# Prediction of Corona Virus (COVID-19) by using X-Ray images of lungs in Orange Tool

**Misbah Akbar**
Student, Riphah International University
Alis81535@gmail.com

Arshad Sarfraz
Student, Riphah International University
Arshadsarfraz45@gmail.com

Kainat Noor
Student, Riphah International University
nkainat801@gmail.com

Muhammad Sohail
Student, Riphah International University
maliksohail284@gmail.com

Muneeb Ahmad
Student, Riphah International University
Ahmadmuneeb847@gmail.com

Shahzaib Iqbal
Student, Riphah International University
shahzaibiqbal43@gmail.com

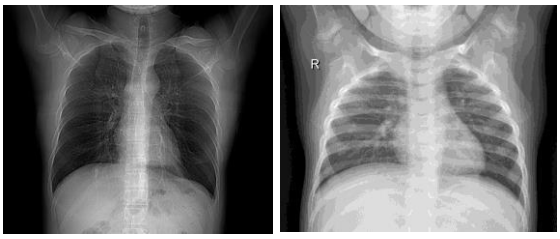
***Abstract - In 2019, Coronavirus (COVID-19) spreads all over China, and very rapidly it took the whole world under its belt and it became a pandemic [1]. Accordant to the World Health Organization (WHO), this COVID pandemic is giving the toughest time even to the best health organizations in the world. Detection of this disease at the initial level will help to recover the patient soon and also reduce the pressure of healthcare [2]. This virus is spreading day by day and our healthcare centers have a limited amount of COVID test kits and we have a huge amount of people who are having this virus [3]. It is ambitious to find people having positive test reports. So, we need a system that can detect the positive cases as soon as possible and treated the afflicted people [1]. We are living in an era in which development is rapid, so we can make an intelligent system by using deep learning algorithms that can easily predict positive COVID patients. We will apply features of deep learning neural networks on X-ray scan images of lungs and recognize whether it is affected or normal [4]. In this paper, two algorithms of the orange tool are used. The first is 'KNN' and the second is 'Neural network'. KNN stands for 'K-Nearest Neighbors', which explores for K nearest training object in component space and used their standard as prediction. The neural network is used 'sklearn's Multi-layer Perceptron algorithm', it can learn both linear and non-linear models. Dataset that is used in this paper is taken from Kaggle.***

***Keywords- Coronavirus, Orange tool, KNN, Neural network, X-ray images***

## I. INTRODUCTION

In 2019, a virus is spread in Wuhan and then it spread all over China and very rapidly it spread all over the world. This virus is very new. Its symptoms are very similar to influenza. The detection of this virus is difficult in a patient. But if it detects in its initial stage then its treatment is possible. Because if disease progresses then its treatment will not be possible. In March, 2020 World Health Organization (WHO) announced Corona Virus (COVID-19) a pandemic [4]. And healthcare originations have no enough kits to test the virus. And some countries have a lack of medical aid to cover this virus.

X-ray is the source from which disease can be predicted early and it is a cheap way as well. X-ray machines are also available in every hospital and healthcare centers. X-rays are not much harmful as laboratory tests are. Laboratory tests include discerning patients repository systems [5]. The concentrate of this research paper is to detect weather lungs X-ray helps detect the virus or not. Here are two images of X-ray, one is not and the other one is infected with COVID.




We also have some contemporary researches on these lung X-rays to diagnosis the disease. By using these datasets we can predict good sensitivity to Corona Virus [5].

In this paper, I select my dataset from Kaggle. And for prediction, I chose the Orange tool, because I found Orange is the most efficient and easiest tool. We have so many algorithms in the orange tool but in this paper, I select two algorithms 'KNN' and 'Neural Network'. We compute the accuracy of both algorithms, KNN is giving us the accuracy of 96% and Neural Network is giving the accuracy of

96% as well. But their AUC is different that we'll discuss in conclusion.

## II. LITERATURE REVIEW

In this paper [4] they want to predict the COVID-19 by using Deep Convolutional Neural Networks (DCNN). Inception V3 model is used and it gave 96% accuracy.

In this paper [3] three convolutional based on neural networks are used. Their names are ResNet50, Inception V3 and InceptionRestNetV2. Confusion matrix and ROC analyses with 5 fold cross-validation are used. RestNet50 model gives the highest accuracy of 98% in this paper.

In this paper [2] they want to make an intelligent system that can separate normal and COVID affected patients via X-rays. They apply machine or deep learning algorithms, ResNet152 is used. Measure the balance or imbalance of data points between COVID or normal patients SMOTE is used.

In this paper [6] 4 convolutional networks are used including ResNet18, DensNet121, ResNet50, and SqueezeNet. By using these algorithms they predict either X-ray is normal or COVID infected.

In this paper [5] they want to know either X-ray machines are useful in detecting the coronavirus or not and they are working on an X-ray of children that come under five years.

In this paper [5] deep neural network ResNeet50 is used with 10 fold of cross-validation. CNN classifier is used.

In this paper [1] they make a model that will give an accurate diagnosis of disease for binary classification and multi-class classification. As the classifier DarkNet model is used. Seventeen (17) convolutional layers are used and on each layer different filter is applied.

## III. PROBLEM STATEMENT

We are going through a disease that we do not have any treatment. We can only avoid it if we can follow some precautions. This virus is pandemic because it involves the whole world inside. And still, we do not have any medicine for this virus. It only can be cured In its first stage, and if it exceeds from the first stage then it will become more swear and very difficult to control or overcome this virus. First, our medical centers have not enough kits to test the patients either they are infected with coronavirus or not and they don't have safety gear as well. So we need such an efficient system that can predict the virus at its early stages so we can save more lives and also predict that how many people are infected with the virus.

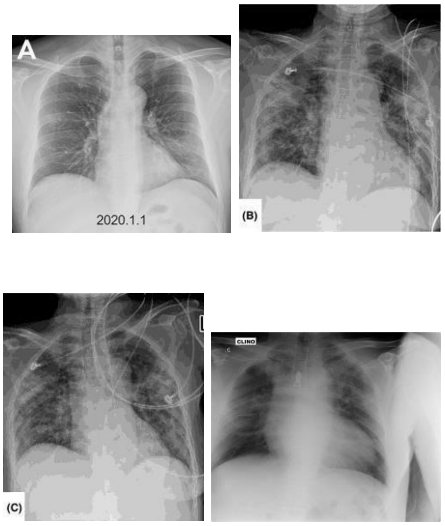
For this purpose, we select two algorithms of Orange tool, KNN and Neural Network. We also compare the accuracy of both algorithms and select the best algorithm that gives us the best accuracy.

## IV. Methodology

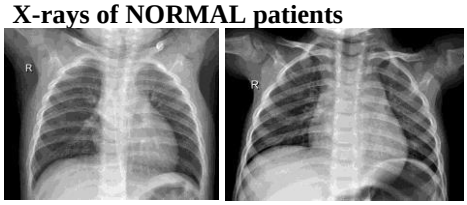
### *a. DATASET*

| Split | NORMAL | COVID |
|---|---|---|
| Training set | 28 | 70 |
| Testing set | 10 | 10 |

**X-rays of COVID patients**



**X-rays of NORMAL patients**



### *b. ALGORITHMS*

**Neural Network**

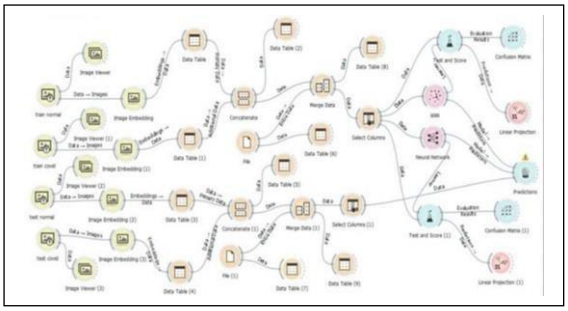
NN in the financial world help in the development of processes such as weather prediction, Disease predictions and algorithms, allocation of credit risk, and classification also. The network carries a strong affinity to statistical ways such as curve fitting and regression analysis. A neural network contains layers of complementary nodes. The simple NN consists of three layers 1st layer are the 'Input' layer, the second is the 'hidden' layer and the third layer is the 'output' layer. 1st layer takes the input design. The Output layers give us the final output of the design that was collected by the input layer. NN is active through statistical ways for example acute network errors and regression scrutiny. The neural network consists of nine intermediate components or nodes.

**KNN**

KNN algorithm is manageable and it is simple to appliance on Supervised ML algorithms which are used to solve classify and regression problems. It can suppose and collect that the alike things. It uses the feature of similarity to predict the data. Mostly it is used to solve classification problems in different industries. Two major properties give a better definition of KNN. 1st is Lazy Learning algorithm. Because KNN doesn't have special training and it can use the whole data during the training phase for classifying. 2nd property of KNN is a non-Parametric Learning algorithm because it cannot suppose anything related to fundamental data.

## V. Results

### *Experimental Results*



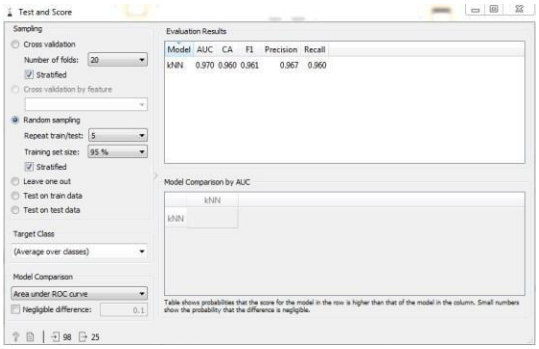
In this section we will discuss the workflow:

We devised our dataset into two different parts: the TEST dataset and TRAIN dataset. In both Train and Test dataset we have two folders named NORMAL and COVID INFECTED. In the NORMAL folder, we have X-rays of that people who don't have COVID. In the COVID folder, we have X-rays of that people who are suffering from COVID.

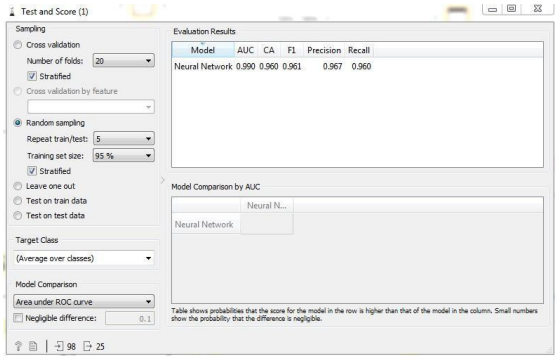
First, we take the widget of IMPORT IMAGE, and upload our dataset. Then we take the widget of IMAGE VIEWER to view the images of our dataset. Then IMAGE TMBEDDING is connected with the Import image, it is used because it expends the details about our dataset and images that are in our dataset. Now, we want the show the details about the dataset so we connect the DATA TABLE widget. Now we CONCATENATE both data tables and TEST dataset.

MERGE them with TRAIN FILE (file that we build in excel). The same process is done with after this; we will apply the algorithms, which give us accuracy. In our paper, we use two algorithms KNN and NEURAL NETWORK. We attach the TEST AND SCORE widget with this algorithm widget to view the accuracy/results.

**Test and scope of KNN**

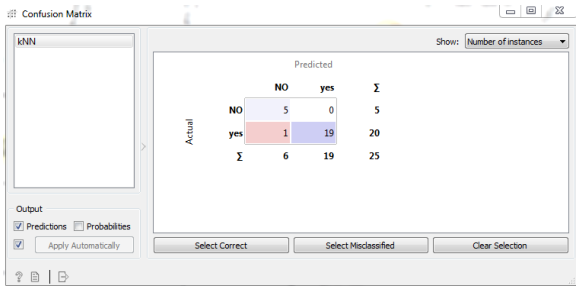



**Test and scope of Neural Network**

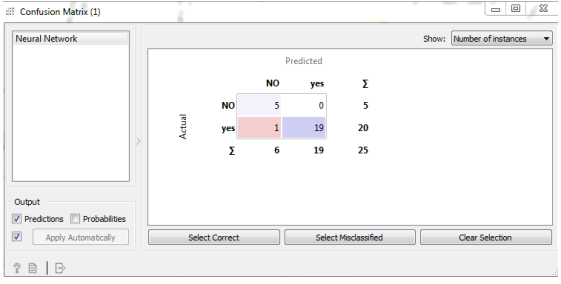



We also add a widget of CONFUSION MATRIX. We use the confusion matrix because it tells us about the instances of data that have COVID and the instances of data that don't have instances.

**Confusion matrix of KNN**

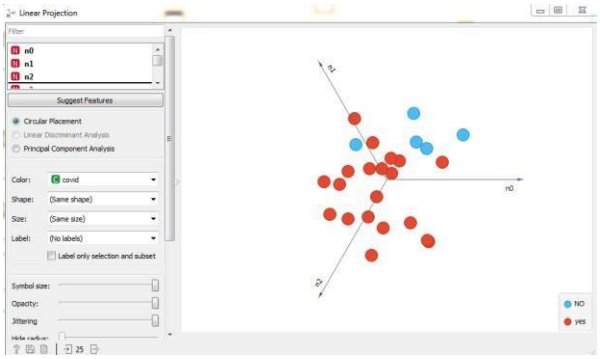



**Confusion matrix of Neural Network**

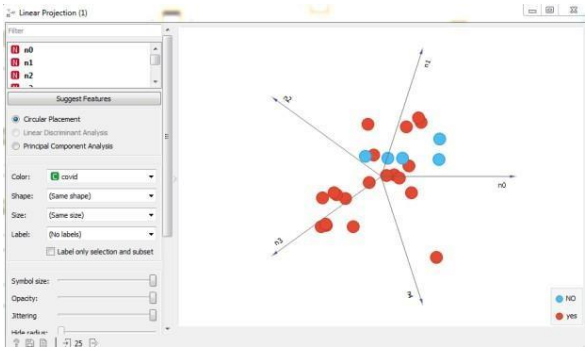



We also add visualization of our results. For visualization purposes, we use the LINEAR PROJECTION widget. Visualization is used to show our data in graphical form, so we can better understand the data or result.

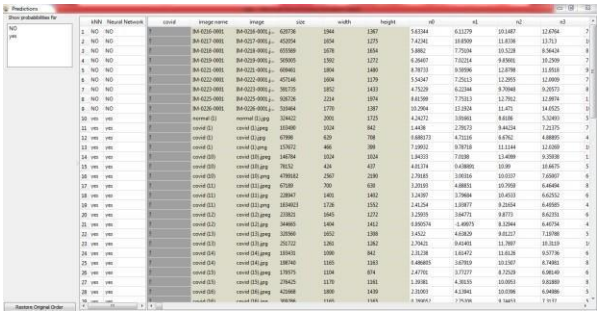
**Visualization of KNN**




**Visualization of Neural Network**




**Prediction**



## VI. CONCLUSION

We have 2 algorithms used in this paper KNN and Neural Network. With the same parameters, 20 numbers of folds and in Random Sampling 'Repeat train/test' is 5, and training test size is 95%. KNN is giving us 96% accuracy and Neural Network is giving us 96% accuracy as well. But AUC (Area under Curve) is different from both algorithms. KNN's AUC is 97% and Neural Network's AUC is 99%. Both algorithms are probably giving the same results. Or we can say that both algorithms are giving the best accuracy and predicting the best results.

We can change the results by changing the number of folds. Training test size and number of repeat train/test size. By changing these we have different results.

## Vll. FUTURE WORK

We can take different datasets or select heavy datasets. We have so many other algorithms in Orange tool, we also change the algorithms. In test and scope, we have different parameters like the number of folds, etc. so we can also change them to have different results.

## REFERENCES

[1] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Computers in Biology and Medicine,* p. 103792, 2020.

[2] R. Kumar *et al.*, "Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers," *medRxiv,* 2020.

[3] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849,* 2020.

[4] S. Asif and Y. Wenhui, "Automatic Detection of COVID-19 Using X-ray Images with Deep Convolutional Neural Networks and Machine Learning," *medRxiv,* 2020.

[5] L. O. Hall, R. Paul, D. B. Goldgof, and G. M. Goldgof, "Finding covid-19 from chest x-rays using deep learning on a small dataset," *arXiv preprint arXiv:2004.02060,* 2020.

[6] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," *arXiv preprint arXiv:2004.09363,* 2020.