DIAGNOSING COVID-19 THROUGH CHEST RADIOGRAPHIC IMAGES

We take immense pleasure in presenting our proposal for the development and use of our deep learning model that can diagnose COVID-19 through chest radiographic images.

About our COVID19-AI Model

The 2019 novel coronavirus (COVID-19) presents several unique features. While the diagnosis is confirmed using polymerase chain reaction (PCR), infected patients with pneumonia may present on chest X-ray and computed tomography (CT) images with a pattern that is only moderately characteristic for the human eye. Bilateral multiple lobular and sub-segmental areas of consolidation can be observed in COVID-19 patients. The model has an accuracy of 90%. Of the true positive patients, the model had a sensitivity of 100%, and of the true negative patients, the model had a specificity of 80%. The model is predicted to have an accuracy of 98% with chest CT-Scans instead of X-Rays
We are committed to using this model to diagnose COVID-19, due to its low cost and extremely high response time.

About the creators of COVID19-AI

Sumeet Motwani: Sumeet has been programming since the age of 12, and has developed a multitude of applications over the years. He has conducted seminars at Engineering Colleges throughout Mumbai and is the youngest Certified Ethical Hacker. Sumeet is proficient in Java, Python, and Haskell, and he has created several machine learning models for real world applications. He has also won several hackathons, including one at IIT Bombay, and some of his websites have thousands of users.

Siddharth Parekh: Siddharth Parekh has been studying Deep Learning since the past 3 years and has worked on several projects such as Image Detection, Movement Detection, Facial Recognition, Medical Imagery, and Deep Fake Recognition. He is the lead coder in the First Robotics Team and will be going to Sydney as part of the competition.

The Times of India has published an article regarding our model, and we have already been approached by several other media houses and start-ups.

Class XI students work on model to spot X-rays of Covid-19 patients

Vinamrata Borwankar & Yogita Rao | TNN

Mumbai: Two city teens have joined several others across the world in the fight against the Covid-19 pandemic. Class XI students of Dhirubhai Ambani International School, Sumeet Motvani and Siddharth Parekh, have used artificial intelligence to build a model they claim can differentiate between X-rays of Covid-19 patients and those of others in less than a second.

Over the past three weeks, the two have scoured the Internet for open source information to help them build their project. "We used over 150 X-rays available on Github and Kaggle, websites hosting Covid-19 related data sets. The Lancet journal had published a study to show how different the X-ray of a Covid-19 patient is from a normal one. Based on this and our knowledge with machine learning, we began work," said Parekh.

Siddharth & Sumeet

Through the model, the two have till now tested 50 X-rays, thrice, and claim an accuracy of 81%.

"Among the Covid-19 positive patients, the accuracy was 100%, but there were a few false negatives in the other set. We are trying to further build and update our model by gathering as much data as possible," said Motvani. While not peer reviewed or validated yet, the two are contacting the BMC to present their work in the coming week.

A classic case of Covid-19, with typical symptoms, can be diagnosed using a chest X-ray and blood test, said Dr VA Sajit Babu, consulting chest physician. "In the UK, doctors working in a heavy-load Covid-19 ICU use X-ray along with clinical signs and symptoms to diagnose the infection. They do not wait for the Covid-19 test to label a case. That scenario is yet to happen here. X-rays are compulsory here for people showing respiratory symptoms. But the nasopharyngeal swab test is a must to confirm all Covid-19 suspect cases," said the doctor.

Amit Sethi, an expert in medical image analysis from IIT-Bombay, said, "Many people are working on machine learning models for Covid-19 diagnosis, but a lot more needs to be done before the model is deployable." He, however, said the efforts of the Class XI students were commendable.

About our COVID-19 Detection Model

1. Current Diagnostic Methods

Molecular tests

They usually involve taking a sample from the back of the throat with a cotton swab. The doctor then sends the sample off for testing.

Serological tests

The antibodies exist in blood and tissues throughout the body. A serological test usually requires a blood sample.

Both these tests have an accuracy much lower than 90%, and present a series of drawbacks such as:

- Delays in Testing
- Increased Costs
- Multiple tests required to increase the accuracy and decrease the probability of false results
- Improper sampling may result in false negatives
- Inconsistent testing kits-The CDC recently admitted test kits they distributed resulted in inconsistent results due to a problematic reagent required for the test.

Hence, failure to detect cases after testing may prove detrimental to the patient and the people around him/her. The reason this disease has been classified as

Spar fill this up and comment on the drawbacks

Aim of our project

- To initially act as a supplementary testing option to reduce false detections and increase the accuracy.
- To act as a cheap, sustainable, and viable option for testing COVID-19 cases
- To act as an extremely fast testing solution. Our model gives results in 0.2 seconds.
- To enable open access for a much wider outreach.
- Develop a Web Application where a user can upload a radiographic image and get immediate results

Results and use cases of our model

Post developing and training our model, we tested it on a set of 31 images each COVID-19 positive and negative, and have listed the results, evaluation, and uses below:

Evaluation of our model

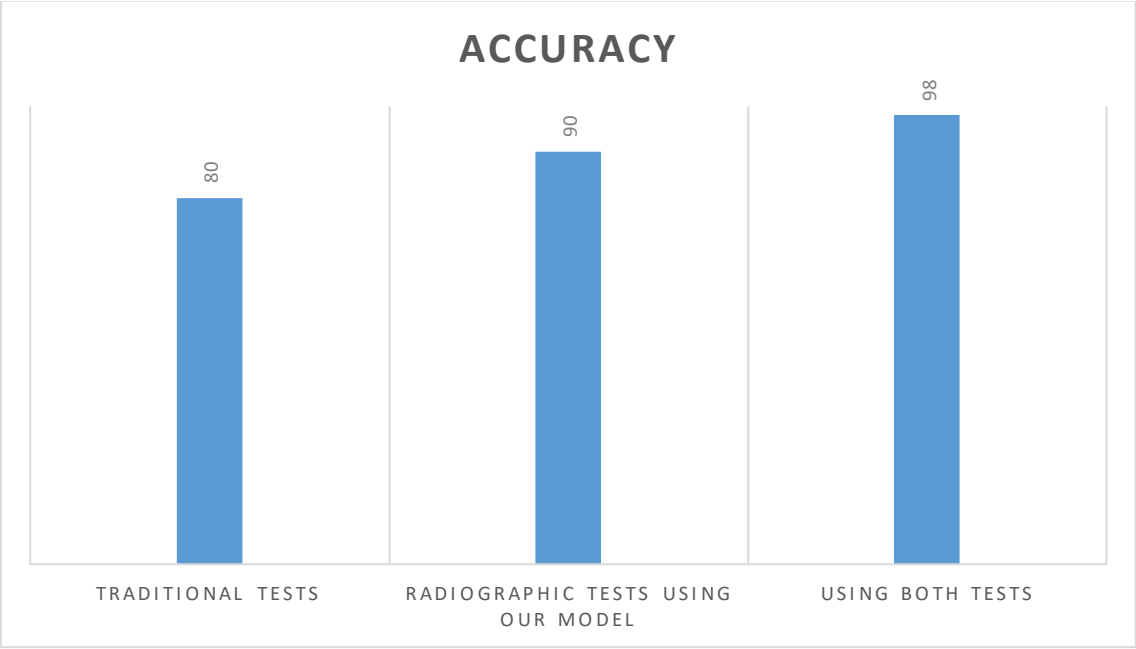
While an accuracy of 90% is very high, to act as an independent testing platform, our model needs more data for better training and development. However, with regards to diagnosing COVID-19 through radiographic images, the accuracy must be improved for large scale implementation of the model.

The model has proved to be very decisive which means that it will always be over 95% confident in its result. However, it must be noted that even the smallest deviations should be prevented. Therefore, once the model is implemented at a large scale, it should be used as a supplementary test rather than an independent test. As its use progresses, with enough data, the model can be made extremely efficient and accurate, with the accuracy touching more than 97%.

If we look at the model as a supplementary test, it reduces the probability of a false diagnosis to a very large extent.

https://www.livescience.com/covid19-coronavirus-tests-false-negatives.html

The average accuracy of traditional tests is 80%. Therefore, the rate of false diagnosis is 20%. Now, if the radiographic test is introduced as a supplementary test, since it has an accuracy of 90%, it gives a false diagnosis 10% of the time. Now, if both tests are conducted together, the probability of a false diagnosis becomes only 2%. **Therefore, up to 98% accuracy can be achieved if the model is also used as a supplementary test.**

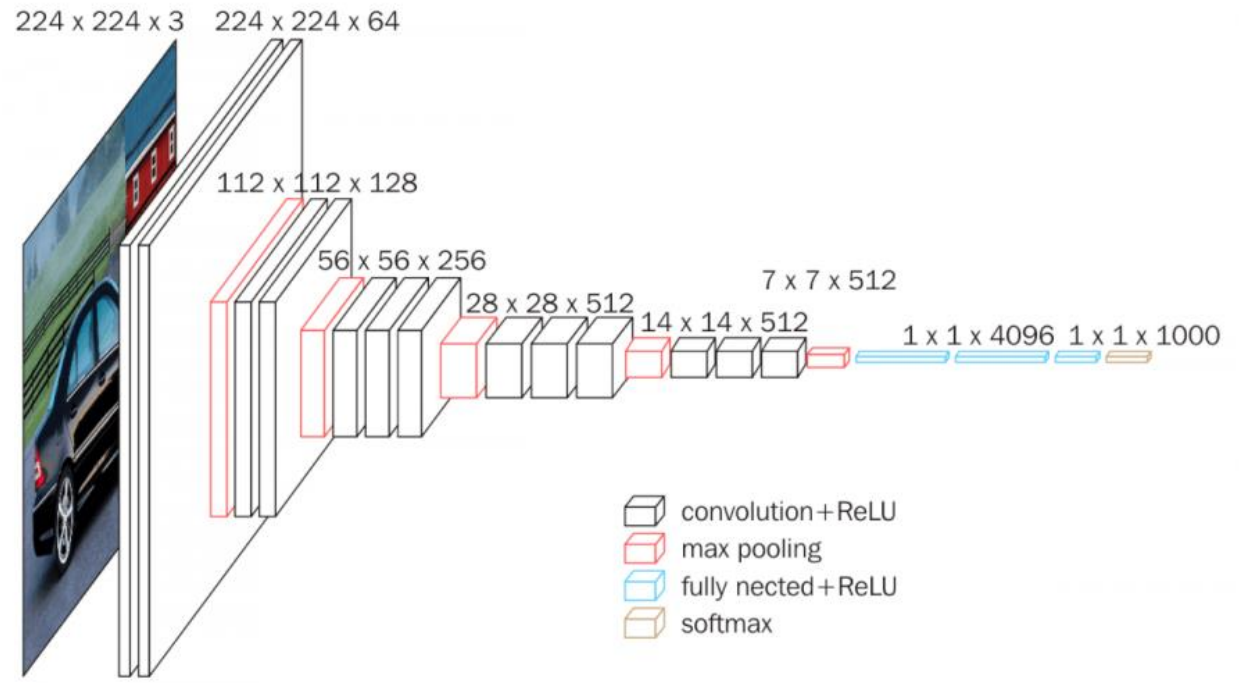


For this reason we need the help of reliable agencies to provide us with a varied and large enough dataset so that we can train our model to be extremely accurate and reliable. With enough data to work with, the model will definitely even outperform traditional testing methods in diagnosing patients.

Specifics of our model

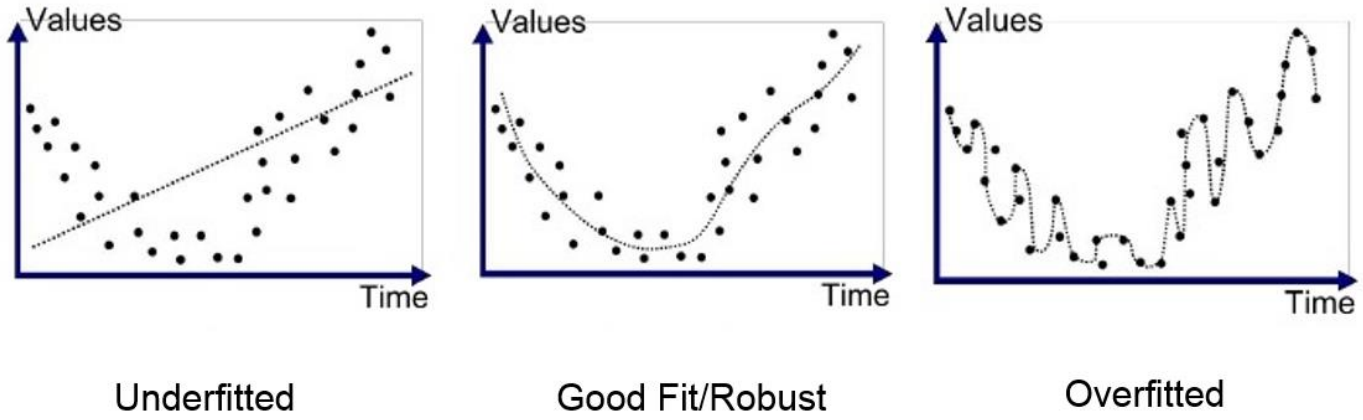
Neural Network Architecture:

We are using transfer learning which means that we are building onto a pre-trained model. The model we have chosen to build onto is the VGG16 classification model from Oxford's Visual Geometry Group. This model consists of several layers and has been trained on several different classes of images, and is quite adept at detecting features that are not very obvious to the human eye. It is considered one of the best models to use in transfer learning for image classification.



We have added a few layers of our own onto this model that are available in the Keras library in Python. We have added a MaxPooling layer,a Flatten layer, along with two Dense layers and two Dropout layers in alternation. MaxPooling reduces the data to be processed by representing a set of pixels by the one with the maximum value. This makes the features more obvious to the model. The flatten layer is used to convert the data into a single vector which is what the upcoming layers take as input. The dense layers are computational neurons that are deciding which features to pay more attention to and which to regard unimportant. The dropout layers in the middle are to reduce overfitting.

To simplify this, here is an example of overfitting along with a good



Now this example is based on regression but something similar also occurs in image classification. The one on the right is clearly overfitting, in terms of classification this would mean that there are some features that are being paid too much attention to and that are too specific and may not always be present in other images belonging to the same class.

Dropout purposely leaves out some of the data to reduce overfitting. However, this might lead to the exclusion of important features as well. So the solution to this problem, is a larger dataset with enough variance which is what we are trying to obtain. Our current model is slightly over fit to the training data since there is a significant difference in accuracy when it is tested on the training set (97%) and on the test set (89%). The accuracy on the test set is a much more reliable metric to measure a model's performance since the model has not been exposed to those images before, just like it would be in real life.

There is a final dense layer which uses the sigmoid function to give a value between 0 and 1. 0 being COVID-19 negative with 100% confidence and 1 being COVID-19 positive with 100% confidence.