COVID-19 Detection using Chest X-Ray

Sagar Raheja
Student, Department of Computer Engineering,
Vivekanand Education Society’s Institute of Technology
Chembur, India.
2017.sagar.raheja@ves.ac.in

Jatin Chhabria
Student, Department of Computer Engineering,
Vivekanand Education Society’s Institute of Technology
Chembur, India.
2017.jatin.chhabria@ves.ac.in

Gaurang Wadhwa
Student, Department of Computer Engineering,
Vivekanand Education Society’s Institute of Technology
Chembur, India.
2017.gaurang.wadhwa@ves.ac.in

Samay Ahuja
Student, Department of Computer Engineering,
Vivekanand Education Society’s Institute of Technology
Chembur, India.
2017.samay.ahuja@ves.ac.in

Anjali Shrikant Yeole
Assistant Professor, Department of Computer Engineering,
Vivekanand Education Society’s Institute of Technology, Chembur, India.
anjali.yeole@ves.ac.in

Abstract: Over the past few months, the exponential increase in COVID-19 cases has been overwhelming for many healthcare systems across the world. With 114 million cases globally as of 28th February 2021, with India itself having 11.1 million cases, it has challenged us with the testing, quarantine, and safety measures. Having limited testing kits, not all patients that have symptoms of respiratory illness can be tested using conventional techniques (RT-PCR). In this project, we propose the use of chest X-Ray to prioritize the selection of patients for further RT-PCR testing. It would also help in identifying patients with a high likelihood of COVID with a false negative RT-PCR who would wish to repeat testing. Further, we propose the utilization of recent AI techniques to detect the COVID-19 patients automatically using X-Ray images, particularly in settings where radiologists aren't available, and help make the proposed testing technology scalable.

Keywords: COVID-19, healthcare systems, X-Ray, RT-PCR, patient, AI.

I. INTRODUCTION

The sudden spike in the number of patients with COVID-19, a new respiratory virus, has put an unprecedented load on healthcare systems across the world. In many countries, healthcare systems have already been overwhelmed. As there are limited kits available for diagnosis, along with limited hospital beds for admission of such patients, and limited personal protective equipment (PPE) for healthcare personnel, it is thus very important to differentiate which patients with severe acute respiratory illness (SARI) could have COVID-19 infection to efficiently utilize the limited resources. In this work, we propose the use of chest X-Ray to detect COVID-19 infection in patients exhibiting symptoms of SARI. Our tool can classify a given X-Ray as one among the three classes: normal, pneumonia, and COVID pneumonia. The main contribution of this work is in proposing a unique deep neural network-based model for highly accurate detection of COVID-19 infection from the chest X-Ray images of the patients. Therefore this automated tool can serve as a guide for those at the forefront of this analysis.

As we are seeing now too, (February 2021), despite many vaccines in use and development, our country is facing a second wave of the virus and the virus is spreading again, to combat this, our system will provide faster results and help in detection and prevention of further spread faster.

II. Literature Review

2.1 respiratory illness detection in Chest X-Rays

Various deep learning-based approaches square measure developed to identify completely different diseases like respiratory illness [7, 8, 11, 13]. The model is trained to classify X-Ray pictures into fourteen completely different sickness classes, as well as respiratory illness. Seeing the similarity of the input samples, we tend to found this to be the nearest pre-trained backbone to develop a model for characteristic COVID-19 respiratory illness.

2.2 COVID-19 detection in Chest X-Rays

Their square measure solely restricted such ASCII text file applications obtainable to be used [1,5,10] that use chest X-Ray pictures.

COVID-Net [10] has AN ASCII text file and actively maintained tool which can determine COVID-19 moreover as different respiratory illness whereas showing respectable sensitivity for COVID-19 detection.
have given a fast, absolutely parameterizable GPU implementation of Convolutional Neural Network variants. This paper explains the basics of a system of logic and its use for image classification. It uses Matlab’s system of logic tool cabinet at intervals the definition of a system of logic illusion rules. These rules square measure tested and verified through the simulation of classification procedure indiscriminately sample areas. [15]

III. IMPLEMENTATION DETAILS

We have used the COVID chest x-ray dataset for COVID-19 frontal-view chest X-Ray images and chest x-ray pneumonia dataset for frontal-view chest X-Ray images with bacterial/viral pneumonia as well as of normal lungs. We will use the pre-trained CheXNet model, thus implicitly using robust features obtained after training on the Chest X-ray dataset. The two types of images in the obtained dataset are normal and pneumonia. The images in the Normal category belong to the Non-Covid case. But not all the images in Pneumonia come under the Covid category; only images in which the chest is affected due to the SARS-CoV-2 virus are referred to as COVID cases. This is because COVID-19 pneumonia is a subset of pneumonia diseases. The dimensions of all images are not fixed and had to be resized to 220 x 220 pixels for training and evaluation. The details about the downloaded dataset are shown below.

As the dataset was obtained in a raw form wherein the images are classified in three different folders i.e. training, testing, and validation, the major challenge we faced was to prepare a single dataset that contains the images of all the three categories (Normal, Pneumonia, and Covid +ve). This was achieved by combining the images present in three different folders into one folder and adequately renaming the images by the category they belong in so that images can be differentiated from each other.

Dataset Visualization

Pre-processing
The images are resized or scaled right down to 220x220x3 size. After resizing, on flattening all the images’ arrays, the entire no. of features is 145200 which might cross the available RAM size while training the model on this dataset leading to a session crash. To beat this, we applied a feature reduction technique called Histogram of Oriented Gradients (HOG) which reduced the feature size of images to a good extent i.e., 24336 features. So, in this way, we obtained the reduced feature size of every image within the dataset.
After plotting both the classification tasks using the 2D-distributed stochastic neighbor embedding (TSNE), we conclude that the knowledge isn't fully linearly distributed, there is a big amount of nonlinearity associated with the dataset. In both the 2D TSNE plots, the space occupied by the COVID-19+ve class is extremely less because the amount of samples belonging to the current class within the dataset is incredibly less as compared to other classes, binary classification task, the number of COVID+ve samples are 1493 and, for COVID-ve, it is 4363. However, for the multiclass classification problem, the quantity of samples belonging to COVID+ve is identical to the binary classification task. The other two classes are: NORMAL and PNEUMONIA, each containing 1583 and 2780 samples respectively.

Methodology
We have created the truth table of the X-ray images dataset by identifying the category of the image from its filename. We see that for binary classification tasks, we have two classes 0 for COVID-ve and 1 for COVID+ve. Similarly, for the multiclass classification task, there are three classes 0 which would mean NORMAL, 1 would mean PNEUMONIA, and 2 would stand for COVID+ve.

The dataset is split into 80:20 ratio to make the training and testing sets. The number of instances for training and testing is 4684 and 1172 respectively. We are visiting apply various ML algorithms (or techniques) to educate the model. For the binary classification task, we apply the following ML algorithms like Decision Tree, regression toward the mean, Support Vector Machine (SVM), K-Nearest Neighbours (KNN). For the multiclass classification tasks, we have applied every algorithm mentioned for binary classification, except KNN to training the model and evaluating its performance. The complete dataset goes to be divided into 5 folds and so the performance of the model goes to be analyzed fold-wise in terms of accuracy, Precision, F1-score, Sensitivity, and Specificity.

Experimental Results:

1. Binary Classification task (Covid+ve vs Covid-ve):
The fold-wise performance of the model created by applying various ML techniques is shown below in the form of classification reports.

| Decision Tree | Performance Metrics | Sensitivity | Specificity | Precision | F1-score | Accuracy |
|---------------|---------------------|-------------|-------------|-----------|----------|----------|
| Fold 1        | 0.45                | 0.80        | 0.44        | 0.45      | 0.72     |
| Fold 2        | 0.40                | 0.80        | 0.42        | 0.41      | 0.70     |
| Fold 3        | 0.50                | 0.81        | 0.50        | 0.51      | 0.74     |
| Fold 4        | 0.50                | 0.79        | 0.48        | 0.49      | 0.72     |
| Fold 5        | 0.43                | 0.78        | 0.38        | 0.41      | 0.70     |
| Average       | 0.46                | 0.80        | 0.46        | 0.46      | 0.72     |

| Logistic Regression | Performance Metrics | Sensitivity | Specificity | Precision | F1-score | Accuracy |
|---------------------|---------------------|-------------|-------------|-----------|----------|----------|
| Fold 1              | 0.50                | 0.89        | 0.62        | 0.56      | 0.80     |
| Fold 2              | 0.54                | 0.89        | 0.65        | 0.59      | 0.81     |
| Fold 3              | 0.54                | 0.89        | 0.61        | 0.55      | 0.79     |
| Fold 4              | 0.50                | 0.88        | 0.59        | 0.54      | 0.78     |
| Fold 5              | 0.52                | 0.87        | 0.60        | 0.56      | 0.79     |
| Average             | 0.52                | 0.88        | 0.61        | 0.56      | 0.79     |
The confusion matrices that have been obtained after performing binary classification generated fold wise by the SVM model are shown below:

2. Multiclass Classification task (3 classes):
   After applying various ML techniques, the fold-wise performance of the model created is shown below:

As the number of samples of class COVID +ve (minority class) is very less as compared to other classes, so there exists a class imbalance problem in our dataset because of which the accuracy of our model in case of any ML techniques applied above is still below 85% and is unable to improve much. To solve this problem, we use another ML technique.

Dealing with Class Imbalance Problem

Most machine learning techniques will ignore, and in turn have poor performance on, the minority class in case of a class imbalance problem because it is the performance of the minority
The simplest approach to solve the imbalance issue is to duplicate entries in the minority class, although they don’t add any new information to the model. We can synthesize new examples from the existing examples, this is a type of data augmentation for the minority class and is referred to as the Synthetic Minority Oversampling Technique, (SMOTE).

After applying the SMOTE technique on the dataset to equalize the number of samples of both minority and majority classes, we again applied SVM for the binary classification task and observed the results.

![Classification Report for SV SMOTE](https://medium.com/swlh/automated-detection-of-covid-19-cases-with-x-ray-images-f5b9557b36d9)

**Figure 13: Classification Report for SV SMOTE**

**Metrics used:**

\[
\begin{align*}
\text{precision} &= \frac{tp}{tp + fp} \\
\text{recall} &= \frac{tp}{tp + fn} \\
\text{accuracy} &= \frac{tp + tn}{tp + tn + fp + fn} \\
F_1 \text{ score} &= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\end{align*}
\]

![Metrics Used](https://medium.com/swlh/covid-19-detection-from-x-ray-using-deep-learning-a8b8046fc077)

**Figure 14: Metrics Used**

![Screenshot of test Report](https://medium.com/swlh/covid-19-detection-from-x-ray-using-deep-learning-a8b8046fc077)

**Figure 15: Screenshot of test Report**

IV. Results

a. Normal Lungs
Figure 16: Screenshot of test Report 2

Figure 17: Screenshot of the opening screen, from where you select the image to be tested

Figure 18: Screenshot of a Normal result in green
Figure 19: Screenshot of a Normal result in Python Shell as well

Figure 20: Screenshot of a Covid-19+ve result in Python Shell as well

V. Application

Looking at the current situation, the most viable and efficient solution is that this could be used to ramp up the testing with no delays and treatment could be started with immediate effect.

Comparison with other tests:

1) Antigen test - Not accurate, nowhere accepted to get hospitalized, requires raw material and not accessible everywhere

2) RT PCR – Accepted (Atleast 2days result delay), but you also need to attach HRCT scan to get admitted which is expensive

3) Our project – Accurate enough plus no further test needed, can be used to understand seriousness with no time delay, Cost efficient. No raw material required and almost available everywhere.
VI. Future Scope

- **Intelligent chatbots**
  chatbots use natural language to communicate with patients, identify an issue and resolve the issue thereby we can enhance the quality of e-commerce service[6].

VII. Conclusion

As we know the detection of Covid-19 from chest X-ray images is very important for both doctors and patients to decrease the diagnostic time, reduce financial costs and be able to save more lives. For recognizing images for the tasks thought, artificial intelligence and deep learning are capable. As we have seen, several experiments were performed for the high-accuracy detection of COVID-19 in chest X-ray images.

When the number of images in the database and the detection time of COVID-19 (average testing time = 0.03 s/image) are considered, it can be suggested that the considered architectures reduce the computational cost with high performance. We got the results that showed the convolutional neural network with minimized convolutional and fully connected layers is capable of detecting COVID-19 images within the two-class, COVID-19/Normal and COVID-19/Pneumonia classifications with a mean accuracy of 85%.

VIII. Acknowledgement

*We wish to express a true sense of gratitude towards* our Mentor Dr.(Mrs.) Anjali Shrikant Yeole for giving us this opportunity to work on this project, using our skills and knowledge. We would like to extend our gratitude to all team members for their time-to-time support and hard work because of which we can build a well-written paper. With all respect and gratitude, we appreciate our success to the writers of reference papers that are referred by us in completion of this paperwork activity which will be useful in presenting our survey paper.
We wish to express a true sense of gratitude towards knowledge. We would like to extend our gratitude to all team Mentor Dr.(Mrs.) Anjali Shrikant Yeole for giving us this opportunity to work on this project, using our members for their time-to-time support and hard work because of which we can build a well-written paper. With all respect and gratitude, we appreciate our success to the writers of reference papers that are referred by us in completion of this paperwork activity which will be useful in presenting our survey paper.

Intelligent chatbots
- able to save more lives. For recognizing images for the When the number of images in the database and the decrease the diagnostic time, reduce financial costs and be and fully connected layers is capable of detecting COVID-19 images within the two-class, COVID-19 images is very important for both doctors tasks thought, artificial intelligence and deep learning are detection time of COVID-19 (average testing time = 0.03 s/image) are considered, it can be suggested that the capable. As we have seen, several experiments were performed for the high-accuracy detection of COVID-19 in considered architectures reduce the computational cost with a mean accuracy of 85%.

X-ray images.

VI. Future Scope

VII. CONCLUSION

IX. REFERENCES

[1.] Bennett, J.M.: Smart ct scan based covid19 virus detector. https://github.com/JordanMicahBennett/SMART-CT-SCAN_BASED-COVID19_VIRUS_DETECTOR
[2.] Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection. arXiv:2003.11597 (2020), https://github.com/ieee8023/covid-chestxray-dataset
[3.] Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
[4.] Mooney, P.: Kaggle chest x-ray images (pneumonia) dataset. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia (2018)
[5.] Nisar, Z.: https://github.com/zeeshannisar/covid-19. https://github.com/zeeshannisar/COVID-19 (2020)
[6.] Petsiuk, V., Das, A., Saenko, K.: Rise: Randomized input sampling for the explanation of black-box models. In: Proceedings of the British Machine Vision Conference (BMVC) (2018)
[7.] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., et al.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
[8.] Ranjan, E., Paul, S., Kapoor, S., Kar, A., Sethuraman, R., Sheet, D.: Jointly learning convolutional representations to compress radiological images and classify thoracic diseases in the compressed domain (12 2018). https://doi.org/10.1145/3293353.3293408
[9.] Ruiz, P.: Understanding and visualizing densenets. https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a (2018)
[10.] Wang, L., Wong, A.: Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images (2020)
[11.] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8:Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2097–2106 (2017)
[12.] Weng, X., Zhuang, N., Tian, J., Liu, Y.: Chexnet for classification and localization of thoracic diseases. https://github.com/arnoweng/CheXNet/ (2017)
[13.] Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D., Lyman, K.: Learning to diagnose from scratch by exploiting dependencies among labels. arXiv preprint arXiv:1710.10501 (2017)