International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Covid Detection Using Lung X-ray Images

Aditya Mandaliya[1], Aayush Panikkar[1], Anurag Kamath[1], Nileema Pathak[2]

[1]Dept. of Information Technology, Atharva College of Engineering, Maharashtra, India
[2]Assistant Professor, Dept. of Information Technology, Atharva College of Engineering, Maharashtra, India

---***---

Abstract - *In the current times, automated disease detection has become a critical issue in the field of medical science because of the rapid population. An automated disease detection model is required to assist the doctors in the diagnostics of disease and provide consistent, exact and fast results and help in reducing the death rate. This project tackles the problem of covid-19 detection by utilizing deep learning CNN (convolutional neural networks). A critical phase in the battle against COVID-19 is successfully screening contaminated patients. This project aims to detect COVID-19 patients using digital chest x-ray pictures while increasing the accuracy of detection using deep convolutional neural networks. Conventional testing is costly and takes a lot of time. This leads to excessive time and money expenses on the diagnosis. In this project, we have addressed both problems using the deep neural model. The dataset we've used is publicly available on the internet. We used a total of 5837 images for both training and testing our network, out of which 2837 Images were of Covid affected lungs and the remaining 3000 were of normal. Our model achieves a precision of 0.94, F1 score of 0.94, recall of 0.94 and accuracy of 0.94.*

Key Words: **Machine Learning, COVID-19, VGG16, Augmentation, X-Ray, Convolutional Neural Network**

1. INTRODUCTION

In recent times COVID-19 has had a tremendous effect on the lives of individuals globally. In the past, there were many such viruses but nothing as strong as this Coronavirus. People sure did not anticipate facing a virus like this in the present age of technology. Evaluating the time required for the diagnostic process and the expense of the kits needed to test cases, AI and deep learning-based application was required to help support doctors working on constraining the spread of the virus and treating patients combating the illness.

1.1 Motivation

The Covid 19 pandemic has been spreading rapidly throughout the world since last year. The rapid spread of Covid-19 and the intensity of the symptoms have claimed the lives of numerous humans. Individuals have lost employment, family members and friends. It has put society in a quarantine. One of the issues we are facing while fighting this pandemic is the lack of quick and affordable diagnosis. COVID19 diagnosis can take a lot of time and can be expensive for some people. In this project, we intend to use Lung X-Ray scans to identify and classify them into COVID and non-COVID images. This reduces both the cost and time taken for COVID19 diagnosis.

1.2 Aim

To build a deep learning model that could extract COVID-19's pictorial features from the X-ray image of the lungs, thus saving crucial time for disease control. Accuracy is the most significant factor in this problem, so by taking a huge number of pictures for training the network and by increasing the number of iterations, the model's accuracy can be improved.

1.3 Scope

The scope of this project is to classify X-ray images and distinguish them as Covid positive or Normal using machine learning algorithms based on image datasets. The dataset of this work has been collected from the Kaggle repository, which has been collected from multiple sources and contains Chest X-Ray scans of Covid-19 affected and normal lungs. This collected dataset is not meant to claim the diagnostic ability of any Deep Learning model but to research various possible ways of efficiently detecting Coronavirus infections using computer vision techniques.

2. LITERATURE REVIEW

In this paper [1], the un-augmented dataset which was used had achieved an average F-measure of 95% meanwhile the augmented data set achieved a score of 99%. The experiment initially contained only 128 images, on which they performed augmentation and extended it to about 1000 images. Paper [2] brought up an interesting observation, where the researchers found out that lung x-ray images provided better accuracy in comparison to CT scan images using the same algorithm. With the limited dataset that they had at hand, using the VGG19 model they attained 86% accuracy for X-ray images and 84% accuracy for CT scans. CNN models can struggle to provide consistent results if the datasets are too small.

The researchers for [3] created a temporary fix for the lack of data by implementing transfer learning in their project. In this paper, multiple models had been used, with VGG19 and Mobilenet v2 providing the best accuracy. Even though

VGG19 provided a higher accuracy, Mobilenet v2 was selected as the better model due to the lesser number of false negatives achieved for the detection of COVID. A false negative means that a person carrying COVID has been diagnosed as normal, which may be potentially harmful.

[4] uses a ResNet-50 model on a relatively small dataset. The algorithm can be easily re-trained with new sets of labelled images to enhance the performance further. In paper [5], a combined deep CNN-LSTM network was developed. The sample size is relatively small and needs to be increased in size to test the generalizability of the developed system. The CNN architecture achieved 95.3% AUC. The researchers for [6] employed domain extension transfer learning (DETL) with pre-trained CNNs and used fivefold cross-validation to estimate the feasibility of the use of chest X-Rays to diagnose COVID-19. The overall accuracy was measured to be 90.13% ± 0.14.

3. METHODOLOGY

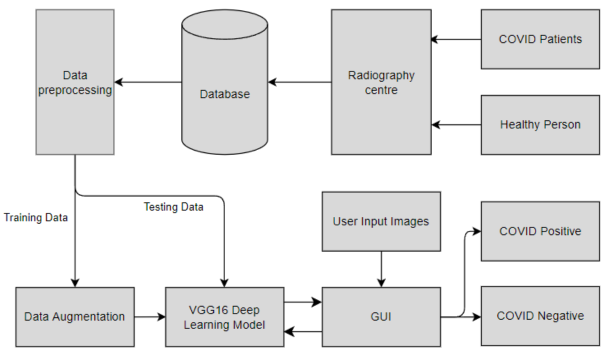

Fig-1: System Design

The proposed method used in this project uses Deep learning algorithm VGG-16. The detection of COVID-19 requires different stages, as shown in Figure 1. The original X-ray image is preprocessed, including size adjustment, width shift, height shift, rotation, and scaling and horizontal flipping of the images. The dataset is then split into training and validation (test) sets. The preprocessed data is used to extract the image features from the X-ray images. The performance of the method is assessed by indices such as accuracy, recall rate, precision, and F1-score.

3.1 Algorithm

VGG-16:

VGG16 is a convolution neural net (CNN) model that was used to win ILSVR(Imagenet) contest in 2014. It is believed to be one of the best vision model architectures to date. The most extraordinary thing about VGG16 is that instead of having a large number of hyper-parameters they focused on having convolution layers of a 3x3 filter with a stride 1 and always used the same padding and max pool layer of a 2x2 filter of stride 2. It follows this system of convolution and max pool layers consistently throughout the whole model. Finally, it has 2 FC(fully connected layers) followed by a softmax for output. The 16 in VGG16 points to the model having 16 layers that have weights. This network is pretty large and it has about 138 million (approx.) parameters.

4. DATASET

This project is using chest X-ray images that are collected from different publicly accessible datasets, online sources and published papers. This dataset is divided into two categories Covid and Normal.[7][8][9]

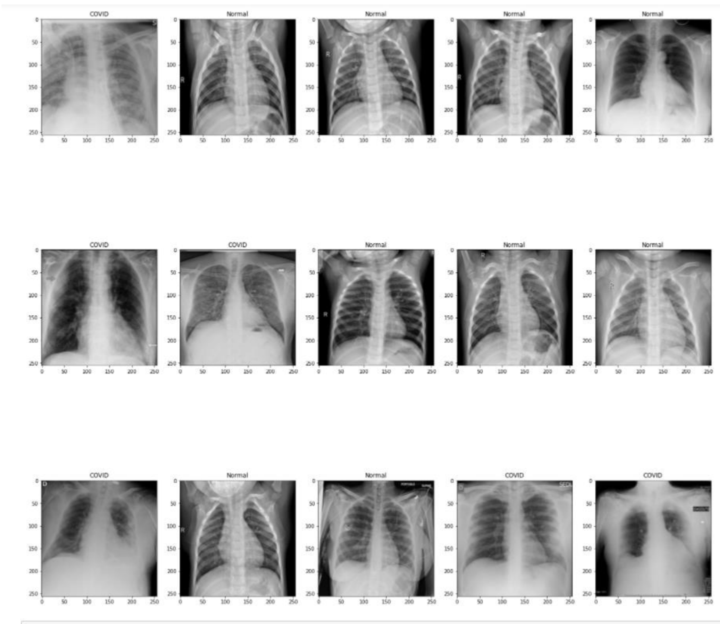

Fig-2: COVID infected and Normal X-Rays

Specifications of the data set:

Covid - The Covid data contains a total of 3,615 images

- 2,473 CXR images are collected from the padchest dataset
- 183 CXR images from a German medical school.
- 559 CXR image from Github, Tweeter, Kaggle and SIRM
- 400 CXR images from another Github source.

Normal - The Normal data contains a total of 10,192 images are collected from two different datasets.

- 8851 RSNA
- 1341 Kaggle

5. IMPLEMENTATION

The implementation of this project is the combination of the following segments: Loading image data, splitting of the data, data augmentation, Training of the model, Evaluation of the model and lastly the Graphical User Interface (GUI).

A. LOADING IMAGE DATA

Firstly, the model collects the paths of each of the image of the dataset folder, then it stores the image data into a variable, then it is ensured that the images are according to our specifications i.e. all images are of the same color and same input size. Normalizing is then performed on the image data. Lastly label binarizer is called upon the label to convert them into 0 or 1 according to the label.

B. SPLITTING OF THE DATA

The procedure which is termed test-train-split is employed to estimate the fulfillment of the machine learning algorithms once they are required to make predictions. 20% of the dataset is reserved for testing and the remaining 80% of the data is further split into training data and validation data in the ratio of 80:20 respectively.

C. DATA AUGMENTATION

In data analysis, data augmentation refers to approaches for increasing the quantity of data by adding slightly changed copies of current data or creating new synthetic data from existing data. When training a machine learning model, it functions as a regularizer and helps to minimize overfitting. In augmentation we changed the parameters of the images such as height, rotation, horizontal-flip and width.

D. TRAINING OF THE MODEL

The model is trained with the augmented dataset and is validated after each epoch. This model uses an early stopping callback because if the model's validation accuracy stops improving within the next 5 epochs, it will revert back to the epoch with the best validation accuracy. This saves time and resources while training of the model.

6. EVALUATION OF THE MODEL

The performance of a classification algorithm is shown and summarized using a confusion matrix. The model was able to predict as follows: True positive - 575, True negative – 526, False positives – 25 and False negative - 42

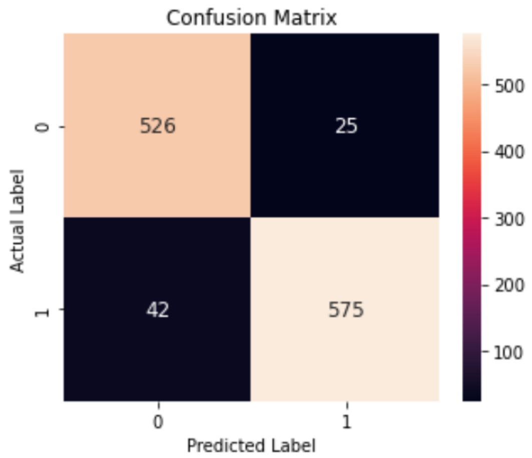

Fig-3: Confusion Matrix

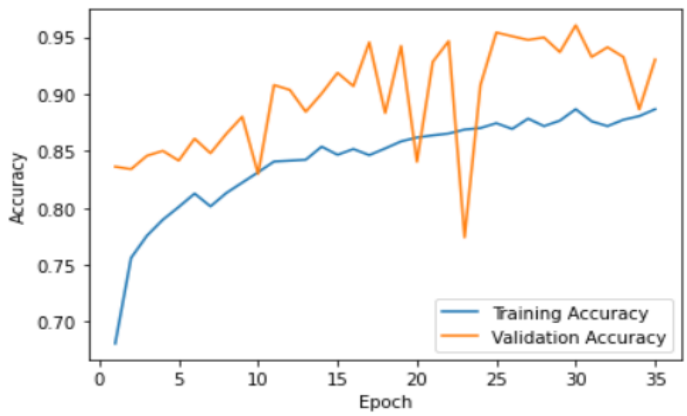

Fig-4: VGG16 Training and Validation accuracy

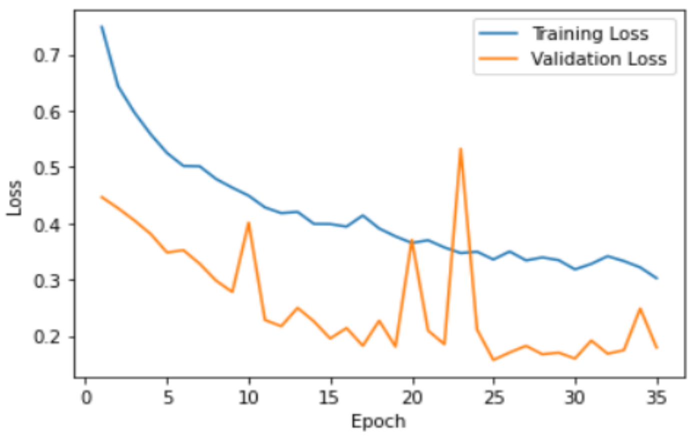

Fig-5: VGG16 Training and Validation loss

The model was evaluated on the basis of parameters such as precision, recall, F1 score, and support. The model was able to achieve the accuracy of 0.9426, sensitivity of 0.9546, specificity of 0.9319.

Table-1: Evaluation report of VGG16 model

	Precision	Recall	F1-Score	Support
COVID	0.93	0.95	0.94	551
Normal	0.96	0.93	0.94	671
Accuracy			0.94	1168
Macro Avg.	0.94	0.94	0.94	1168
Weighted Avg.	0.94	0.94	0.94	1168

7. GRAPHICAL USER INTERFACE (GUI)

A graphical user interface (GUI) allows people to interact with the system. Users will primarily be given the choice to pick a file from which to submit their lung X-Ray image.

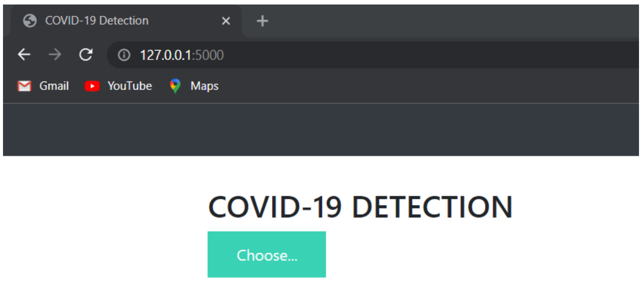

Fig-6: Primary Interface

On clicking the 'Choose' button, the user will be prompted with their respective file manager to select and upload the lung X-Ray.

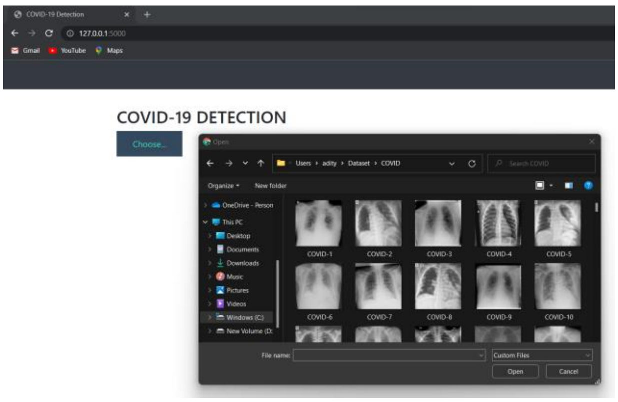

Fig-7: Select the Lung X-Ray

On uploading the image model will process the image and give the prediction of 'COVID' or 'Normal' accordingly.

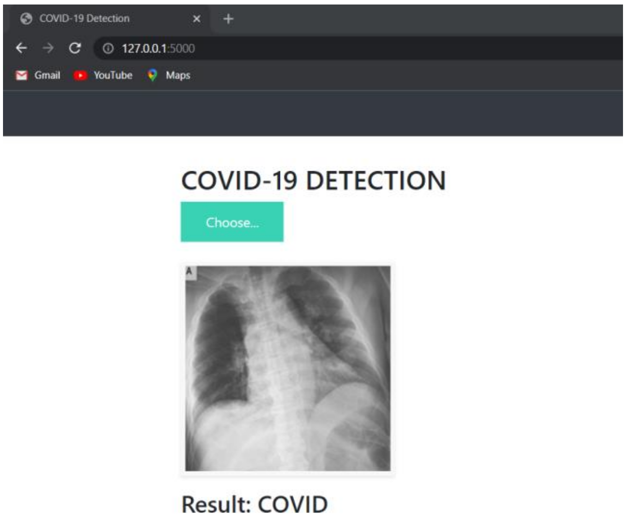

Fig-8: Predicted output: COVID

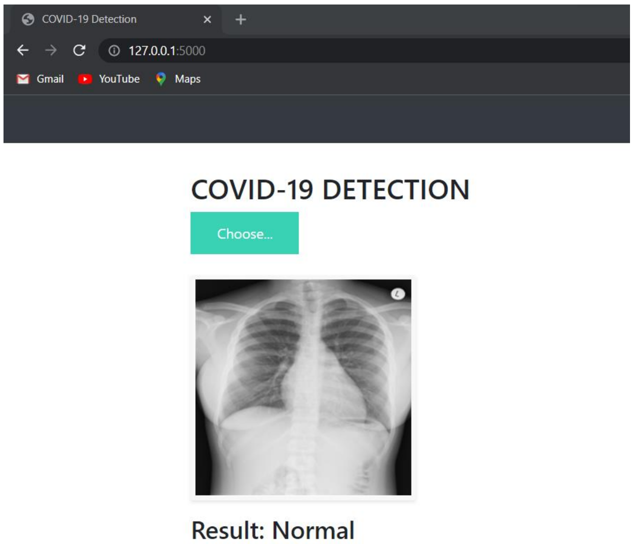

Fig-9: Predicted output: Normal

8. CONCLUSION

The concept of a deep learning model approach to the augmented dataset for COVID-19 detection is presented in this research. Our active experimentation is the effective implementation of the system, evaluation parameters, and GUI execution. COVID-19 was detected utilizing chest X-ray pictures using a diagnostic algorithm based on VGG16. With the use of an augmented dataset, the model was able to detect COVID-19 quickly and reliably, with an F1 score of 0.94.

9. REFERENCES

[1] Alazab, M., Awajan, A., Mesleh, A., Abraham, A., Jatana, V. and Alhyari, S., 2020. COVID-19 prediction and detection using deep learning. International Journal of Computer Information Systems and Industrial Management Applications, 12, pp.168-181.

[2] Horry, M.J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., and Shukla, N., 2020. COVID-19 detection through transfer learning using multimodal imaging data. IEEE Access, 8, pp.149808-149824.

[3] Apostolopoulos, I.D., Mpesiana, T.A. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43, 635–640 (2020). https://doi.org/10.1007/s13246-020-00865-4

[4] Deep Learning for Medical Imaging: COVID-19 Detection Posted by Johanna Pingel, March 18, 2020 [BLOG].

[5] Islam, M.Z., Islam, M.M. and Asraf, A., 2020. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Informatics in medicine unlocked, 20, p.100412.

[6] Basu, S., Mitra, S. and Saha, N., 2020, December. Deep learning for screening covid-19 using chest x-ray images. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 2521-2527). IEEE.

[7] Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O. and Acharya, U.R., 2020. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in biology and medicine, 121, p.103792.

[8] Hussain, E., Hasan, M., Rahman, M.A., Lee, I., Tamanna, T. and Parvez, M.Z., 2021. CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. Chaos, Solitons & Fractals, 142, p.110495.

[9] Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Al Maadeed, S., Zughaier, S.M., Khan, M.S. and Chowdhury, M.E., 2021. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Computers in biology and medicine, 132, p.104319.

[10] Poojary, R., Raina, R. and Mondal, A.K., 2021. Effect of data-augmentation on fine-tuned CNN model performance. IAES International Journal of Artificial Intelligence, 10(1), p.84.