useful. I will use both of these concepts to predict with high accuracy.

### *1.1. Feature extraction*

The Transfer Learning method allows us to create the convolutional neural network model we want to create more quickly with a small amount of data and one of the important factor is feature extraction. When constructing a convolutional neural network model, we can use the properties of the pre-trained model instead of starting from scratch. Replacing the classifier part of pre-trained model with new classifier is enough to make feature extraction. Feature extraction enables reducing requested data and computing resources.

### *1.2. Fine-tuning*

The model that we have created start to learn basic shapes in the very first layers as we go deeper layers the model learns more complicated shapes and the model become more specialized for data. We can decide which layers we are going to start the model learning process and this is called fine-tuning.

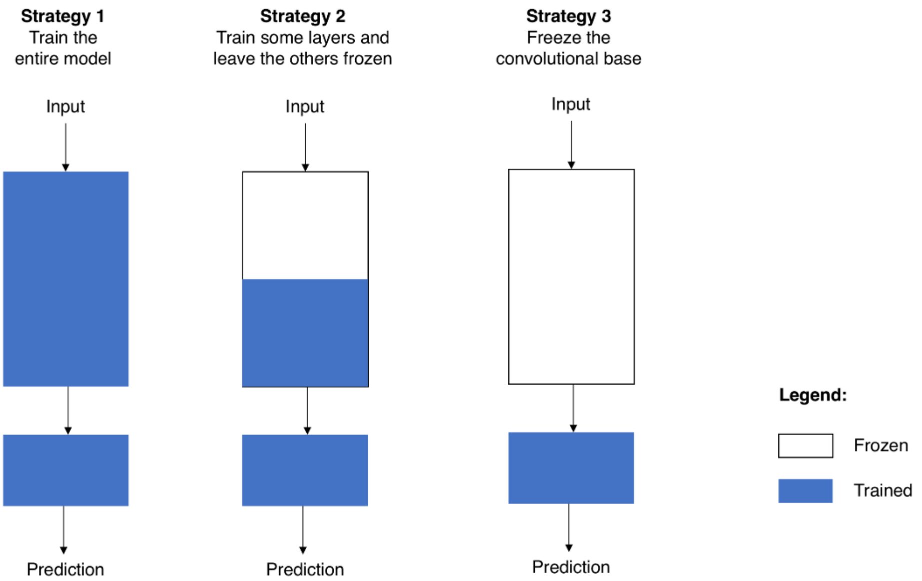


Figure 1. Some fine-tuning strategies

### *1.3. Models*

State-of-the-arts models are used because they provided sufficient benefits like high accuracy and flexibility on changing the number of layers. In this study, some of the models are trained from scratch and others are trained from last layers of bottom layer of state-of-the-arts model.

#### *1.3.1. VGG16*

This model is proposed by K. Simonyan and A. Zisserman with the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition” [15]. VGG16 consists of 16 convolutional layers and the model achieved 92.7% top 5-test accuracy in ImageNet. The default input size is 224x224 but this can be changed if the classifier of the model is removed. The number of class is 1000.

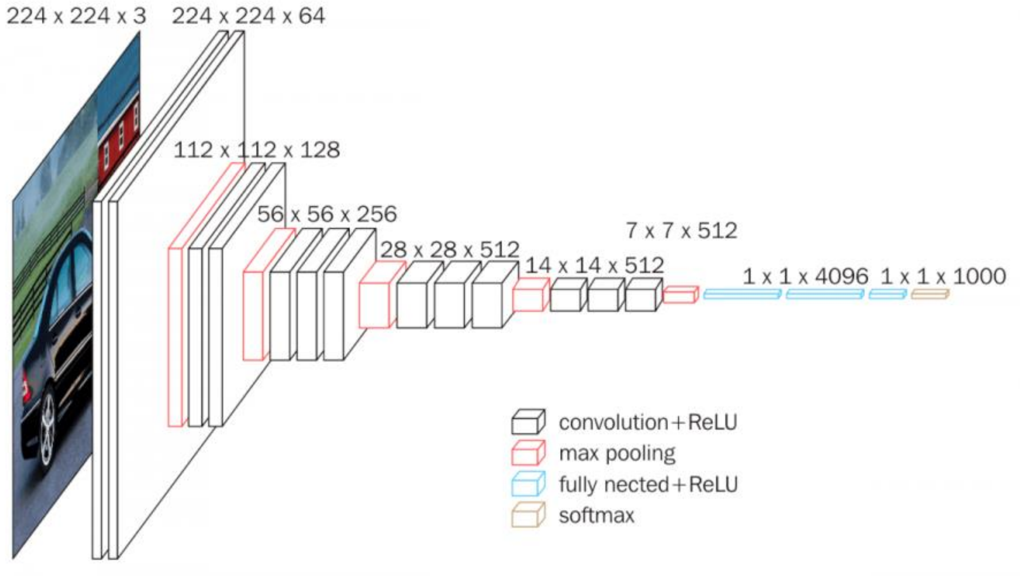


Figure 2. An overview of VGG-16 model

#### *1.3.2. VGG19*

VGG19 is a deep learning model that consists of 5 pooling layers, 16 convolutional networks and 3 fully-connected layers and is defined by ReLU activation functions.

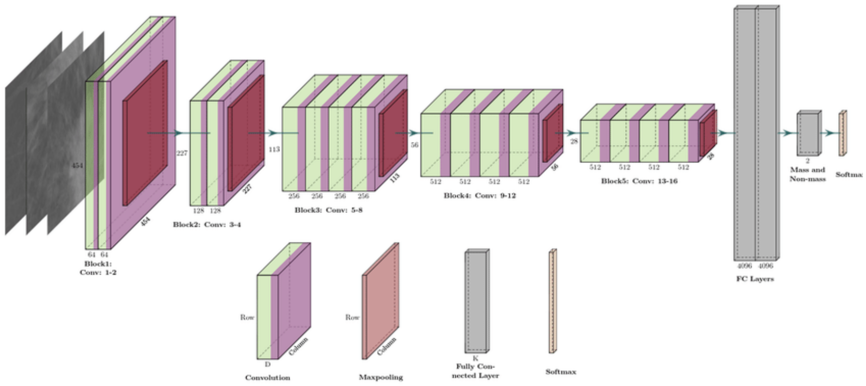

Figure 3. VGG19 model

### *1.3.3. InceptionV3*

This model is firstly showed up at “Rethinking the Inception Architecture for Computer Vision(CVPR 2016)” paper[16]. The only feature that distinguishes it from Inception v2 is the addition of not only convolution layers, but also batch-normalized (batch-normalized) and fully connected (FC) layers as auxiliary classifiers. Optionally it can be used with weights which is pre-trained on ImageNet dataset. Input size is 299x299x3 but it can be changed if the classifier is removed.

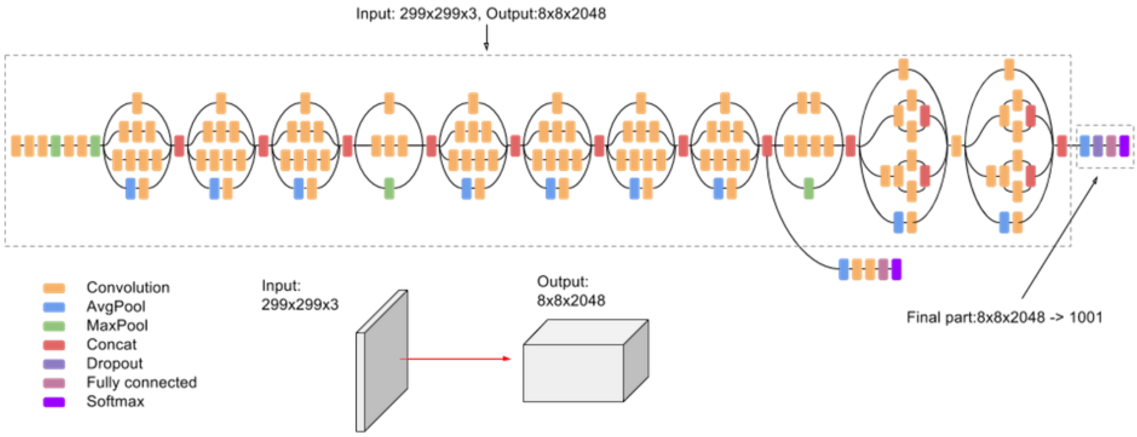


Figure 4. An overview of InceptionV3 model

### *1.3.4. ResNet50*

Resnet50 is a 50-layer network trained on the ImageNet dataset. Instead of using 2 (3x3) convolutions, the Resnet model uses convolution layers (1x1), (3x3), (1x1).

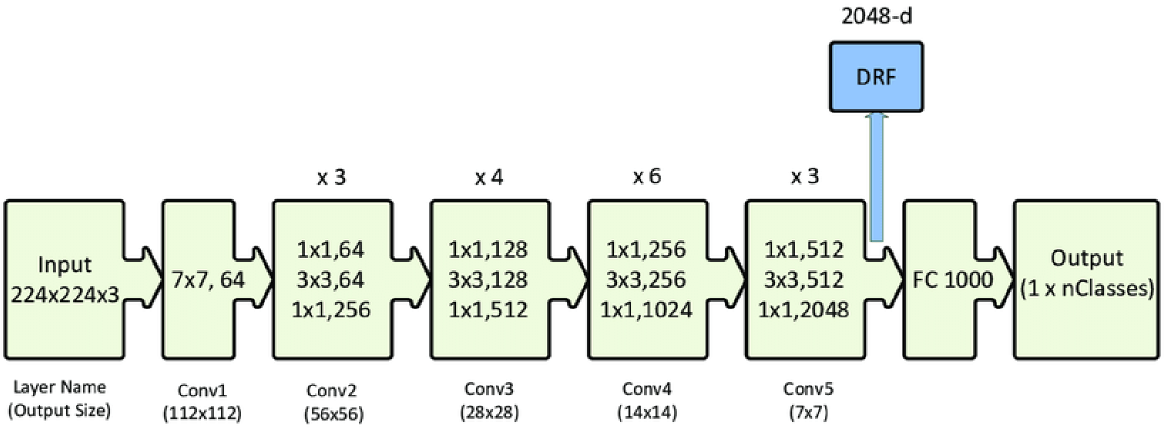


Figure 5. ResNet50 schematic

### *1.3.5. MobileNetV2*

MobileNetV2 was released with the paper “MobileNetV2: Inverted Residuals and Linear Bottlenecks”[17] in 2018. It reached better performance values than its previous version, MobileNetV1 [18].

### *1.3.6. InceptionResNetV2*

This model is constructed with Residual Inception Block algorithm. The input size is 299x299 by default.InceptionResNetV2 trained on ImageNet dataset and optionally the weights on dataset can be used.This CNN has 164 layers and can classify 1000 categories.

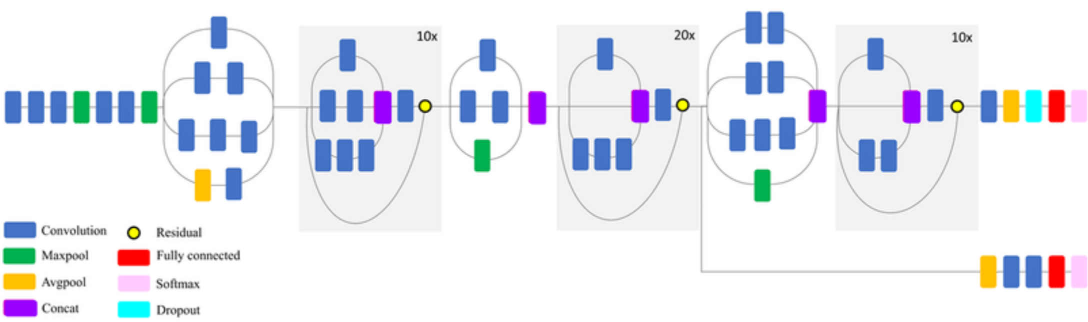


Figure 6. InceptionResNetV2 architecture

## 2. RELATED WORK AND DATASET

### 2.1. RELATED WORK

There are lots of researches which are aiming to develop accurate diagnostics of COVID-19 using binary and multi-class classification. Sethy & Behera [13] are found that ResNet50 + SVM is reached the best performance compared to their other trained models with 95.38% accuracy. Their dataset contains only 25 number of COVID-19 X-ray images and 25 number of normal(healthy) X-ray images. Although their unsufficient amount of data (totally 140 X-ray images, 100 for training purposes and 40 for test) Sahinbas and Catak [14] are reached 80% accuracy with VGG16 state-of-the-art model between other VGG19, ResNet, DenseNet and InceptionV3 models for binary classification.

Ozturk et al. [8] created a model which is inspired by the DarkNet[9] architecture. Their model produced 98.08% accuracy for binary classification(COVID vs. No-Findings) of COVID-19 and 87.02% accuracy for multi-class classification(COVID vs. No-Findings vs. Pneumonia). Narin et al. [10] proposed a diagnose using ResNet50, ResNet101, ResNet152, InceptionV3 and Inception-ResNetV2 with different binary classification. They used four classes (COVID-19 , healthy, viral pneumonia and bacterial pneumonia) and created three different binary classifications. Also, they utilized k-fold cross validation method with k=5. They observed that ResNet50 gives the highest performance for three datasets ( 96.1% / 99.5% / 99.7%). Apostopolus et. al. [11] have reached 98.75% accuracy for binary classification and 93.48% accuracy for multi-class(3) classification. They used two different datasets one of them was consisted of 224 confirmed COVID-19 X-ray images, 700 X-ray images with confirmed bacterial pneumonia and 504 X-ray images for normal(healthy). The other dataset was comprised with 224 COVID-19 images, 714 bacterial pneumonia images and 504 images for healthy conditions. State-of-the-art models(VGG-19,MobileNet v2, Inception, Xception and Inception ResNet v2) are used to make binary and multiple classifications.

A dataset includes 850 COVID-19 X-ray images, 500 pneumonia images and 915 healthy X-ray images is proposed at the study of Albahli et al. [12]. Three state-of-the-art model(Inception ResNetV2, InceptionNetV3 and NASNetLarge) are used with fine-tuning method of transfer learning algorithm. The InceptionV3 model was gave the best performance with 98.63% accuracy when the data augmentation is not used. The accuracy of 99.02% is acquired when data augmentation is used.

2.2. DATASET

Ozturk et al. [8] used 125 COVID-19 X-ray images. Narin et al. [10] used 341 COVID-19 X-ray images with various number of other diseases and healthy X-ray images in their study. Apostopolus et. al. [11] worked with 3905 X-ray images and these images was consisting of 6 different diseases and 455 X-ray images were COVID-19.

In this study,the increased number of COVID-19 X-ray images are utilized. Totally 2541 X-ray images which contains 1200 COVID-19 positive X-ray images and 1341 healthy X-ray images. The dataset is splitted 20% for test and 80% for train. After that split the train data is splitted again as 20% validation and 80% train data. The dataset is avaible at here: [https://www.kaggle.com/kamildinleyici/cov-dataset].

This dataset was composed from another Kaggle dataset which is added here: [https://www.kaggle.com/tawsifurrahman/covid19-radiography-database]. Just the one of the first version of dataset is used after our data set created the new versions are added. This version is found more effective.

The increased number of COVID-19 positive images has solved the data imbalance problem.

Figure 7. X-ray image

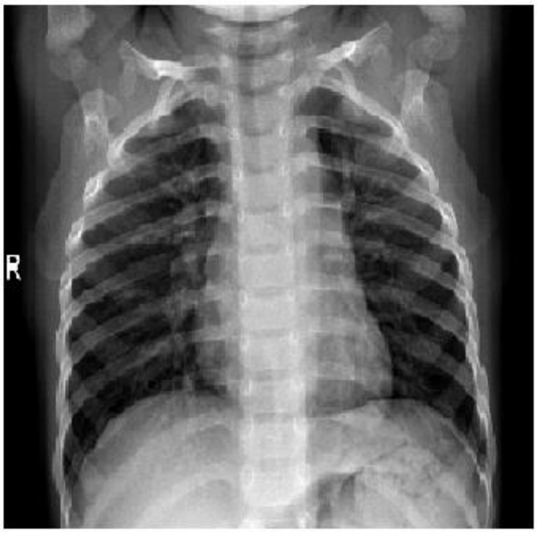

Figure 8. X-ray image

Figure 7. shows(left) COVID-19 positive case and Figure 8. shows(right) COVID-19 negative case(healthy)

# 3. METHODOLOGY

## 3.1. IMPLEMENTATION OF TRANSFER LEARNING TO CNNS

The Convolutional Neural Network(CNN) is inspired from the movement of the neurons in the human brain and used to differentiate an image from another. It is basically an Deep Learning algorithm which can detect features such as curves and edges. In other terms, CNN is an artificial neural network that can detect patterns and make these patterns useful.

CNN processes the image with different layers such as convolutional layer, pooling layer, flattening layer and fully-connected layer. Convolutional layer applies some filters to the image to extract low and high level features from the image. Pooling layer is to reduce the spatial size of the representation and the number of the parameters and calculations within the network. Flattening layer is simply to prepare the data for the most important layer, fully connected layer. Generally, neural networks take input data from a one-dimensional array. The data in this neural network is the one-dimensional array of matrices from the convolutional and pooling layers. Fully-connected layer is the last and the most important layer of CNN. It takes the data from the flattening layer and performs the learning process through the neural network.

Transfer Learning method on CNNs is showed a big impact on classification problems. Dataset size is no longer problem with the help of Transfer Learning because small datasets can be useful using these datasets with pre-trained models. There are few options to use Transfer Learning effectively.

Some of the models are trained from the scratch in this paper. It is observed that changing the fine-tuning methods also changes the accuracy and other metrics.

## 3.2. MODEL STRUCTURE

The proposed models are constructed using pre-trained models and the new classifiers that is changed with the previous pre-trained model's classifiers. Firstly, image augmentation is applied to the input X-ray images. As it is shown in Figure 9., after pre-trained models GlobalAveragePooling2D, dense and dropout layers are added. At the end there is a dense layer to binary classification to detect if the image is COVID-19 positive or healthy.

Dropout layer is added to prevent overfitting issue. Overfitting is a big problem for classification processes and it occurs when the model learns almost every path in training dataset. This situation causes to low validation accuracy ratio because the model memorizes only the training dataset. When any other image is given as input the model will not recognize successfully.

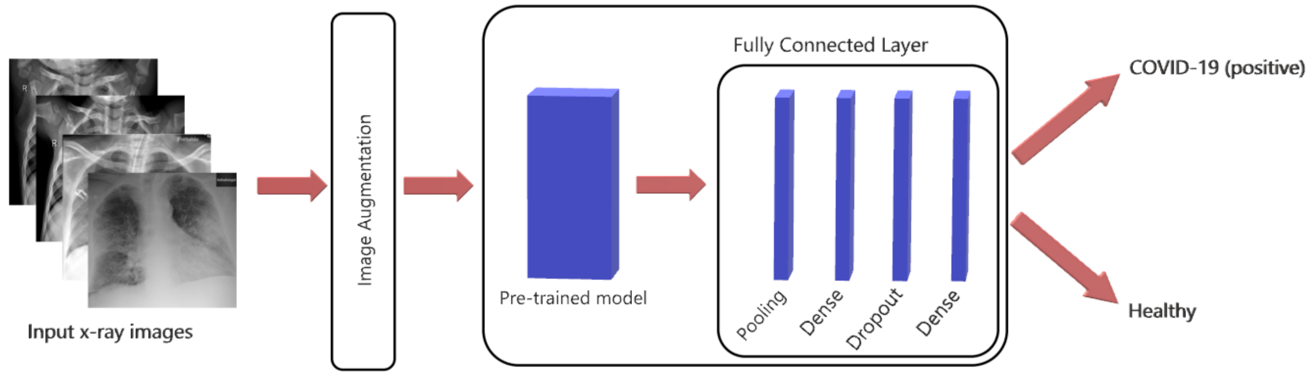


Figure 9. The model structure

## 4. PROCESSING ON CNN MODELS

All the training is done at Google COLAB and Kaggle is also used to get data. The models are trained with batch size of 32 and learning rate is 1e-5. Adam optimizer is used and data augmentation is applied. Epoch is 50 for every models.

The models were trained using various transfer learning methods such as feature extraction, fine-tuning and training the model from the scratch. For this paper, 6 state-of-the-arts models are trained. It is available to see how they are trained in Table 1.

| MODEL | TRANSFER LEARNING METHOD |
|---|---|
| VGG16 | Feature extraction(Only classifier is trained) |
| VGG19 | Fine tuning(The model is trained from 19th layer) |
| InceptionV3 | Trained from the scratch |
| ResNet50 | Trained from the scratch |
| MobileNetV2 | Trained from the scratch |
| InceptionResNetV2 | Feature extraction(Only classifier is trained) |

Table 1. The models and how they are trained

### 4.1. PERFORMANCE MEASUREMENTS

Sensivity, specificity and accuracy are the performance measurements for this paper. There is a confusion matrix for every model to observe performance except the accuracy at models.

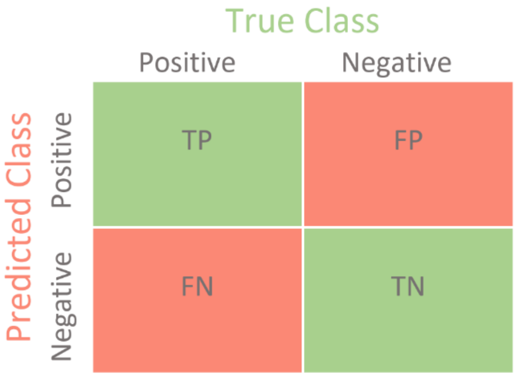


Figure 10. Confusion matrix

**Sensivity** = TP / (TP + FN)

**Specificity** = TN / (TN + FP)

**Accuracy** = TP + TN / (TP + TN + FP + FN)

5. RESULTS

The state-of-the-arts models(VGG16, VGG19, InceptionV3, ResNet50, MobileNetV2, InceptionResNetV2) which are used in study are trained with fine-tuning methods and from the scratch. To make the study more clear the best score among these methods is written.

From Figure 11. to Figure 16. the training, validation accuracy and loss graphics are shown. Although the best score is used some of the models tend to overfit.

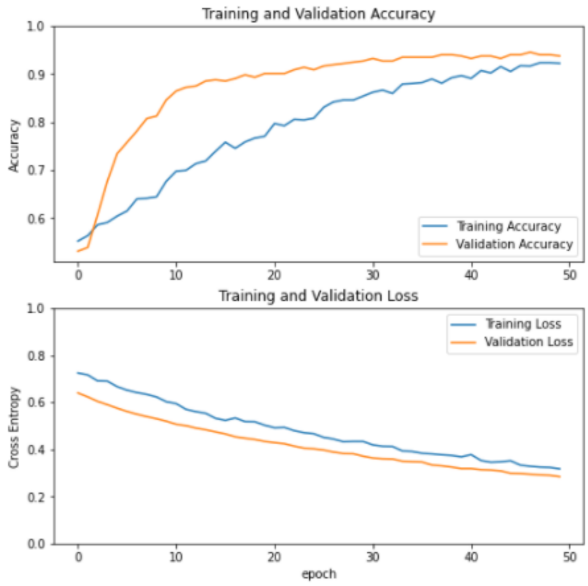


Figure 11. VGG16

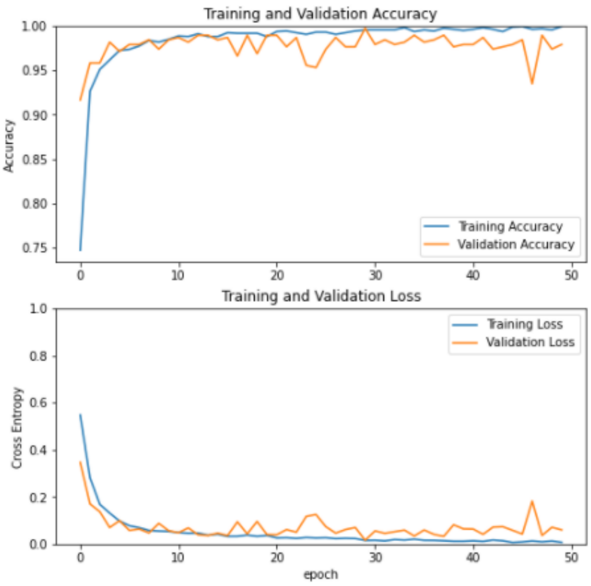


Figure 12. VGG19

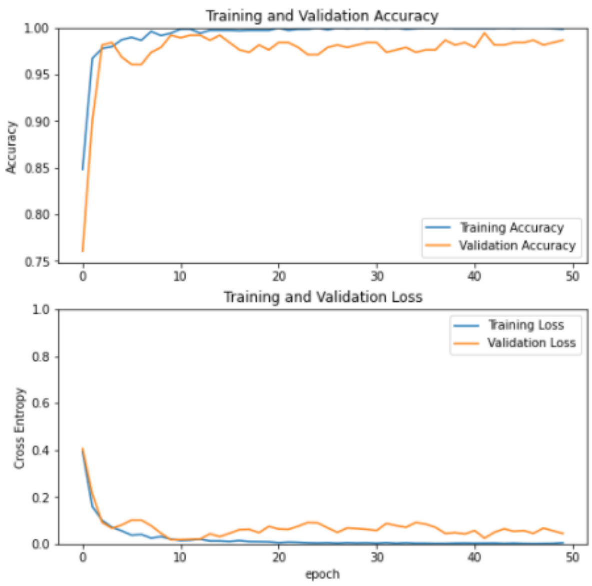


Figure 13. InceptionV3

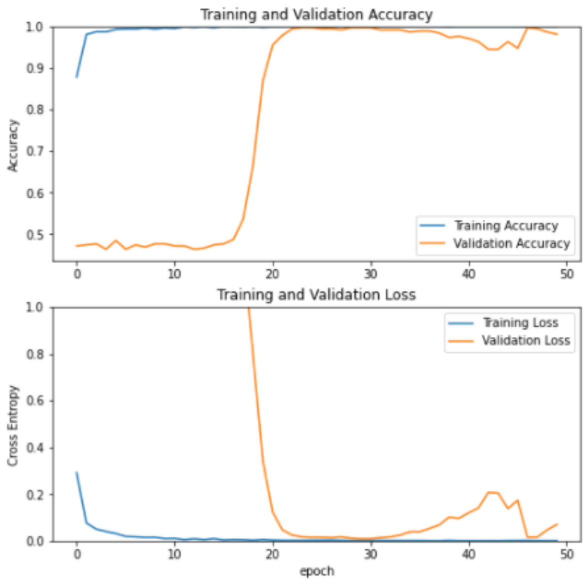


Figure 14. ResNet50

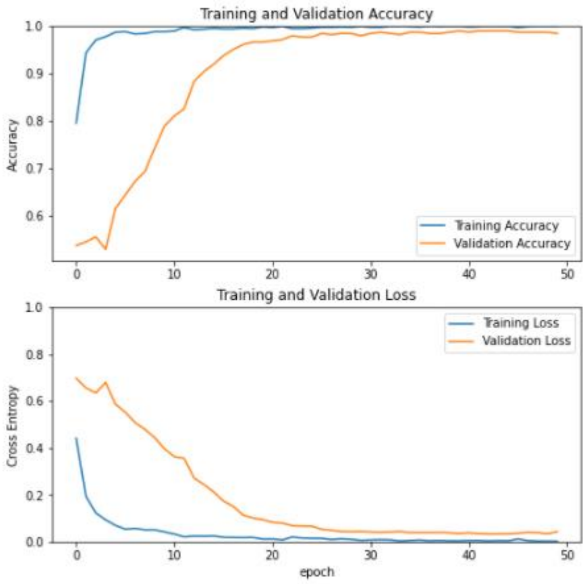


Figure 15. MobileNetV2

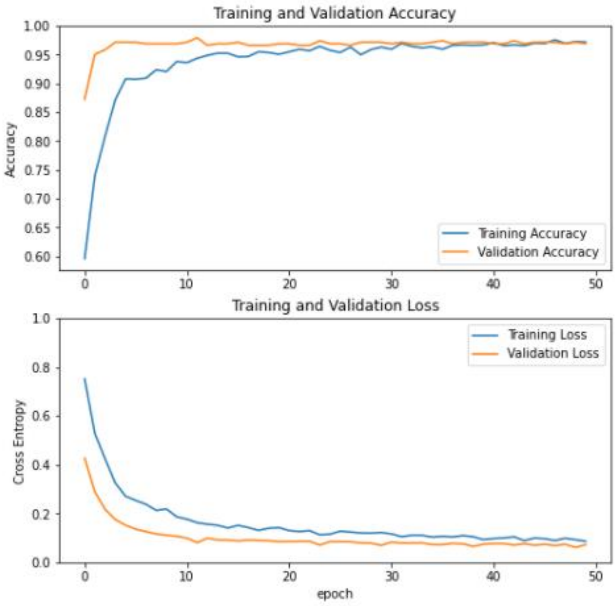


Figure 16. InceptionResNetV2

Confusion matrices are created to see the performance of model over dataset.

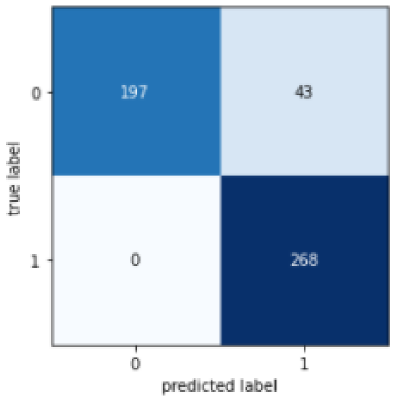


Figure 17. VGG16

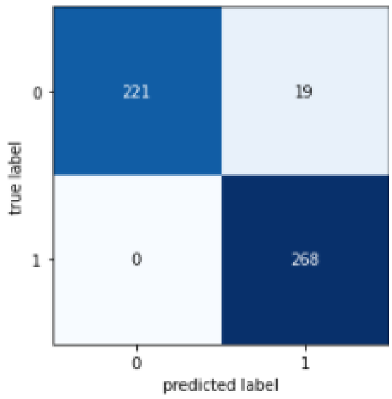


Figure 18. VGG19

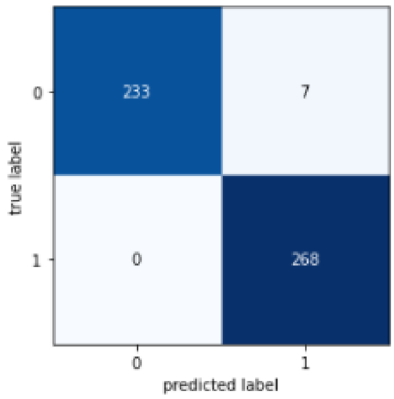


Figure 19. InceptionV3

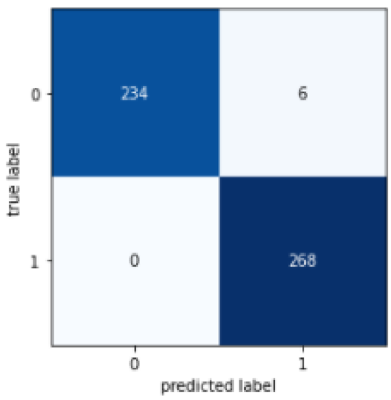


Figure 20. ResNet50

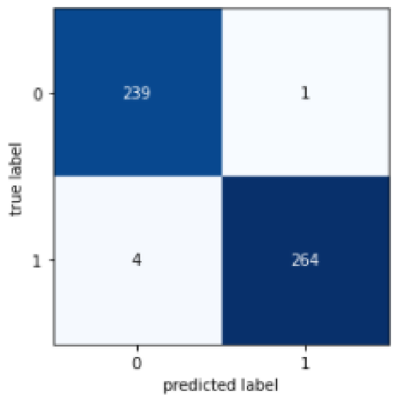


Figure 21. MobileNetV2

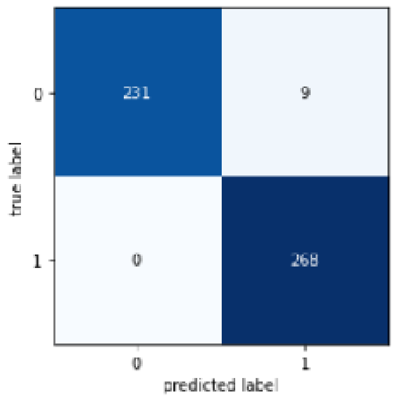


Figure 22. InceptionResNetV2

Binary classification results with different models are given in Table 2.

| MODEL | ACCURACY | SENSITIVITY | SPECIFICITY |
|---|---|---|---|
| VGG16 | 0.9153 | 1 | 0.8208 |
| VGG19 | 0.9625 | 1 | 0.9208 |
| InceptionV3 | 0.9862 | 1 | 0.9708 |
| ResNet50 | 0.9881 | 1 | 0.9750 |
| MobileNetV2 | **0.9901** | 0.9850 | **0.9958** |
| InceptionResNetV2 | 0.9822 | 1 | 0.9625 |

Table 2. Accuracy,sensitivity and specificity scores

6. DISCUSSION

As seen in the previously shared article [19], the accuracy has increased with the increasing number of data. It has been stated before[11] that training the models from the scratch will be beneficial for some models, and it has been revealed that the number of data sets increased with this study also contributes positively to the performance of the model.

Compared to similar studies, this paper shows the impact of the size of dataset. Using the increased number of X-ray images and MobileNetV2 model it is obviously shown that accuracy and specificity scores are relatively higher than the researches that shares the same aim(Table 3). Although accuracy was expected to be higher than other studies, it was observed that it remained lower. However, specificity score is remarkably higher than other studies.

To sum up, in future studies more state-of-the-arts models should be trained by working with data sets whose number has been increased even more.

| AUTHOR | DATA STRUCTURE AND SIZE | BEST MODEL | ACCURACY | SENSIVITY | SPECIFICITY |
|---|---|---|---|---|---|
| Apostolopoulos et al. (2020)[11] | 455 COVID-19 patients; 910 viral pneumonia; 2,540 other pulmonary diseases (X-ray images) | MobileNet v2 | 99.18% | 97.36% | 99.42% |
| Sethy et al. (2020)[13] | 25 COVID-19 patients; 25 Healthy | ResNet50 plus SVM | 95.38% | 97.29% | 93.47% |
| Vaid et al. (2020)[20] | 181 COVID-19 patients; 364 Normal patients (X-ray images) | VGG-19 | 96.3% | 97.1% | - |
| Waheed et al. (2020)[21] | 403 COVID-19 patients; 721 normal patients (X-ray images) | VGG16 | 95% | 90% | 97% |
| **Proposed Study** | 1200 COVID-19 positive X-ray images; 1341 healthy X-ray images | MobileNet v2 | **99.01%** | 98.50% | **99.58%** |

Table 3. Other similar researches

## 7. CONCLUSION

The importance of using large datasets has emerged in this research, as in previous state-of-the-arts models. Compared to previous researches the bigger dataset has shown better performance with MobileNetV2 model(Table 3).

The successful operation of these models will not give healthy results on their own, a definite result should be reached after being examined by a radiologist. Instead making binary classification multiple classification is also possible and can be utilized to seperate the COVID-19 positive,healthy and other diseases.

REFERENCES

1. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/

2. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA. 2020 Apr 7;323(13):1239-1242. doi: 10.1001/jama.2020.2648. PMID: 32091533.

3. Chan, J. F. W.; Yuan, S.; Kok, K. H.; To, K. K. W.; Chu, H.; Yang, J.; ... and Tsoi, H. W. 2020. A Familial Cluster of Pneumonia As-sociated with the 2019 Novel Coronavirus Indicating Person-To-Person Transmission: A Study of a Family Cluster. *The Lan-cet*, *395*(10223), 514-523.

4. McKeever A. Here's what coronavirus does to the body. National Geographic: https://www.nationalgeographic.com/science/2020/02/here-is-what-coronavirus-doesto-thebody/, 2020.

5. Zi Yue Zu, Meng Di Jiang, Peng Peng Xu, Wen Chen, Qian Qian Ni, Guang Ming Lu, and Long Jiang Zhang. Coronavirus Disease 2019 (COVID-19): A Perspective from China Radiology 2020 296:2, E15-E25

6. https://ayyucekizrak.medium.com/medikal-g%C3%B6r%C3%BCnt%C3%BClerde-bilgisayarl%C4%B1-g%C3%B6r%C3%BC-3b7dd2abe641

7. https://community.nxp.com/t5/eIQ-Machine-Learning-Software/Transfer-Learning-and-the-Importance-of-Datasets/ta-p/1109248

8. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med. 2020 Jun;121:103792. doi: 10.1016/j.compbiomed.2020.103792. Epub 2020 Apr 28. PMID: 32568675; PMCID: PMC7187882.

9. J. Redmon, A. Farhadi, Yolo9000: Better, Faster, Stronger, 2017, arXiv:1612.08242.

10. Narin, A., Kaya, C. & Pamuk, Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *Pattern Anal Applic* (2021). https://doi.org/10.1007/s10044-021-00984-y

11. Apostolopoulos ID, Aznaouridis SI, and Tzani MA. Extracting possibly representative COVID-19 Biomarkers from X-Ray images with Deep Learning approach and image data related to Pulmonary Diseases. Journal of Medical and Biological Engineering, 40:462-469, 2020.

12. Albahli S, Albattah W. Detection of coronavirus disease from X-ray images using deep learning and transfer learning algorithms. J Xray Sci Technol. 2020;28(5):841-850. doi: 10.3233/XST-200720. PMID: 32804113; PMCID: PMC7592683.
13. Sethy, P.K.; Behera, S.K. Detection of Coronavirus Disease (COVID-19) Based on Deep Features. Preprints 2020, 2020030300 (doi: 10.20944/preprints202003.0300.v1).
14. Sahinbas K, and Catak FO. Transfer Learning Based Convolutional Neural Network for COVID-19 Detection with X-Ray Images. https://www.ozgurcatak.org/publication/2020-deep-learning-covid19/
15. Simonyan, K. and Zisserman, A., 2014. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1409.1556>
16. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2015. *Rethinking the Inception Architecture for Computer Vision*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1512.00567> [Accessed 1 June 2021].
17. Sandler, Mark, et al. “MobileNetV2: Inverted Residuals and Linear Bottlenecks.” *ArXiv.org*, 21 Mar. 2019, arxiv.org/abs/1801.04381.
18. https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c
19. https://community.nxp.com/t5/eIQ-Machine-Learning-Software/Transfer-Learning-and-the-Importance-of-Datasets/ta-p/1109248
20. Vaid S, Kalantar R, Bhandari M. Deep learning COVID-19 detection bias: accuracy through artificial intelligence. Int Orthop. (2020) 44:1539–42. doi: 10.1007/s00264-020-04609-7
21.Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR. CovidGAN: data augmentation using auxiliary classifier GAN for improved Covid-19 detection. IEEE Access. (2020) 8:91916–23. doi: 10.1109/ACCESS.2020.2994762