Detection of Covid-19 using Chest X-Ray

Divya Dwivedi Siddhesh Bagwe

University College Dublin

Background

The first documented instance of a highly transmissible illness caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was detected in Wuhan, China, in December 2019. The World Health Organization (WHO) designated it as a Public Health Emergency of International Concern or an epidemic in January 2020. Despite efforts to contain it, the disease continued to spread, leading to thousands of deaths. In February 2020, the WHO officially named the illness "COVID-19," and by March 2020, it was declared a pandemic, rapidly spreading across the globe.It has caused a devastating effect on both daily lives, public health, and the global economy. Recent findings obtained using radiology imaging techniques suggest that such images contain salient information about the COVID-19 virus. Application of advanced artificial intelligence (AI) techniques coupled with radiological imaging can be helpful for the accurate detection of this disease, and can also be assistive to overcome the problem of a lack of specialized physicians in remote villages.

Methodology

The urgency of the pandemic led to many studies using datasets that contain obvious biases or are not representative of the target population, for example, paediatric patients. Before evaluating a model, it is crucial to report the demographic statistics for their datasets. The data set used for this study is clinical data for individuals aged between 10-90 years with 65% Male and 35% Female ratio.

Model

This model is designed for automatic COVID-19 detection using raw chest X-ray images. The model uses three convolutional layers with an input shape of (100, 100, 3) and Rectified Linear Unit (ReLU) activation function to introduce non-linearity in the model. The first layer has 32 filters to convolve over the input image. It also employs dropout and L2 regularization techniques to improve generalization and prevent overfitting. In the final dense layer, the softmax activation function is used to obtain the class probabilities for multi-class classification.

The following are some of the key findings of this study:

(i) CNN with three convolutional layers performs best in COVID-19 diagnosis

(ii) CNN models require a sufficient amount of images for efficient and more accurate image classification. The dataset contains 6536 X-ray images.

(iii) Data augmentation techniques are very effective to improve the CNN model performance remarkably by generating more data from an existing limited size dataset.

(iv) Data augmentation is also effective in image classification as it gives the ability of invariance to CNNs.

(v) CNN-based diagnosis using X-ray imaging can be very effective for medical sector to handle the mass testing situations in pandemics like COVID-19

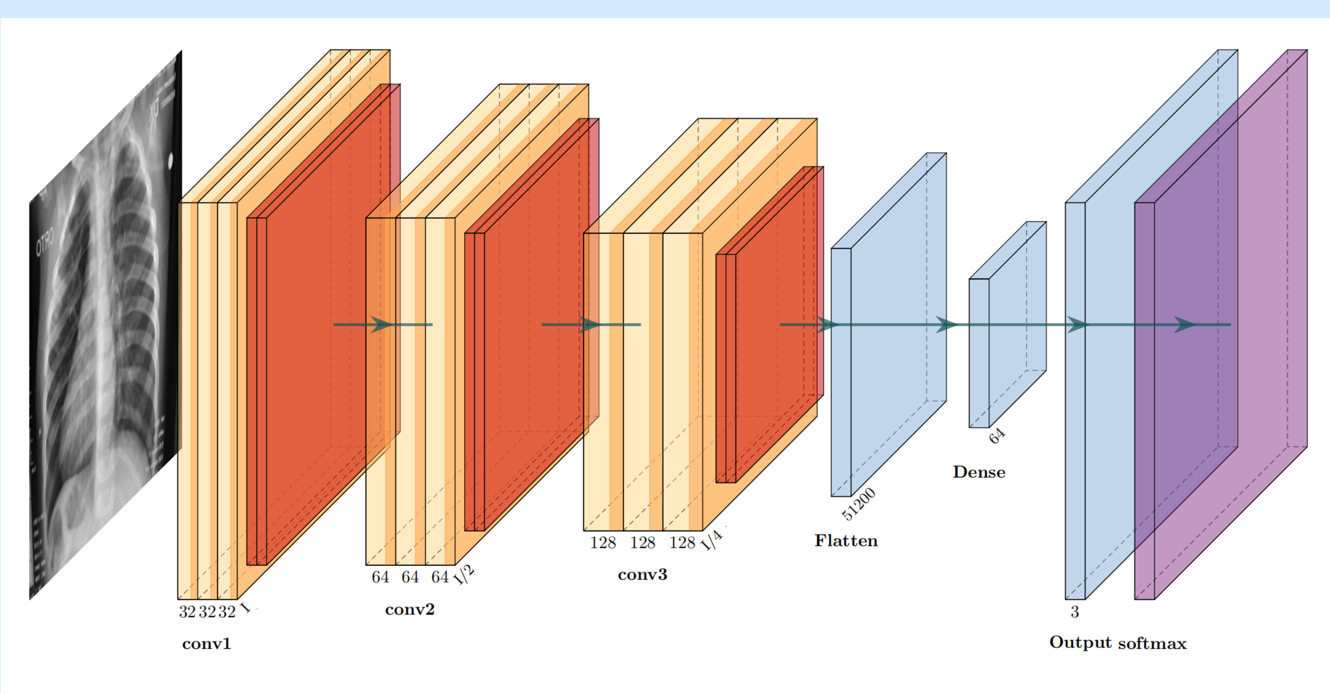

Figure 1. CNN Model

Results

The goal of this research is to identify the presence of COVID-19 in the unseen Chest X-rays that may be presented to the model in the future. We have split the data in Training, Validation and Test Data set with a test size of 20% and validation size of 10% of the data set. The figure 2 shows the classification report from the our model training and testing. We are interested in the F1 score of the model. This score provides the balance between precision and recall or in other words it is the accuracy for individual class. From the output below, we can see the overall accuracy is 95.9% for validation and 95.4% for test data. The support count for each class represents the number of images on which the model training and testing was performed.

	precision	recall	f1-score	support
COVID19	0.95312	0.92424	0.93846	66
Normal	0.92857	0.95413	0.94118	109
Pneumonia	0.97118	0.96839	0.96978	348
accuracy			0.95985	523
macro avg	0.95096	0.94892	0.94981	523
weighted avg	0.96002	0.95985	0.95987	523

	precision	recall	f1-score	support
COVID19	0.96296	0.96296	0.96296	135
Normal	0.91772	0.91483	0.91627	317
Pneumonia	0.96616	0.96729	0.96673	856
accuracy			0.95413	1308
macro avg	0.94895	0.94836	0.94865	1308
weighted avg	0.95409	0.95413	0.95411	1308

Figure 2. Classification Report of Validation and Test Data

To analyze the model classification we look into the Confusion matrix of our proposed model. We can see that the sensitivity (Recall) of Covid-19 (96.2%) is at par with sensitivity of Pneumonia (96.7%). Due to the fatality of the problem in hand, we aim to focus on the False Negatives of the model which is only 2 case out of the total dataset. This is likely due to the overlapping imaging characteristics.

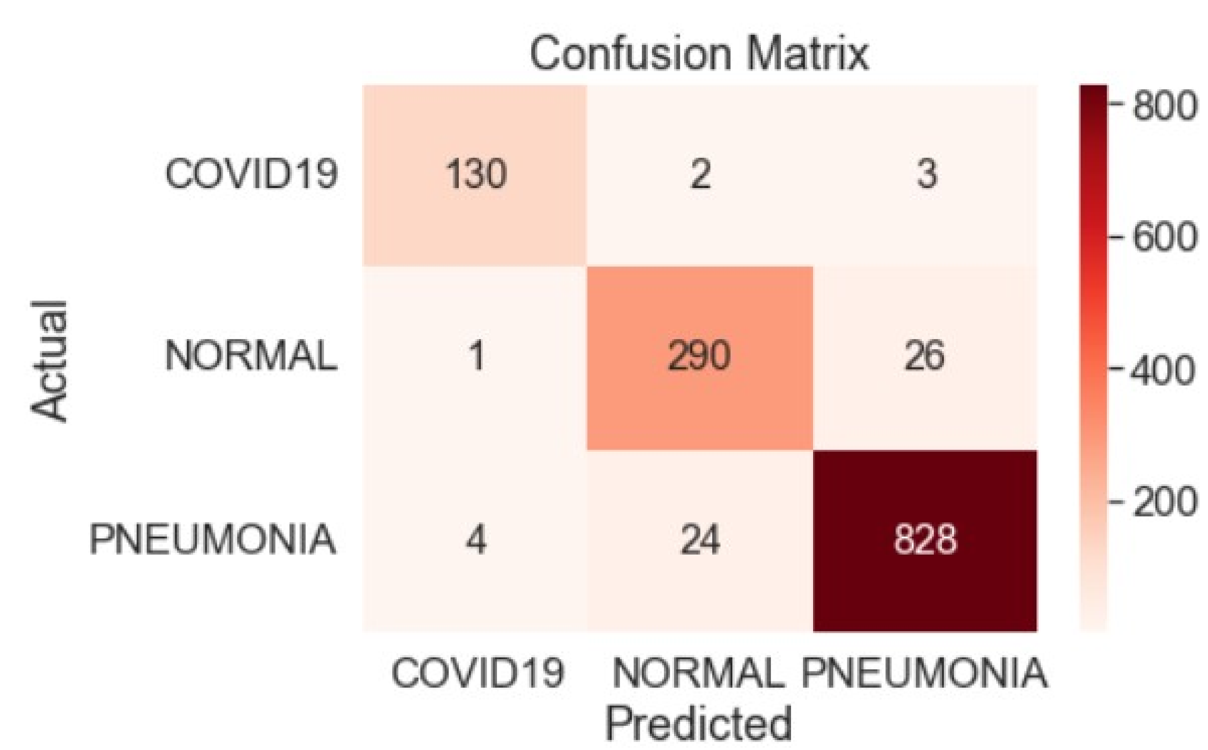

Figure 3. Confusion Matrix

In addition to this, we shall have an additional check to rule out model over-fitting as well as to validate whether the regions of attention correspond to the right features from a radiologist's perspective. We shall use Saliency map visualization for the same.

Visualization

Given the severity of the issue at hand, it is imperative that we create a visualization of the scare tissue presence. Gradient-wighted Class Activation Mapping uses the gradients from any target convolutional layer to create a local map which highlights important feature that the model has learned.The images on the left column are the actual input images of chest X-ray. heatmap of class activation is then generated from the image based on the detected features from the image. Finally, the heatmap is superimposed on the actual image to clearly show the presence of the COVID-19 scare tissue on the X-ray image.

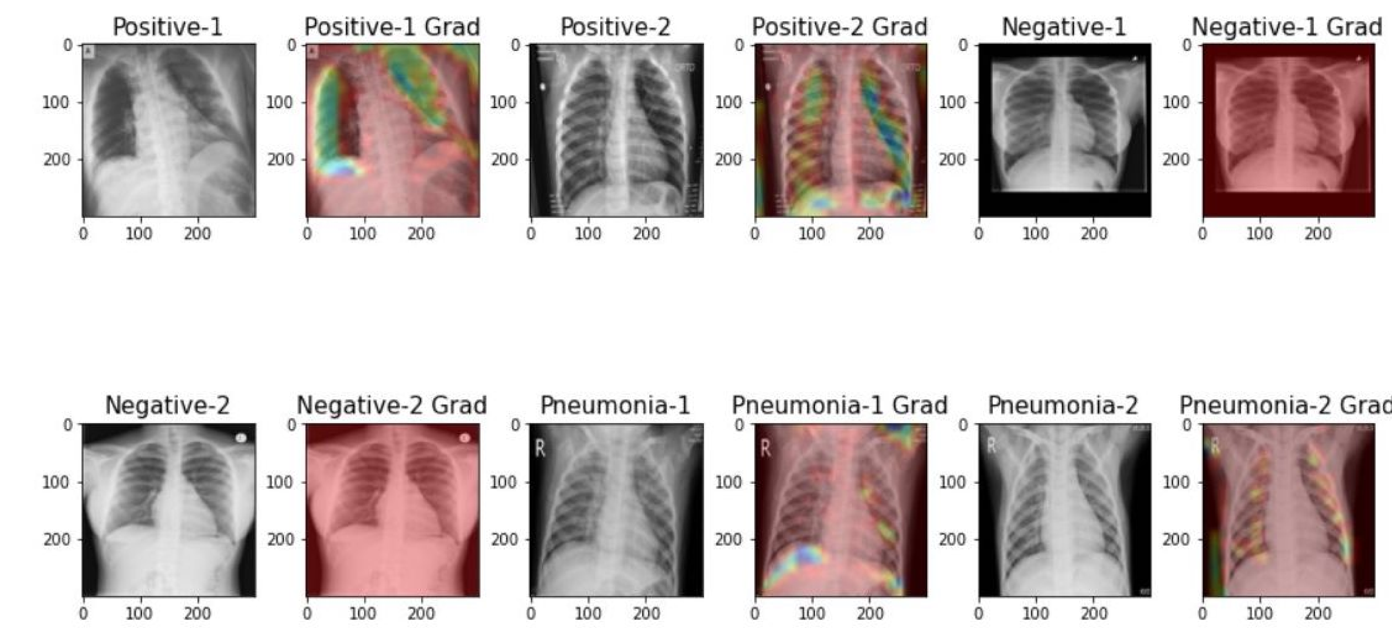

Figure 4. Gradient-wighted Class Activation Mapping

The high-intensity visuals (blue and green) reflects the area of interest to our model at the time of prediction

Challenges and Opportunities

Models for diagnosing and prognosticating from radiological imaging data face limitations due to the quality of their training data. Public datasets available for training deep learning models in these areas are often insufficient in size and quality, leading to unreliable models with a high or unclear risk of bias. To address this, we propose establishing a systematic review database where researchers can submit data for public review, continuously improving dataset size and quality.For successful translation, AI algorithms for COVID-19 detection, diagnosis, or prognosis must be closely linked to clear clinical needs. This necessitates collaboration between computational and clinical experts, utilizing high-quality healthcare data.

Conclusion

A reliable and automatic mechanism for COVID-19 diagnosis is presented using chest radiography images to differentiate between patients with pneumonia, and COVID-19 infections. The system incorporates image enhancement techniques to improve Chest X-ray image intensity and eliminate noise. Notably, the proposed system surpasses existing models, as evidenced by classification report. Furthermore, this model can assess newly acquired clinical data sets to detect the presence of COVID-19.

References

[1] D. Cucinotta and M. Vanelli. "who declares covid-19 a pandemic". *Acta Biomedica: Atenei Parmensis*, 91:157–160, 2020.

[2] Humayun M Jhanjhi NZ Gouda W, Almurafeh M. Detection of covid-19 based on chest x-rays using deep learning. *Healthcare (Basel)*, page 343, February 2022.

[3] A. S. Lundervold and A. Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift fur Medizinische Physik*, 29:102–127, 2019.

[4] P. K. Ratha P. K. Sethy, S. K. Behera and P. Biswas. Detection of coronavirus disease (covid-19) based on deep features and support vector machine. 5:643–651, 2020.

[5] S. Kasaei R. Rouhi, M. Jafari and P. Keshavarzian. Benign and malignant breast tumors classification based on region growing and cnn segmentation. *Expert Systems with Applications*, 42:990–1002, 2015.