**Pneumonia detection from COVID-19 Chest X-Ray**

**Abstract**

Pneumonia is a lung infection that causes tissue inflammation in one or both lungs. It's usually caused by a bacterial infection. A virus, such as the coronavirus, can potentially cause it (COVID-19). It can be detected by chest X-rays. Radiologists can detect whether the patient has pneumonia or not. But in the age of technology, there are some other ways. A computer-aided detection system (CAD) is one of them. It can identify pneumonia on Chest X-rays (CXRs). However, existing CAD systems for identifying pneumonia from Covid-19 patient chest x-rays are not that trustable. Nowadays, it’s getting better and better. The performance of the CAD was evaluated using the area under the receiver operating characteristic curves (AUCs) that were compared with standard results of the RT-PCR and the presence of findings of pneumonia on chest X-Rays. Our study is to diagnose pneumonia from Covid-19 chest X-rays and evaluate the effectiveness of the pretrained model of Convolution Neural Networks (CNNs). The dataset contains 5800+ Chest X-Ray images that are labelled with virus, bacteria or ARDS from Covid-19 or healthy individuals. In this project we tried identification of Pneumonia from Chest X-Ray images based on different pre-trained deep learning algorithms and fine-tuned to maximize detection accuracy and identify the best algorithm. Deep learning is very useful with X-Ray images. We tried four models and collected their results. Mobilenet\_v2 and vgg16 obtained the highest AUCs of 98% and 92%. All other models were outperformed by Mobilenet\_v2. This pretrained model can be used to identify pneumonia and Covid-19 and can be significantly improved. Availability of more datasets is required for more accurate and reliable identification when using deep learning. In this pandemic period, a CAD system would be extremely beneficial.

**Introduction**

A Covid-19 patient is mainly infected in the lungs. Pneumonia is one of them. Pneumonia patients may show various symptoms after they are affected. It can be detected from chest x-rays by radiologists. But radiologists are not available everywhere and every time. Therefore, early identification is important. Using the Chest X-Ray dataset to train a Deep Learning Model to classify X-Rays of healthy and Corona Affected Patients, and then using this model to power an AI application to test for Pneumonia in a faster phase. Deep learning projects are replacing many physical systems. Deep learning is all about how a computer program can learn through observation and make decisions based on its experience. If enough datasets are available, deep learning can be used to make a meaningful system. Recent studies have reported that artificial intelligence systems using deep learning techniques can detect various diseases on CXRs, showing comparable performance to expert radiologists. In this pandemic situation, this kind of system is very important.

As these systems are AI-based, they can predict the wrong information. Deep learning-based systems for many pneumonia detection systems are available. But people still can't rely on these systems. There are many systems available that detect pneumonia from normal chest x-rays vs. pneumonia chest x-rays. As it is the time of COVID-19 and a large number of people are infected with COVID-19. We developed a system that can detect pneumonia from COVID-19 chest x-ray. We tried four architectures of CNN models: VGG16, Resnet50, Resnet101, and Mobilenet\_v2. These are all pre-trained models. We collected a dataset of COVID-19 infected with pneumonia, and the chest x-ray was collected. We gained the highest accuracy of 98% from the mobilenet\_v2 model.

**Related works**

There are many fields where deep learning can be used. Image classification is one of them and deep learning is very good for it. Many authors have already proposed several image detection techniques.

In 2016, Redmon et al. proposed YOLO, which does not require a separate region proposal network, so its detection speed is extremely fast and can reach 45 FPS. [1] In the same year, Liu et al. proposed the SSD algorithm. [2] Both SSD and YOLO win in detection speed, but SSD uses a multiscale feature map to detect independently, the spatial resolution of images in deep networks has been significantly reduced, and it may not be possible to locate small targets that are difficult to detect in low resolution, reducing the accuracy of detection. YOLO does not use multiscale feature maps for independent detection. It polishes the feature map and splices it with another lower-resolution feature map, but it treats the detection only as a regression problem and the detection accuracy is low. In 2014, Girshick et al. proposed R-CNN, which greatly improved the speed of training. On the PASCAL VOC 2010 dataset, the map improved from 35.1% to 53.7%. In 2015, Ren and others proposed the Faster R-CNN algorithm, which uses RPN (region proposal network) to generate proposals on the feature map. [3]

In 2018, Lee et al. proposed DetNet, which was designed specifically for target detection and achieved better detection results with fewer layers. [4] To avoid the large computational complexity and memory consumption caused by the high-resolution feature map, the network adopts a low-complexity dilated bottleneck structure; a higher resolution of the feature map is ensured while obtaining a higher subtractive field. This paper draws on the idea of DetNet and the framework of Faster R-CNN to study the detection of pneumonia.

In 2017, Rajpurkar et al. [10] proposed a classical deep learning network named DenseNet-121, which was a 121-layer CNN model to accelerate the diagnosis of pneumonia. [5] In contrast to experienced doctors, the framework obtained a higher F1 score. Besides, in order to alleviate the effect of imbalanced classes, the team introduced Weighted Binary Cross-Entropy loss, whose difference between the Binary Cross Entropy loss was the different weights of imbalanced classes according to the number of each class. However, the proposed loss did account for the different training difficulty levels of classes. In order to solve the problems of poor generalization ability caused by over-fitting and the problem of spatial sparseness caused by ordinary convolution operations, residual connection networks and dilated convolution were used by Liang et al. in the backbone network model. [6] The final recall rate and F1 score of their model reached 96.7% and 92.7%, respectively.

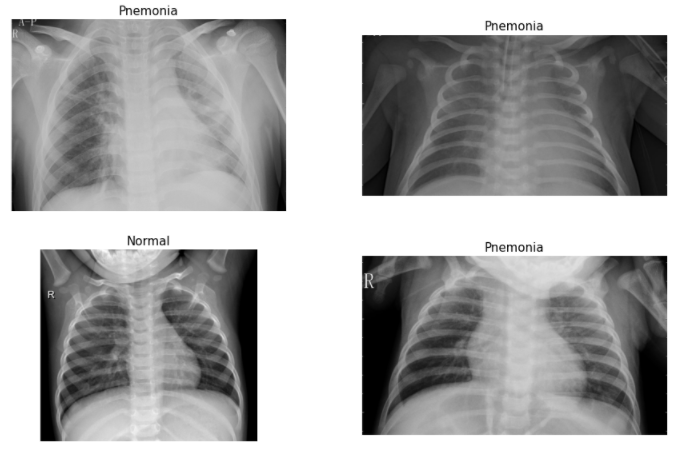
The CNN model proposed by Jain et al. combined with transfer learning effectively used the image features learned in a large dataset, sped up the training procedure of the model and made it more difficult to fall into local minimum points. [7] In addition, two models were proposed to train. Moreover, the dataset used by Jain et al. was from the world-famous organization and competition site named Kaggle, which contains numerous competitions and attracts passionate competitors to achieve a higher rank. The dataset is split into three subsets: use the training subset to train the validation subset to adjust the parameters of the model, and the test subset to verify it. Then generalization ability of the model.

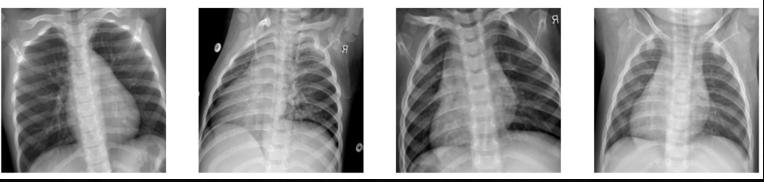
Sharma et al. and Stephen et al. devised simple CNN architectures for the classification of pneumonic chest X-ray images. [8] [9] They used data augmentation to compensate for the scarcity of data. Sharma et al. obtained a 90.68% and Stephen et al. a 93.73% accuracy rate on the dataset provided by Kermany et al., hereafter called the Kermany dataset. [10] Data augmentation, however, provides only a limited amount of new information from which the CNNs can learn and thus may not significantly boost their performance.

**Methodology**

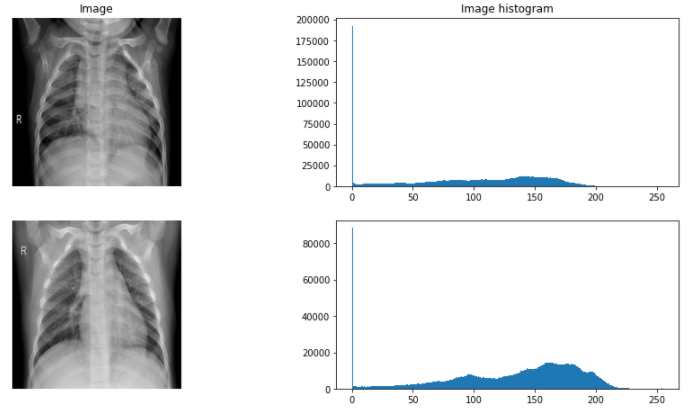
**Data Preprocessing:**

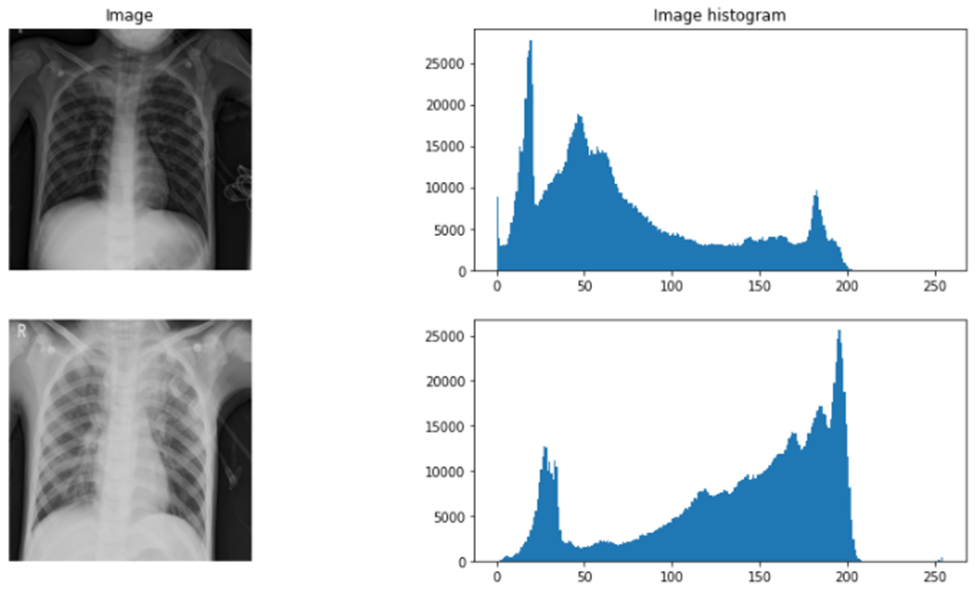
As our dataset is small, augmentation is used to prevent overfitting. Images are also normalized and downsized (224,224). There are various image enhancement methods based on histogram equalization that increase image contrast to make non-linearities more distinguishable. An example of enhancement:





**Comparison between original and enhanced images:**

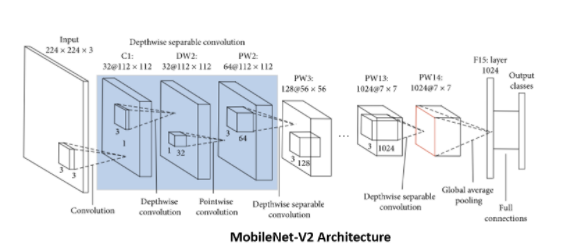
****



**Proposed Method:**

In this section, model development is explained in different steps. At first, a base convolutional model is designed and trained on different portions of the dataset. Then, pretrained models based on the Mobilenet\_v2 dataset are discussed. Finally, a pretrained model on a similar image type is explained.

The image illustrates the architecture of mobilenet\_v2:



Transfer learning is to benefit from a pretrained model in a new classification task. Some pretrained models are trained on millions of images for many epochs and achieve high accuracy on a general task. We chose the mobilenet\_v2 architecture, which is based on Keras. It is worth mentioning that pretrained models are used for fine-tuning, training on target datasets for a small number of epochs, instead of retraining for many epochs. Since ImageNet images and labels are different from the CXR dataset, a pretrained model on the same data type should also be considered.

**Classification evaluation Metrics:**

In this subsection, several evaluation metrics, accuracy, precision, recall and so on, are described. According to the outputs of the model, four indices, True Positive, True Negative, False Positive, and False Negative, are used to analyze and identify the performance of the model. The True Positive means that the chest X-ray images, which suffer from pneumonia, are signed as pneumonia as well by the model. The True Negative means if the chest X-ray images do not show pneumonia as well as the model predicts. The metrics are given as follows:

Accuracy =TP + TN/TP + TN + FP + FN

Precision =TP/TP + FP

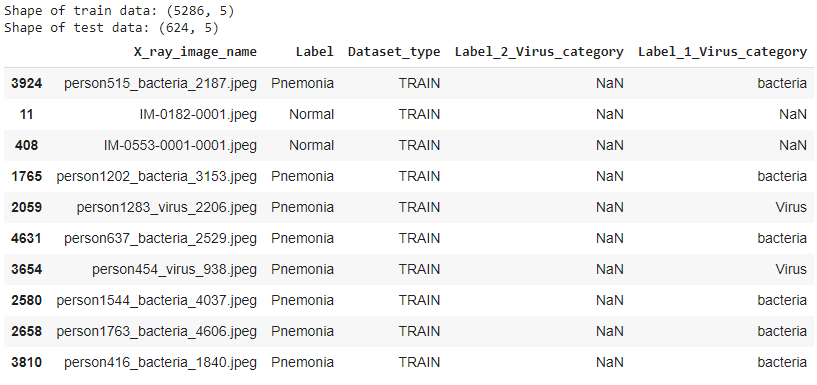
Recall =TP/FP + FN

**Experiments**

**Dataset:**

We collected a dataset from kaggle. It is a public dataset. It is Collection of Chest X Ray dataset as jpeg images. It has healthy people and people affected by Pneumonia (Bacteria or Viral) infections There are 5800+ images in this dataset. [11]

**Shape of the dataset is given below:**



**Convolutional Neural Network and Models:**

We have chosen models such as MOBILENETV2, RESNET50, VGG16, and RESNET101.

**VGG16:** VGG16 is a convolutional neural network model. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. [12]

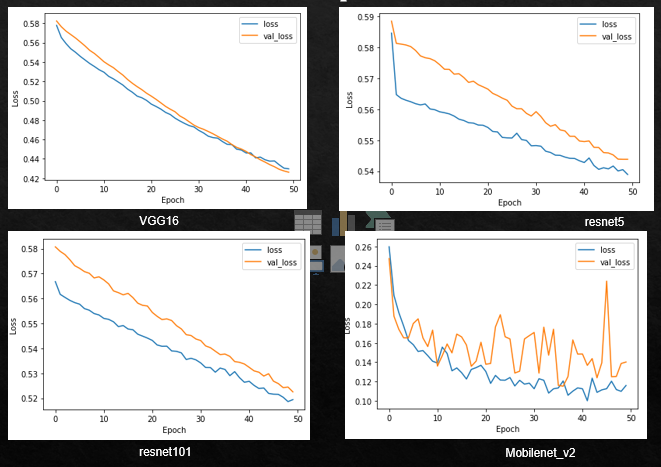
**RESNET50:** ResNet-50 is a convolutional neural network that is 50 layers deep. The pretrained network can classify images into 1000 object categories. The network has an image input size of 224-by-224. [13]

**RESNET101:** ResNet-101 is a convolutional neural network that is 101 layers deep. The pretrained network can classify images into 1000 object categories. The network has an image input size of 224-by-224. [14]

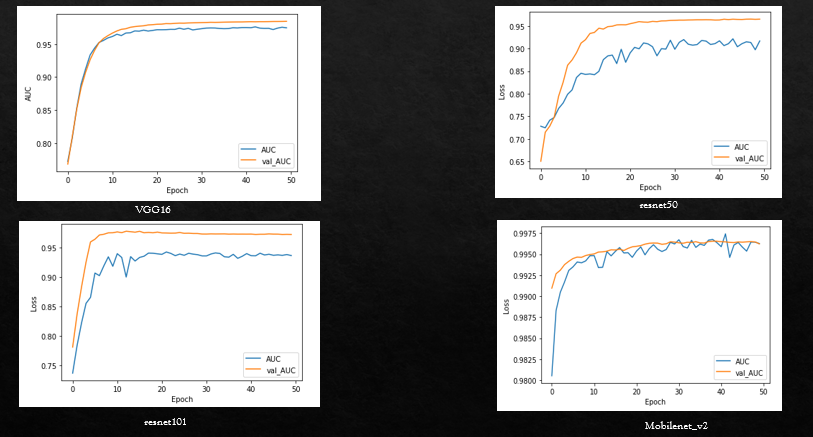
**MOBILENET\_V2:** MobileNetV2 is a convolutional neural network architecture that is optimized for mobile devices. The architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers. [15]

**Evaluation and Comparisons of the models:**

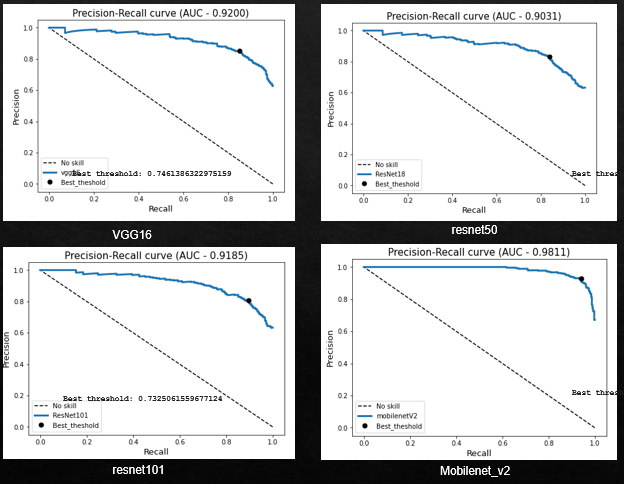
**loss vs val\_loss:**



**AUC vs val\_AUC:**



**Precision recall Curve:**

****

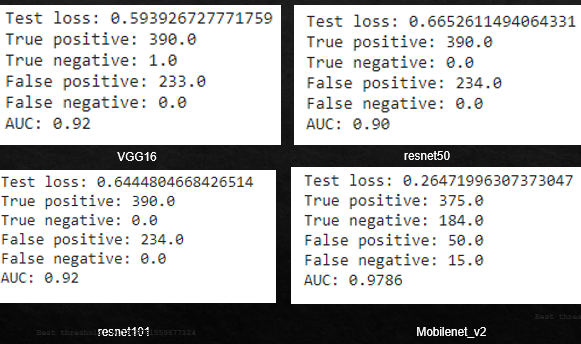
**Evaluation Metrics:**

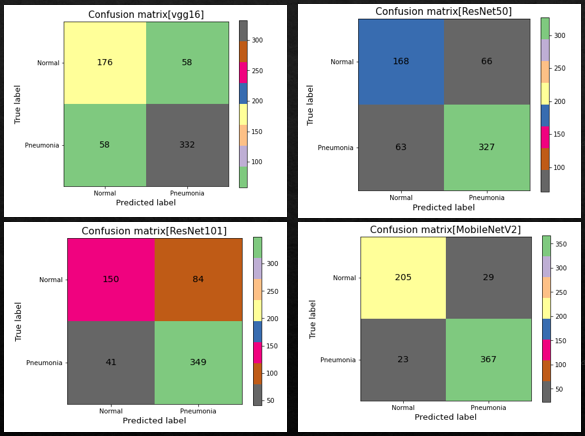
Evaluation metrics evaluate the classifier with different aims. Furthermore, all types of evaluation metrics are scalar group methods where the entire performance is presented using a single score value. Thus, it makes it easier to do the comparison and analysis, although it could mask subtle details of their behavior. In practice, the threshold and ranking metrics are the most common metrics used by researchers to measure the performance of classifiers. In most cases, these types of metrics can be employed in three different evaluation applications. [16]

The evaluation metrics were used to evaluate the generalization ability of the trained classifier. The evaluation metric is used to measure and summarize the quality of the trained classifier when tested with unseen data. Accuracy or error rate is one of the most common metrics in practice used by many researchers to evaluate the generalization ability of classifiers. Through accuracy, the trained classifier is measured based on total correctness, which refers to the total number of instances that are correctly predicted by the trained classifier when tested with the unseen data.

The evaluation metric task is to determine the best classifier among different types of training data which focus on the best future performance (optimal model) when tested with unseen data information

The accuracy metric is employed to discriminate against every single solution and select the best solution that is produced by a particular classification algorithm. Only the best solution, which is believed to be the optimal model, will be tested with the unseen data.

****

****

**Result and Discussion**

While choosing the network architecture, several different architectures such as VGG-16, ResNet50, ResNet101, and MobileNetV2 are experimented with. MobileNetV2 has performed the best among all the networks and we include the results based on that architecture. The history of the accuracy and loss of four experimented models is shown in Tab1. Below,

Comparison of Different Deep Learning Models

|  |  |  |  |
| --- | --- | --- | --- |
| **No #** | **Configuration** | **Test Loss (%)** | **Accuracy (%)** |
| **1** | **VGG-16** | **59.39** | **92** |
| **2** | **ResNet50** | **66.52** | **90** |
| **3** | **ResNet101** | **64.44** | **92** |
| **4** | **MobileNetV2** | **26.47** | **97.86** |

**Conclusion**

The prediction of pneumonia using chest X-ray prevents the spread of the disease in the chest and detects the virus faster. In this study, we train, validate, and test four popular CNN models with deep learning. We tested MobileNetV2, VGG16, Resnet50 and Resnet101 as pre-trained models to classify chest X-ray images of normal and pneumonia. The results show that the MobileNetV2 pre-trained model achieves the highest accuracy among the four models with an accuracy of 98% on the test set. The global struggle against COVID-19 is going to take us a step closer to conquering every ounce of technical innovation and intellect to combat that pandemic. COVID-19 rapid expansion and significant rates of disease and mortality have put pressure on some chest-based health systems. Extraordinary virus disinfection and preventative strategies, such as not scanning, practicing a healthy lifestyle, and preventing sick people, as well as social and psychological dissociation, have become necessary to prevent unnecessary collapse and to reduce communication potential. In order to better comprehend and handle pneumonia, AI and DL play a crucial role. ML technology allows computers to imitate human intellect and to swiftly uncover patterns and insights from vast volumes of data. AI and ML techniques, in general, are capable of dealing with pneumonia and its crisis. But a quick analysis of existing studies shows that ML is still in its early stages of combating the pandemic. Previous studies showing that AI technology is still effective at an early level against pneumonia support the findings. In detection of pneumonia, organizations immediately used their ML expertise in a variety of industries, improving patient communications, and expediting research and processing.

**References:**

|  |  |
| --- | --- |
| [1] | S. D. R. G. a. A. F. J. Redmon, "You only look once: unified, real-time object detection," in *In Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, America, 2016, pp. 779-788. |
| [2] | D. A. D. E. e. a. W. Liu, "Ssd: single shot multibox detector," in *In European conference on computer vision*, Cham, 2016, pp. 21-37. |
| [3] | K. H. R. G. a. J. S. S. Ren, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems,* p. 91–99, 2015. |
| [4] | C. P. G. Y. X. Z. Y. D. a. J. S. Z. Li, "Detnet: design backbone for object detection," in *In Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 334-350. |
| [5] | P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya and e. al., "Chexnet:Radiologist-level pneumonia detection on chest x-rays with deep learning," 2017. |
| [6] | G. Liang and L. A. t. l. m. w. a. d. r. n. f. p. p. d. Zheng, "Comput. Methods Programs Biomed," 2020. |
| [7] | R. Jain, P. Nagrath, G. Kataria, V. Kaushik and D. Hemanth, "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," 2020. |
| [8] | J. J. B. P. &. G. S. Sharma H., "Feature extraction and classification of chest x-ray images using CNN to detect pneumonia," *International Conference On Cloud Computing, Data Science & Engineering (Confluence),* pp. 227-231, 2020. |
| [9] | S. M. M. U. &. J. D. Stephen O., "An efficient deep learning approach to pneumonia classification in healthcare," *Journal Of Healthcare Engineering,* 2019. |
| [10] | Z. K. &. G. M. Kermany D., "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," 2018. |
| [11] | "Kaggle," [Online]. Available: https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset. |
| [12] | K. S. a. A. Zisserman, "VGG16 – Convolutional Network for Classification and Detection," *Very Deep Convolutional Networks for Large-Scale Image Recognition,* 2010. |
| [13] | "MathWorks," [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/resnet50.html. |
| [14] | "MathWork," [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/resnet101.html. |
| [15] | A. Z. a. L.-C. C. Mark Sandler. [Online]. Available: https://paperswithcode.com/. |
| [16] | a. P. D. N. Lavesson, "Generic Methods for Multi-Criteria Evaluation," in *in Proc. of the Siam Int. Conference on Data Mining* , USA, Siam Press, 2008, pp. 541-546. |