JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

VIRTUAL COVID-19 PREDICTOR

Ms. Kavita Agarwal Assistant Professor[1], Ms. Vimala Manohara Ruth Assistant Professor [2]
Juhitha Doddoju Student [3], Jyotika Koneru Student[4]
Department of Computer Science and Engineering,Chaitanya Bharathi Institute of Technology, Gandipet,Hyderabad, 500075,India

Abstract: COVID-19 tests are currently hard to come by-they are simply not enough of them and they cannot be manufactured fast enough, which is causing panic. Given that there are limited testing kits, we need to rely on other diagnosis measures. The financial costs of the laboratory kits used for diagnosis, especially for developing and underdeveloped countries, are a significant issue when fighting the illness. Using X-ray images for the automated detection of COVID-19 might be helpful in particular for countries and hospitals that are unable to purchase a laboratory kit for tests or that do not have a CT scanner. This is significant because, currently, no effective treatment option has been found, and therefore effective diagnosis is critical. The paper aims to develop a system that predicts the probability of a person getting infected with the corona virus using various sources of input. To benefit wide number of users we aspire to develop a mobile application which has variety of features like having a personal datastore to ensure all the information like reports are available at a single location. Many other personalised aspects like reminders are provided along with some location-based services.

Keywords: Convolutional Neural Network, Data Augmentation

1.INTRODUCTION

COVID-19 attacks one's respiratory tract, we can use X-rays along with detailed description of symptoms like user's travel history and user's physical contact with any infected person to analyse the patient's health condition by developing deep learning and machine learning models. Automating this procedure significantly reduces the time invested in the analysis process. This paper aims to perform that automation. The Model takes the advantage of TensorFlow and keras deep learning abilities, along with scikit-learn and matplotlib for pre-processing the images in the dataset and developing a model for image classification. Apart from X-Rays, CT Scans are also considered to be a great source to perform the prediction. Finally, it can be concluded based on the inputs if the person is infected with COVID or not. Convolutional neural networks were evaluated as of their ability to detect infected patients from chest X-ray images.

2.LITERATURE SURVEY

Kiran[1]proposed an automatic COVID screening (ACoS) system that uses radiomic texture descriptors extracted from CXR images to identify the normal, suspected, and nCOVID-19 infected patients. The proposed system uses two-phase classification approach (normal vs. abnormal and nCOVID-19 vs. pneumonia) using majority vote-based classifier ensemble of five benchmark supervised classification algorithms. It revealed that radiographic images can be used and examined to know if person is infected with COVID. The study reveals that infected patients exhibit distinct radiographic visual characteristics along with fever, dry cough, fatigue, dyspnea, etc. Chest X-Ray (CXR) is one of the important, non-invasive clinical adjuncts that play an essential role in the detection of such visual responses associated with COVID infection.It has the approach of multi classification as COVID, pneumonia and None.

Alazab[2]proposed the system that examines chest X-ray images to identify such patients.Findings indicate that such an analysis is valuable in COVID-19 diagnosis as X-rays are conveniently available quickly and at low costs. Empirical findings obtained from 1000 X-ray images of real patients confirmed that proposed system is useful in detecting COVID-19 and achieves an F-measure range of 95-99%. Owing to the limited availability of chest X-ray images, dataset was generated using data augmentation. Data augmentation is an AI method for increasing the size and the diversity of labelled training sets by generating different iterations of the samples in a dataset. Data augmentation methods are commonly used in ML to address class imbalance problems, reduce overfitting in deep learning, and improve convergence, which ultimately contributes to better results. The total number of images in the dataset became 1000 after applying augmentation. A concept called Data Augmentation to increase the data available for training purpose. The first use is the ability to generate 'more data' from limited data. The second one is to avoid overfitting: For a network it is somewhat problematic to memorize a larger amount of data, as it is very important to avoid overfitting. This occurs because the model memorizes the full dataset instead of only learning the main concepts underlying the problem. [2] To summarize, if our model is overfitting, it will not know how to generalize and, therefore, will be less efficient.

The key component in deep learning research is the availability of training data sets. With a limited number of publicly available COVID-19 chest X-ray images, the generalization and robustness of deep learning models to detect COVID-19 cases developed based on these images are questionable. Thousands of readily available chest radiograph images with clinical findings associated with COVID-19 as a training data set, mutually exclusive from the images with confirmed COVID-19 cases, used as the testing data set. [3] Deep learning model based on the ResNet-101 convolutional neural network architecture was used, which was pretrained to recognize objects from a million of images and then retrained to detect abnormality in chest X-ray images. The architecture which is most suitable for building an image classifier to detect COVID from the X-ray data set provided. ResNet-101 is the architecture which is the best in terms of validation accuracy and it contains 101 network layers along with certain skip connections to ensure there is no overfit in the model developed.
CNN Architecture is the major foundation for the X-Ray classification model to be built. The different layers that are used in the model for this application is defined in Fig 1.

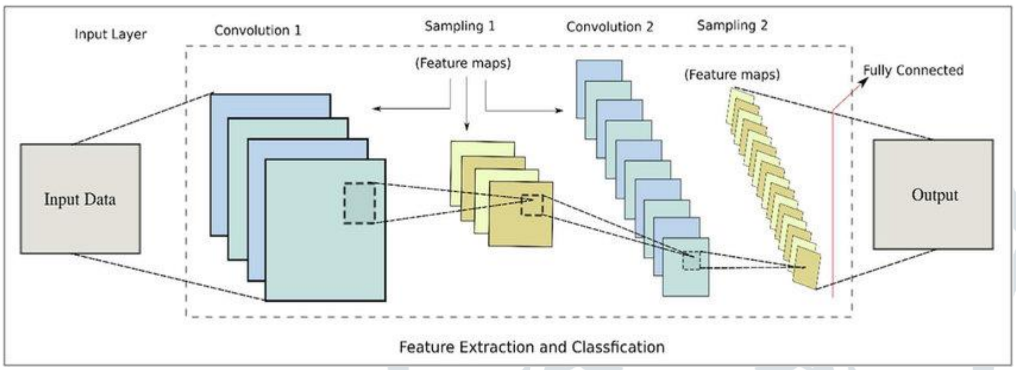


Fig 1 CNN Architecture

There are three types of layers that make up the CNN which are the convolutional layers, pooling layers, and fully connected (FC) layers. Fig 1 gives a brief understating of how of the layers are connected and where are they placed between the input and output flow. When these layers are stacked, a CNN architecture will be formed. In addition to these three layers, there are two more important parameters which are the dropout layer and the activation function

3.DESIGN

The block diagram to represent the flow of training the deep learning model for COVID prediction based on X-Ray of the user and the classification model for COVID prediction based on symptoms given as input. Fig 2 has all the preprocessing steps and the way the results are combined.

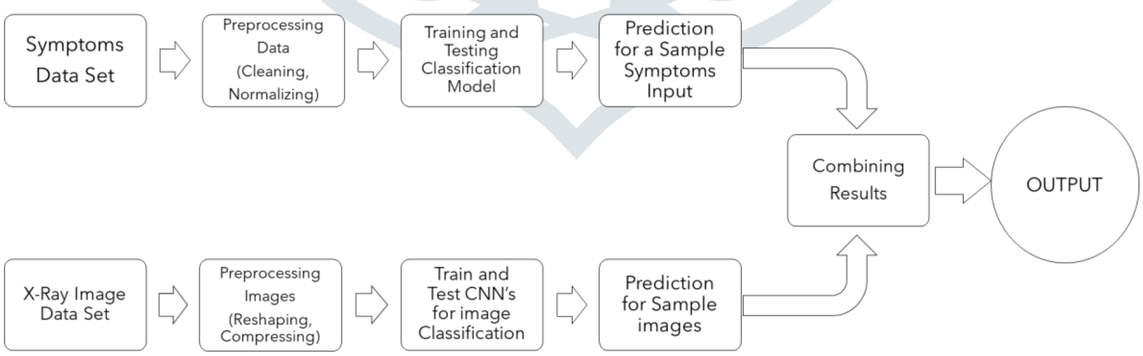


Fig 2 Block diagram representing the model working

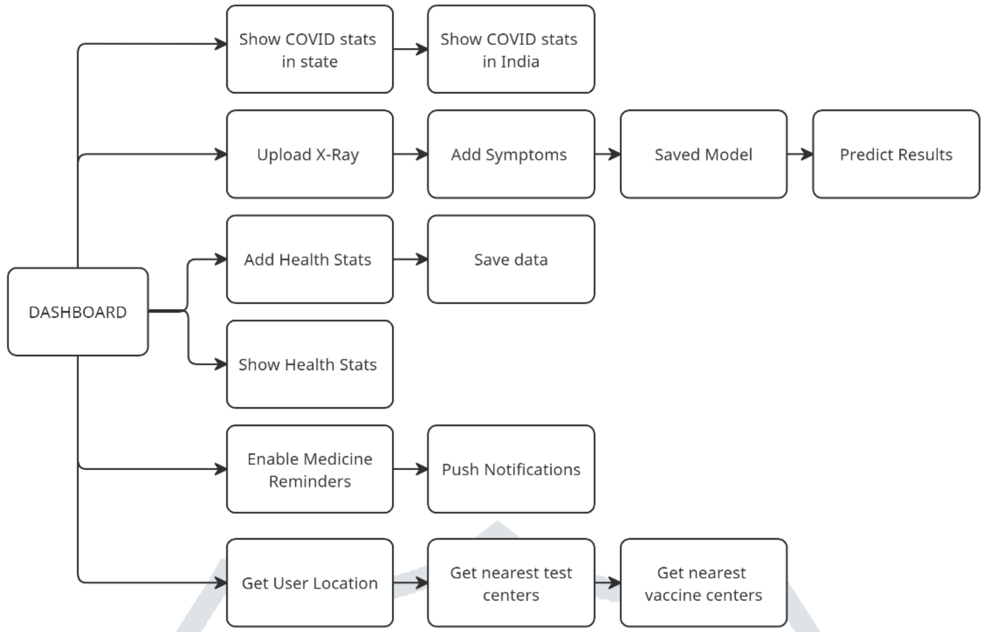


Fig 3 Block diagram for mobile application

MODULE DESCRIPTION

1. Show Real-time Data: Real-time data related to COVID-19 like total number of cases, deaths and recovery rate are represented in a graphical manner. It makes user to understand the statistics easily. It has both country wide and state wise data. The state data is provided by taking the location input from user.

2. COVID-19 Prediction: User must provide two categories of inputs as X-Ray and symptoms. User uploads an image of his/her X-Ray along with input of symptoms, travel history, contact with infected person etc. After the input is given, it is passed to the pre-trained models to fetch the output as COVID-positive or negative.
The output is majorly dependent on X-Ray's model output because there are many cases of asymptomatic patients. The two outputs are combined in the following fashion:

- If X-Ray model returns positive and symptoms classifier also returns positive. The overall result is positive.
- If X-Ray model returns negative and symptoms classifier also returns negative. The overall result is negative.
- If X-Ray model returns positive and symptoms classifier returns negative. The overall result is positive as result is majorly depending on X-ray model.
- If X-Ray model returns negative and symptoms classifier returns positive. The overall result is negative, but a warning is given to the user regarding the symptoms possessed and is asked to take additional care.

3. Health Statistics: The application provides a feature to monitor and keep a track of the user's temperature and oximeter values. This helps the user to find all his data stored at a single place and makes it easy to maintain. The data is saved in firebase and displayed to user upon request.

4. Medicine Reminders: The application also reminds and suggests the effected person regarding medicine dosages. Enabling this feature is user's choice. It is initially disabled and option of enabling it is only visible for users who update their profile as COVID positive. The notification contains the medicine name along with the dosage and is pushed at the time when user is supposed to take the medicine.

5. Nearest centres: User is also notified regarding the nearest test centres and nearest vaccine centres is given the address of it by taking users location as input. They are displayed in maps using the Google Maps API by providing markers. There are two different kind of markers each for test centre and vaccine centre. On tapping upon the markers, the name of the centre, Type and also the distance from current location is displayed in kilometres.

4.METHODOLOGY

The application involves two modules i.e., X-Ray Module and Symptoms Module. X-Ray module is built by using Convolutional Neural Network and Symptoms module is built by using Logistics Regression.

4.1 Algorithm for X-Ray module:

i. Load required packages.
ii. Define the values required for training the model.
iii. Provide the location of the data to train the model.
iv. Initialize the model to build the neural network.
v. Add a Convolutional layer with 8 filters and relu activation function.
vi. Add a pooling layer.
vii. Add a Convolutional layer with 16 filters and relu activation function.
viii. Add a pooling layer.
ix. Add a Convolutional layer with 32 filters and relu activation function.
x. Add a pooling layer.
xi. Add a Convolutional layer with 48 filters and relu activation function.
xii. Add a pooling layer.
xiii. Add a Convolutional layer with 64 filters and relu activation function.
xiv. Add a pooling layer.
xv. Add a dropout layer to avoid overfitting of the model.
xvi. Add a Dense layer to connect all the inputs to the final output with sigmoid activation function.
xvii. Compile the model using categorical_crossentropy with RMSprop optimizer with learning rate of 0.0005.
xviii. Provide the image to the compiled model to detect if the X-ray belongs to Normal person or COVID attacked person.

4.2 Algorithm for Symptoms module:

i. Import required libraries.
ii. Read the dataset.
iii. Clean the dataset to remove null values and replace categorical values into numbers.
iv. Split data for training and testing of the model.
v. Draw an instance of Logistic Regression to train the model.
vi. Provide the data for training the model.
vii. Test the model with testing data.

5.IMPLEMENTATION

5.1 Data Set Description

There are two types of datasets used one for X-Ray and symptoms.

a. X-Ray's Dataset

This dataset is organised into two folders namely COVID and Normal. COVID folder contains X-Ray images of COVID positive patients and Normal folder has X-ray images of COVID negative ones as shown in fig 3.

Huge amount of data was not available as a single dataset and hence this dataset was created by combining many datasets available. The dataset was pre-processed using the technique Data Augmentation.

The model has been trained with 15,153 files segregated into two folders namely Normal and COVID.

Layer (type)	Output Shape	Param #
conv2d_5 (Conv2D)	(None, 148, 148, 8)	224
max_pooling2d_5 (MaxPooling2	(None, 74, 74, 8)	0
conv2d_6 (Conv2D)	(None, 72, 72, 16)	1168
max_pooling2d_6 (MaxPooling2	(None, 36, 36, 16)	0
conv2d_7 (Conv2D)	(None, 34, 34, 32)	4640
max_pooling2d_7 (MaxPooling2	(None, 17, 17, 32)	0
conv2d_8 (Conv2D)	(None, 15, 15, 48)	13872
max_pooling2d_8 (MaxPooling2	(None, 7, 7, 48)	0
conv2d_9 (Conv2D)	(None, 5, 5, 64)	27712
max_pooling2d_9 (MaxPooling2	(None, 4, 4, 64)	0
dropout_2 (Dropout)	(None, 4, 4, 64)	0
flatten_1 (Flatten)	(None, 1024)	0
dense_2 (Dense)	(None, 32)	32800
dropout_3 (Dropout)	(None, 32)	0
dense_3 (Dense)	(None, 1)	33

Total params: 80,449
Trainable params: 80,449
Non-trainable params: 0

Fig 3. Model summary for X-Ray module

Data Augmentation

Data Augmentation encompasses a suite of techniques that enhance the size and quality of training datasets such that better Deep Learning models can be built using them. The image augmentation algorithms discussed in this survey include geometric transformations, color space augmentations, kernel filters, mixing images, random erasing, feature space augmentation, adversarial training, generative adversarial networks, neural style transfer, and meta-learning. To build useful Deep Learning models, the validation error must continue to decrease with the training error. Data Augmentation is a very powerful method of achieving this. The augmented data will represent a more comprehensive set of possible data points, thus minimizing the distance between the training and validation set, as well as any future testing sets.

In the X-Ray dataset which consists of many images, there was still lack of sufficient data. So, we applied the data augmentation technique as shown in Fig. 4 to make sure that model performs well for different kinds of inputs. The images are flipped to left, right and also vertically. They are zoomed for few cases and cropped in others.

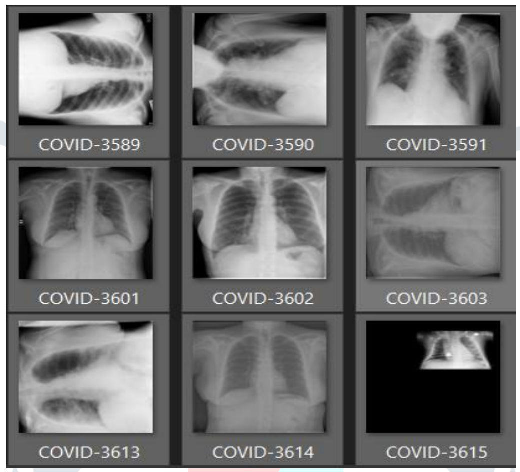


Fig. 4 Data Augmentation for X-Ray Dataset

b. Symptoms Dataset

It is a CSV file with 8 columns and around 206058 rows. The columns in the dataset are cough, fever, sore_throat, shortness_of_breath, head_ache, corona_result age_60_and_above, gender, test_indication. Out of these corona-result is the output column and test_indication contains of the remarks regarding the patient like if he has a contact with infected person or travelled abroad recently. The sample data can be seen in the Fig.5.

The pre-processing steps for this dataset are as follows:

- **Cleaning:** There were null values for few columns in some rows. Those rows were dropped.
- **Converting:** All the categorical values in columns were converted to numerical so that the data can be fed to different types of classifiers for testing.

The data distribution is an important aspect to keep a check on. It is done to ensure that the data is equally distributed, and the model is trained to predict all the possible cases with equal probability. For the symptoms dataset the data distribution is shown in the Fig. 6.

To understand the contribution of the variables to the output column, a heat map can be generated to get an idea of the relationship between the variables. This is depicted in Fig. 7.

	cough	fever	sore_throat	shortness_of_breath	head_ache	corona_result	age_60_and_above	gender	test_indication
0	1	0	0	0	0	0	0	0	2
1	1	1	0	0	0	0	0	0	1
2	0	1	0	0	0	0	1	0	0
3	1	1	0	0	0	1	1	0	1
4	1	0	0	0	0	0	1	1	0

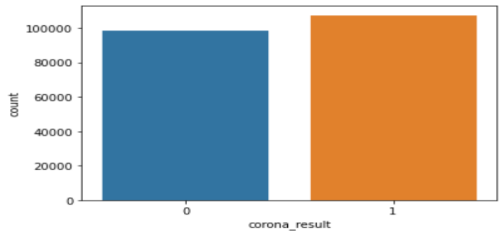


Fig. 6 Distribution of positive and negative results

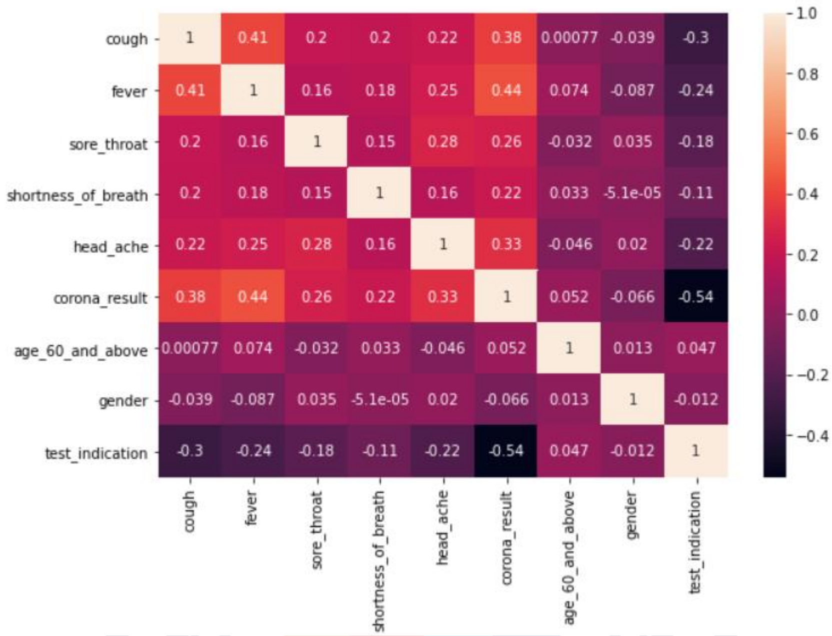
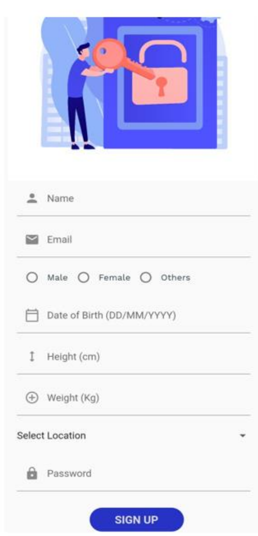


Fig. 7 Relationship among variables

6.RESULTS



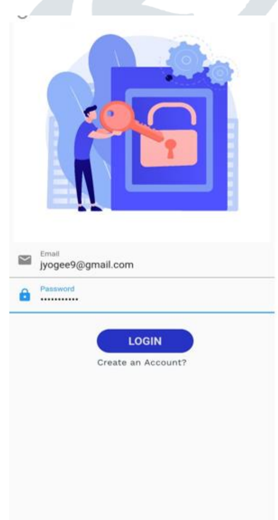


Fig 8 Register screen

Fig 9 Login screen

A user must register and login with the same credentials to use the application. Fig 8 represents the registration page for the application. Fig 9 represents the login page for the application.

Fig 10 COVID cases data screen

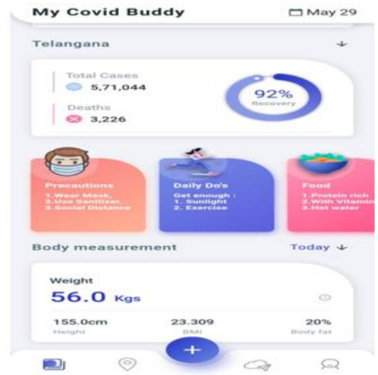


Fig 11 Body measurement screen

Fig 10 & Fig 11 gives the details of the number of covid cases across the country and state. If the user changes the state in his profile, it is reflected in this page showing the details of updated state.

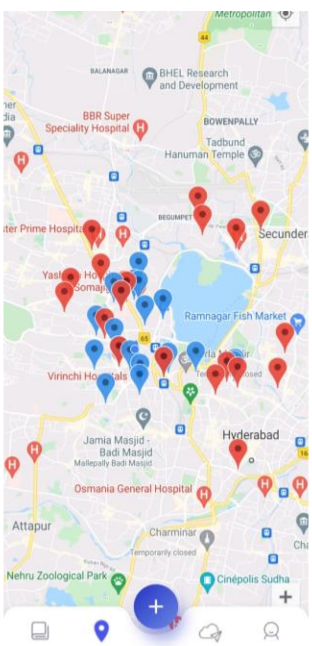


Fig 12 Maps screen

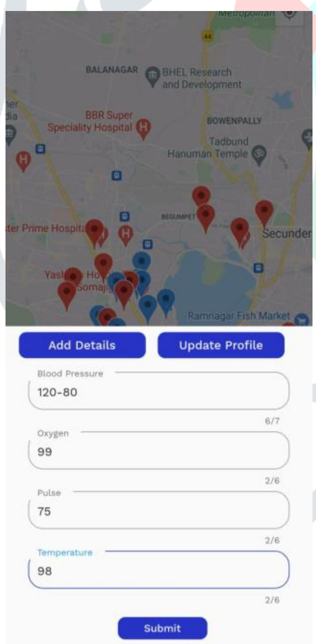


Fig 13 Add details screen.

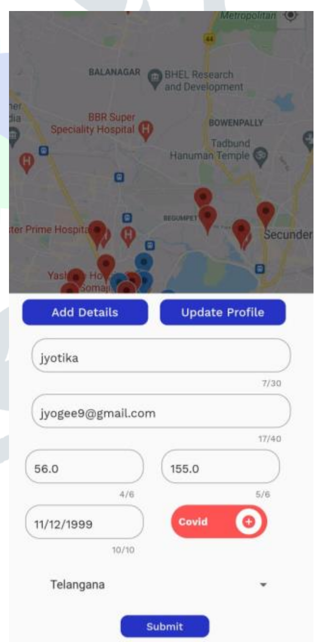
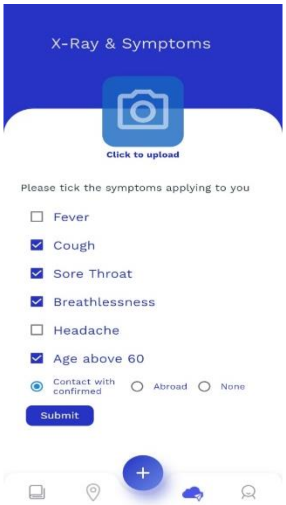


Fig 14 Update profile screen

Fig 12 gives the location of nearest test centers and vaccination centers near to user's location. There are differentiated with the color of the marker and upon tapping the marker there are details which appear. Fig 13 & Fig 14 represents the pages for uploading patient's daily updates and personal details.



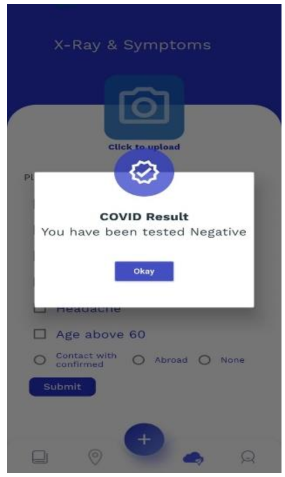
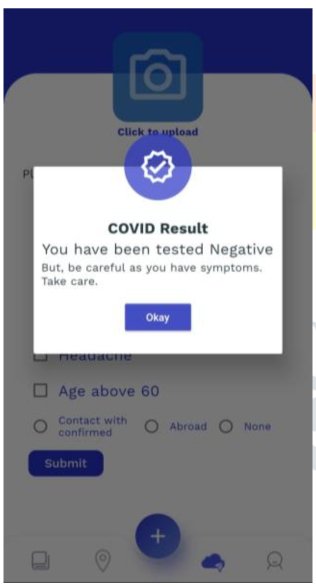


Fig 15 Input screen

Fig 16 Output screen for Covid-19 positive

Fig 15 represents the page for providing input for the models to predict if the user is infected with COVID-19 or not. Fig 16, Fig 17 & Fig 18 represents different output pages for different scenarios.



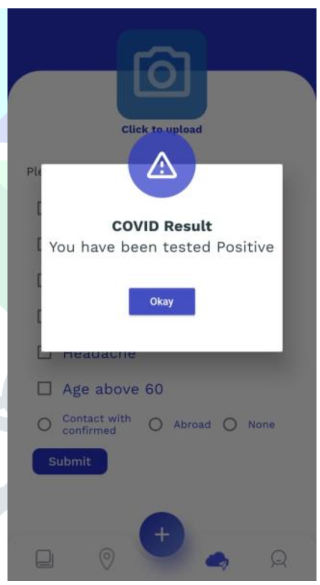


Fig 17 Output screen for Covid-19 negative with symptoms

Fig 18 Output screen for Covid-19

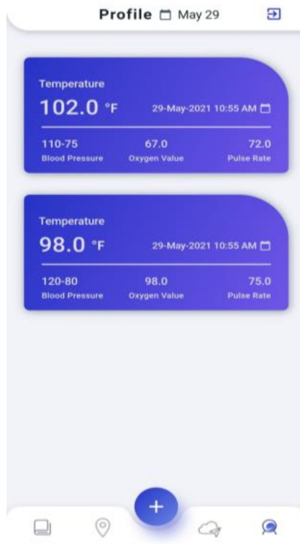


Fig 19. Health Cards screen

Fig 19 shows a page that shows health cards of the individual. These can be deleted if required. This screen also has the logout button at the top-right to log out user from the session.

5.2 Metrics

This section deals with the metrics that were considered in training the models both X-Ray based classification and symptom based one.

For the symptoms model there were multiple measures considered as shown in Fig 20. It consists of measured like precision, recall and support apart from the one used mostly which is accuracy.

	precision	recall	f1-score	support
0	0.74	0.86	0.79	25488
1	0.84	0.71	0.77	27309
accuracy			0.78	52797
macro avg	0.79	0.78	0.78	52797
weighted avg	0.79	0.78	0.78	52797

Fig. 20 metrics for symptoms model

Accuracy was the metric that was considered for the X-Ray based image classification model. The two types of accuracy training and testing are depicted in the graph as show in Fig. 21.

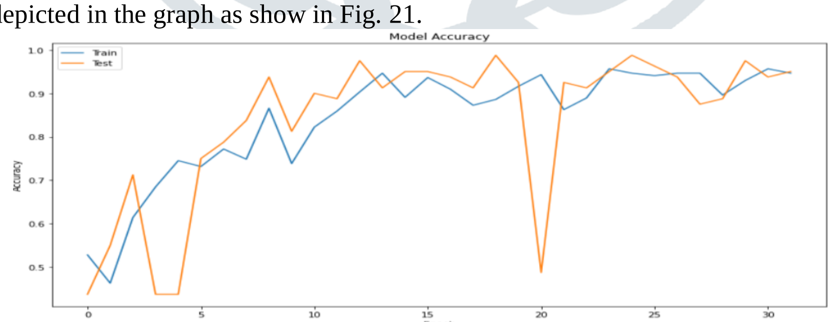


Fig 21 Accuracy levels for X-Ray Model

7. CONCLUSIONS

The application offers all the important features required for a user regarding COVID. If also offers features which help a person in the testing process and after the user is tested positive. It enables the user to have an understanding regarding the situation in his state. The application acts as quarantine buddy and helps user by reminding him of medicines and allowing him to monitor his health statistics. The limitations is the uploaded X-ray image must be of good quality. It should be having great clarity to obtain the correct results from the model. The medicine reminders feature as of now has the medicines given in the COVID home quarantine kit in India. It can only be turned on and off, but the medicines cannot be changed as required by the user.

REFERENCES

[1] Chandra TB, Verma K, Singh BK, Jain D, Netam SS. Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting-based classifier ensemble. Expert Syst Appl. 2021 Mar 1;165:113909.
Doi: 10.1016/j.eswa.2020.113909. Epub 2020 Aug 26. PMID: 32868966; PMCID: PMC7448820.

[2] Alazab, Moutaz & Awajan, Albara & Mesleh, Abdelwadood & Abraham, Ajith & Jatana, Vansh & Alhyari, Salah. (2020). COVID-19 Prediction and Detection Using Deep Learning. International Journal of Computer Information Systems and Industrial Management Applications. 12. 168-181.

[3] Mohd Zulfaezal Che Azemin, Radhiana Hassan, Mohd Izzuddin Mohd Tamrin, Mohd Adli Md Ali, "COVID-19 Deep Learning Prediction Model Using Publicly Available Radiologist-Adjudicated Chest X-Ray Images as Training Data: Preliminary Findings", International Journal of Biomedical Imaging, vol. 2020, Article ID 8828855, 7 pages, 2020. Doi: https://doi.org/10.1155/2020/8828855

[4] Kiran Purohit, Abhishek Kesarwani, Dakshina Ranjan Kisku, Mamata Dalui, “COVID-19 Detection on Chest X-Ray and CT Scan Images Using Multi-image Augmented Deep Learning Model” bioRxiv 2020.07.15.205567; doi: https://doi.org/10.1101/2020.07.15.205567

[5]. Abhishek Kesarwani, Kiran Purohit, M.D., Kisku, D.R.: Measuring the degree of suitability of edge detection operators prior to an application (2020)

[6]. Akiba, T., Suzuki, S., Fukuda, K.: Extremely large minibatch sgd: Training resnet50 on imagenet in 15 minutes. arXiv preprint arXiv:1711.04325 (2017)

[7]. Apostolopoulos, I.D., Mpesiana, T.A.: Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine p. 1 (2020)

[8]. Chen, J., Wu, L., Zhang, J., Zhang, L., Gong, D., Zhao, Y., Hu, S., Wang, Y., Hu, X., Zheng, B., et al.: Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. MedRxiv (2020)

[9] P. Vimala Manohar Ruth, Kavitha Agarwal, Dr.Y.Rama Devi, Sravya Gudipelly, Vaishnavi Kubeer, “Handwritten Telugu Vowel Character Prediction Using Convolution Neural Network”, Journal of Huazhong University of Science and Technology, Volume:50, Issue:6, 2021.

[10] Kavita Agarwal, Vimala Manohara Ruth, Kedarnath Chaturvedula, Vishal Chandra Jongoni, “Analysis of Human Traits Using Machine Learning”, International Journal of Creative Research Thoughts (IJCRT), Volume 9, Issue 6, ISSN:2320-2882, 2021

[11] Kavita Agrawal (2021) “COVID pandemic analysis using Auto- Regression- Based moving average method”, Materials Today: Proceedings, Elsevier, , doi: 10.1016/j.matpr.2021.01.710

[12] P. Rajendra Kumar,Y. Rama Devi,T. Prathima. Text Detection and Localization in Low Quality Video Images through Image Resolution Enhancement Technique. International Journal of Computer Applications 2012 by IJCA Journal Volume 58-Number 6,2012.