ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023

https://doi.org/10.5281/zenodo.12707447

Detection of Covid-19 Using A Deep-LearningCT Scan Image Classifier

VIJAYA BHASKAR MADGULA, DIGALA RAGHAVARAJU,TALARI SIVALAKSHMI

Assistant Professor[1,2,3],

shivalakshmidinesh@gmail.com, raghava.digala@gmail.com, vijaya.bhaskar2010@gmail.com

department of CSE, Sri Venkateswara Institute of Technology

N.H 44, Hampapuram, Rapthadu, Anantapuramu, Andhra Pradesh 515722

Keywords:

XGB-L,XGB-Tree, ResNet, RTPCR

ABSTRACT

The global spread of the COVID-19 epidemic has caused widespread fear and destruction. The very infectious SARS-CoV-2 virus has impacted almost every country on Earth. In the early stages of a viral outbreak, efficient isolation and patient care are necessary to inhibit the spread of the virus. In the present day, RT-PCR testing is used to confirm the medical condition. Nonetheless, studies and other publications have shown that RT-PCR's sensitivity may not be sufficient for the early diagnosis and treatment of individuals suspected of having COVID-19. Recent years have seen remarkable progress in medical imaging employing AI powered by deep learning technologies. Applying deep learning allowed for rapid and precise diagnosis of bacterial and viral pneumonia in chest computed tomography imaging. One imaging method that has been useful in the early detection and screening of COVID-19 is the computed tomography (CT) scan. Our primary objective is to develop and deploy a Deep Learning classifier that can accurately distinguish between COVID and non-COVID patients, allowing us to provide patient outcomes with the highest degree of precision. Prediction sensitivity is good enough to identify early-stage viral infections. Since it is a Deep Learning model, the time it takes to analyse and provide testing results is lightning quick.

ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023
https://doi.org/10.5281/zenodo.12707447

Introduction

Chinese health officials learned about a previously undiscovered pneumonia source in 2019. A new coronavirus known as SARS-CoV-2 causes severe acute respiratory syndrome, which mostly affects the lungs and respiratory system. The transmission of this virus from bats to humans remains a mystery. In addition to several common and moderate symptoms such as a cold, headache, and fever, severe corona symptoms include trouble breathing and chest discomfort. A reverse transcription-polymerase chain reaction (RT-PCR) assay was first used to forecast the coronavirus. There are two main problems with this method: first, it generates false positives, which means that individuals who aren't really sick are informed they are, and second, it delivers detrimental results to those who do have the virus. This method of incorrect classification is woefully inadequate since it would enable several sick individuals to return home and transmit the infection. The time it takes to get RTPCR results makes the virus more contagious, however nowadays there are computer models that can forecast the values and provide correct findings. If the samples are not sent to the lab for the RT-PCR test within two days after collection, a high probability of false positives will be present. As time goes on, the precision drops. Due to this, the procedure cannot be trusted. The CT scan is a non-invasive imaging method that may show the pulmonary symptoms of COVID-19. Consequently, one of the best ways to detect COVID-19 early on is with a CT scan. Our method for virus prediction in CT Scan pictures relies on detecting image distortions, as detailed in our publication. Fluid accumulation in the lung, manifesting as opacities, is a common symptom of pneumonia-related abnormalities. When comparing CT scans of lungs infected with COVID-19 to those of other forms of pneumonia caused by various viruses, the study article on utilising deep learning to identify COVID-19 using CT scans highlights many distinctive traits. There are three such traits listed in the paper: [1]: "...ground-glass appearance, striking peripheral distribution along with the pleura, and usually more than one independent focus of infections for one case." 1. Lung opacities like ground-glass may be seen in clusters, with the bulk of them located near the lung's periphery. We looked at a plethora of other articles and methods that are or might be utilised to detect COVID-19. Here they are: 1) RT-PCR (Reverse Transcription-Polymerase Chain Reaction):[2] this method screens viral RNA directly by identifying antibodies in the body. The laboratory adds reverse transcription polymerase to the swab samples that have been collected. This technique, which is based on nuclear energy, may identify viruses and other pathogens by analysing their genetic content. This involves taking a sample from the patient's nasal passages or throat. False negative findings might happen in as many as 30 percent of PCR testing. Consequently, this method is thought to be less precise. 2) Antibody testing: This doesn't look for the virus, but rather for antibodies. We may learn about the impacted population percentage using this method. The test may miss the mark when looking for antibodies if administered too early in the course of an illness. Antibody testing analyses the immune system in addition to the virus. The precision depends on when and what kind of antibody test was conducted. Therefore, you shouldn't use this method until after two weeks of experiencing symptoms. This technique has an 84% success rate.

Thirdly, a ResNet-based CT-based deep learning model: [4] For COVID-19 patients with moderate to severe symptoms, the diversity severity scale produces reliable findings. Using a pre-trained ResNet network for severity diagnosis is how this strategy is put into action. We used 408 patients' CT scans and a five-fold cross-validation method. Using a CT scan picture, this model accurately predicts the severity and course of the illness. Compared to using individual CT parameters for prediction, the Deep Learning CT model that incorporates five clinic-radiological factors is more effective. An accuracy rate of 87.50% for the severity and 78.54% for the non-severity were found in this research. 4) A bidirectional elastic registration technique was used to segment the lung boundaries using chest CT images [5]. In order to

ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023
https://doi.org/10.5281/zenodo.12707447

identify the virus, the centre parts of the lungs were examined. By using a threshold, anomalies in the lung area may be detected. An accuracy rate of 81.3% for severity and non-severity was reached in this investigation. According to these findings, using computer tools yields reliable outcomes when it comes to find the pneumonic abnormalities in the picture and measure them.

regions associated with covid-19 and generate heatmaps to identify the

1) Using Machine Learning for chest X-rays:**[6]** This method relies upon supervised learning and is mainly composed of two main phases: training, testing. It shows the method's effectiveness by calculating precision and recall equal to 0.965 and distinguishing between covid-19 and pulmonary diseases with symptoms like covid-19. It differs between the output results of deep learning and formal verifications. Deep learning methods have the potential to improve accuracy and efficiency.

2) Machine learning classification using XGB-L, XGB-TREE :**[7]** This technique is aimed at developing an AI imaging analysis tool to classify COVID-19 lung infection based on x-rays. AI classification of texture andmorphological features for COVID-19 vs. regular utilizing five different classifiers: XGB-L, XGB-Tree, CART (DT), KNN, and Naïve Bayes. They used all performance measures and the top four ranked features (i.e., compactness, thin ratio, perimeter, standard deviation).

II THE PROPOSED SYSTEM

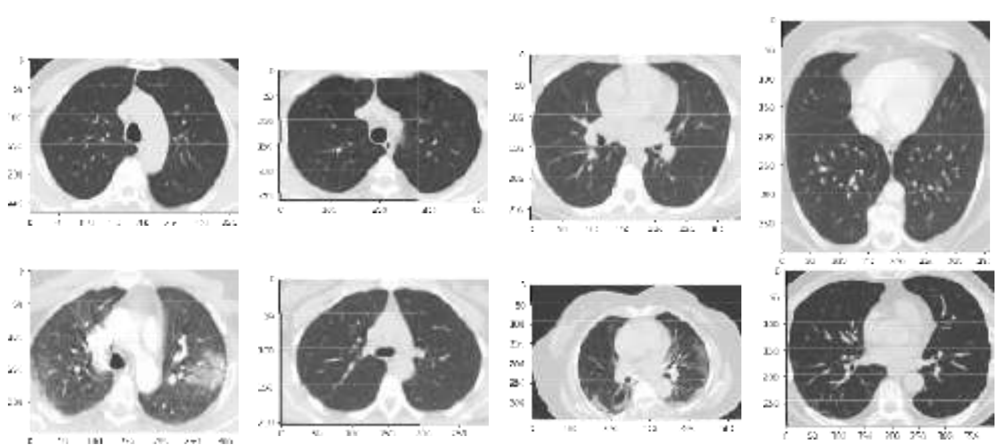
1. Data Set Acquisition:

In our paper, we will be using a large dataset **[8]** of CT scans publicly available to identify covid-19.

The SARS-Cov-2 Ct Scan Dataset. The dataset contains 1252 CT scans of patients infected by the SARS-CoV-2virus and 1230 CT scans of non-infected SARS-CoV-2 patients with pneumonia and other pulmonary diseases.

All this data is collected from different hospitals in Sao Paulo, Brazil.

To train a robust classifier, we need to have information about the non-covid patients who contain other diseases thatmight affect the lungs and cause opacities.



Sample images

ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023
https://doi.org/10.5281/zenodo.12707447

Luckily, this dataset is not skewed toward any of the classes. All images are of different shapes, which implies thenecessity of reshaping. This study takes us to the next module, preprocessing.

2. *Performing Data Preprocessing:*

All the CT scan images undergo preprocessing that standardize the images, and then they are used to train and test the deep learning model for feature extraction. The convolutional neural networks will process these extracted features and classify the CT image as covid or non-covid.

a) The input images are preprocessed to make them compatible with a pre-trained network, i.e., an image which is of size 224 x 224 x 3 converted into 64 x 64 x 3 to maintain uniformity.

b) We labeled all the CT images from train data to 0 (no covid) or 1 (covid).

c) We performed normalization on the image to have a mean of 0 and a standard deviation of 1. After completing data preprocessing, we shall be sending the images to the image classification model, which we build using convolutional neural networks

3. *Image Classification by Extracting Features:*

From the existing systems, we have observed that the Deep Learning models provide better accuracy and results. A Deep Learning-based model is used in this system. Convolutional neural networks function just like a human brain.

A Neural Network has neurons arranged in layers; input passes through the layers from the input layer to the output layer and generates the output. We use convolutional neural networks to deal with inputs having 2D and 3D formats and extract various spatial features.

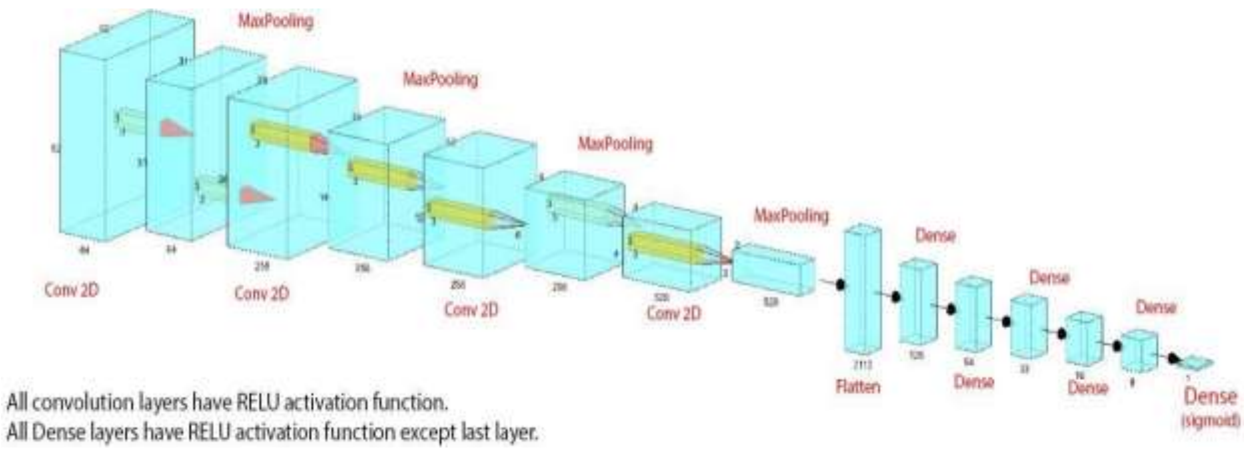
Instead of extracting the features manually, unlike in Traditional Machine Learning techniques, the features are automatically extracted in Neural Networks by the hidden layers. These can also perform well even if the dataset is too large. Neural Networks are generally trained using the Back Propagation algorithm.

We built a customized Convolutional neural network from scratch and performed binary classification (covid or non-covid).

AES

ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023

https://doi.org/10.5281/zenodo.12707447



4. *Building the model:*

We build a 15-layer sequential network containing four convolutional layers, which employed stochastic gradient descent with momentum to update all network parameters. Each followed by a max-pooling layer; finally, a flatten layer is followed by six dense layers for generating the diagnosis.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 64)	1792
max_pooling2d (MaxPooling2D)	(None, 31, 31, 64)	0
conv2d_1 (Conv2D)	(None, 29, 29, 256)	147712
max_pooling2d_1 (MaxPooling2	(None, 14, 14, 256)	0
conv2d_2 (Conv2D)	(None, 12, 12, 256)	590080
max_pooling2d_2 (MaxPooling2	(None, 6, 6, 256)	0
conv2d_3 (Conv2D)	(None, 4, 4, 528)	1217040
max_pooling2d_3 (MaxPooling2	(None, 2, 2, 528)	0
flatten (Flatten)	(None, 211 2)	0
dense (Dense)	(None, 128)	270464
dense_1 (Dense)	(None, 64)	8256
dense_2 (Dense)	(None, 32)	2080
dense_3 (Dense)	(None, 16)	528
dense_4 (Dense)	(None, 8)	136
dense_5 (Dense)	(None, 1)	9

ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023
https://doi.org/10.5281/zenodo.12707447

```
Total params: 2,238,097
Trainable params: 2,238,097
Non-trainable params: 0
```

a) We shuffle the data and split all the available images into train data (80%) and test data (20%).

b) We set the batch size to 64 for train data which means 64 images are processed as chunks and epochs to 30.

c) We used the sigmoid activation function in the last layer to perform binary classification so the output could be classified into two categories. For this case, greater than or equal to 0.5 implies COVID-19 negative. And less than
0.5 corresponds to COVID-19 positive.

d) With 'rmsprop' optimization, the model is compiled with the default learning rate. The loss function used wasbinary cross-entropy.

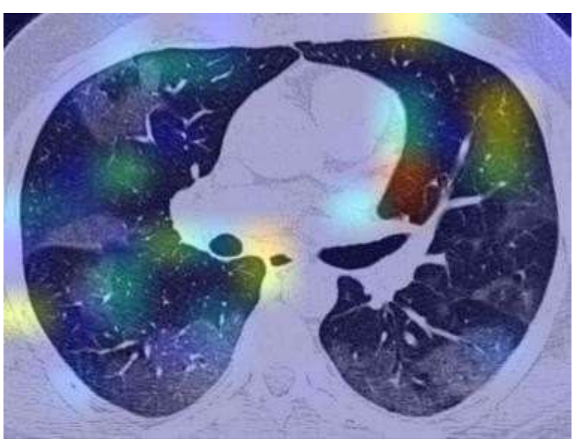
5. *Grad-cam to Visualize Abnormalities:*

After training our model, it is time to test how well it is performing.

Testing is used to evaluate the generalizing ability of the model by giving unseen data to it. The test data is given to the model in this phase.

Now that we have built an image classification model, to know how well our model is predicting, build trust, and integrate the model in our life, we need to have more model transparency to understand why it predicts.

For that, we will be using an algorithm known as the gradient-based class activation map technique (GRAD CAM) to visualize the regions of abnormalities for the target class as predicted by our model.



Highlighted regions in this indicate abnormality

ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023
https://doi.org/10.5281/zenodo.12707447

This algorithm uses the gradient information flowing into the last convolutional layer of the CNN to visualize and understand each neuron for a decision of interest and obtain the class discriminative localization map.

III RESULTS

We, therefore, need to focus on metrics, such as:

1. *Sensitivity*, or the true positive rate. Sensitivity is the proportion of true positives out of the total number of positive samples or, for a better understanding, the number of coronaviruses infected patients that our algorithm correctly classified as positive.

If sensitivity is too low, many people have the virus to whom our algorithm is classifying as negative. Less sensitivity is particularly alarming. Because of that, it would allow many infected people to go home and socialize, spreading the virus.

2. *Specificity*, or the true negative rate: This is the proportion of the true negatives to the total number of negative samples or number of non-infected people that we correctly classify as no covid.

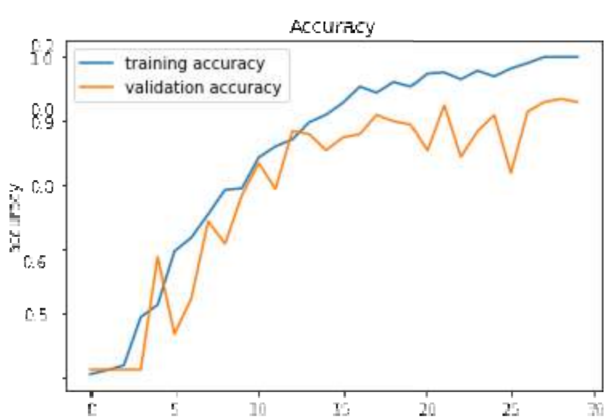
Having a specificity that is too low would mean that we will be telling many people who do not have the virus that they do.

We also use metrics like **precision** (of all the patients we have diagnosed as positive, how any people had the disease; this is useful for measuring how resourceful our test is) and the **F1 Score** (which combines Precision andSensitivity).

3. *Accuracy*

The accuracy is the ratio of the number of correct predictions to all predictions made. Accuracy = No. of correct predictions / Total no. of Input samples

The graph of accuracy on training data and validation data during the training process is as follows:

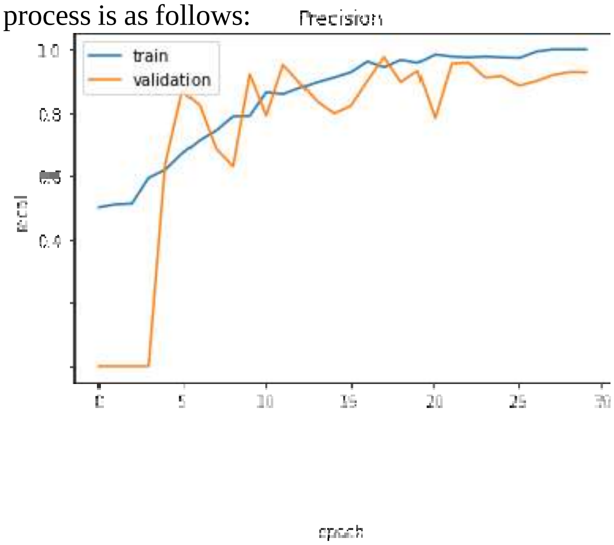


ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023
https://doi.org/10.5281/zenodo.12707447

epochs

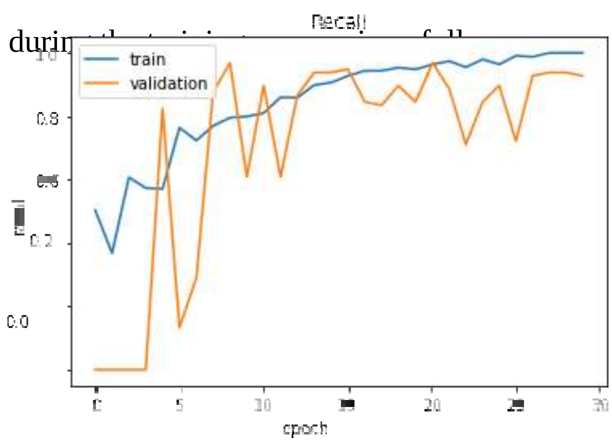
4. Precision

Precision is defined as the ratio of True Positives to the sum of True Positives and False Positives. The graph of precision on training data and validation data during the training process is as follows:



5. Recall

The recall is defined as the ratio of True Positives to the sum of True Positives and False Negatives. The graph of recall on training data and validation data durin



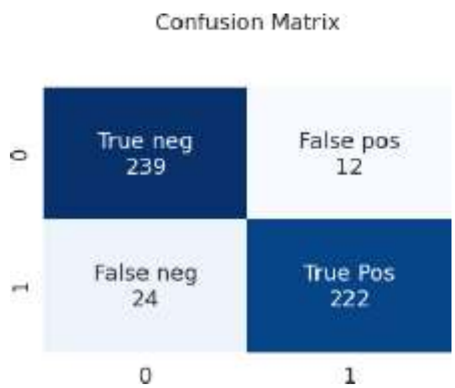
ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023

https://doi.org/10.5281/zenodo.12707447

Confusion Matrix

A confusion matrix is used to describe the performance of a classification model on test data. The confusion matrixuses the actual labels and predicted labels and how many are correctly and wrongly predicted.

The Figure of Confusion Matrix is as follows:



0 indicates no covid 1 indicates covid

Finally, we successfully achieved

f1 score = 0.925
Precision = 0.949
Specificity = 0.952
Sensitivity = 0.902

CONCLUSION

Using CT images of the lungs, this effort aims to reliably forecast the presence of COVID-19. We advocated for the study and implementation of a tailored deep learning model for the comprehensive analysis and classification of lung CT data. The test results showed a very high level of accuracy, 93%. The medical community stands to benefit greatly from this model's potential as a tool for the early identification and diagnosis of COVID-19. Deep learning algorithms have the potential to provide light on how to better identify COVID-19 and enhance testing and therapy for patients. Extending our project to accurately identify and quickly diagnose additional illnesses in other organs using CT scans is a potential future upgrade. The medical sector would greatly benefit from it.

ISSN: 2096-3246
Volume 55, Issue 01, Mar, 2023

https://doi.org/10.5281/zenodo.12707447

REFERENCES

1. An AI-Powered Coronavirus Screening System for 2019 Pneumonia 2. In May 2020, the International Atomic Energy Agency published an article on how the COVID-19 virus is discovered using real-time RT-PCR.3. See https://www.mayoclinic.org/tests-procedures/covid-19-antibody-testing/about/pac-20489696. Dated May 6, 2021, by the Mayo Clinic staff. 4. A System Based on Deep Learning and CT Scan for Coronavirus Disease Severity Prediction in 2019: https://www.nature.com/articles/s41598-020-76282-0
November 5, 2020, by Jun Zhang, Liang Zhang, Lianlian Wu, and Jun Chen. 6. Covid-19 detection using machine learning with chest X-rays: ttps://www.sciencedirect.com/science/article/pii/S1877050920321621. October 2, 2020, LucaBrunese and Fabio Martinelli. 7. A Portable Chest X-Ray for COVID-19 Lung Infection: Classification by Machine Learning: https://pubmed.ncbi.nlm.nih.gov/33239006/ Tony Nguyen, Haifang Li, Lal Hussain, November 25, 2020. 8. Go to www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset.