AI-Enabled Covid-19 Prediction Methods and Anti-Covid Strategies

Anvesh Chitturi
Dept. of Electrical Engineering
Indian Institute of Technology
Dharwad, India
180020005@iitdh.ac.in

Rahul Jashvantbhai Pandya
Dept. of Electrical Engineering
Indian Institute of Technology
Dharwad, India
rpandya@iitdh.ac.in

Sridhar Iyer
Dept. of ECE
KLE MS Sheshgiri Col.of Eng.& Tech.
Belagavi, India
sridhariyer1983@klescet.ac.in

***Abstract*— SARS-CoV-2, also known as the Coronavirus, is a disease belonging to the SARS-CoV family, first reported in December 2019 in Wuhan, China. It has since spread to several countries and has become a global threat. The issues of the Covid-19 pandemic are being tackled with the help of technological breakthroughs. Artificial Intelligence (AI) and Machine Learning (ML) play crucial roles in addressing this problem. Many organizations have been developing various devices to monitor the health parameters of a person from time to time. This work aims to study these parameter values and identify patterns within the data to predict whether a person is infected with Covid-19 or not. Different Classification algorithms such as Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, and Logistic Regression are employed. The Random Forest showed the highest performance among the algorithms mentioned above, demonstrating 99.03% accuracy. This article also suggests Anti-Covid Strategies, such as Mask Detection and Social Distancing. A Mask Detection model is constructed utilizing Transfer Learning and existing Image Classification Networks, and the best one is VGG-19 which has obtained 99.31% test accuracy.**

Keywords— Artificial Intelligence (AI); Convolutional Neural Network (CNN); Covid-19 prediction; Mask detection; Neural Network (NN); Social distancing; Transfer learning.

I. INTRODUCTION

SARS-CoV-2 is the cause of the ongoing pandemic (COVID-19). The pathogens observed were classified as the Novel Coronavirus-2019 (2019-nCoV) [1]. Since then, several mutations have taken place, leading to different strains (or variants). A few mutations have affected the spike protein of the coronavirus, which allows it to attach to the human cells more easily [2]. As a result of the mutations in the new variants, the spike protein now binds tightly to the human cells [2]. At the end of 2021, Omicron is reported as the latest variant, and there have not been many casualties; however, it is spreading faster than the previous variants.

Wearable devices in the healthcare industry can monitor vital signs such as Oxygen Saturation (SpO2), Heart Rate (HR), Body Temperature (BT), and Blood Pressure (BP). These vital signs can be tracked with the help of these gadgets, which can help detect covid indications from time to time. The data can be gathered and analysed from these devices and prevent the spread of infection. "The common symptoms of Covid-19 are fever, cough, tiredness, sore throat and shortness of breath" [3]. Some of the severe symptoms of Covid-19 are lower oxygen saturation (<90%) and higher heart rates (>100 beats/min) [4]. These values will differ depending on the age groupings.

Using the values collected from the wearable device, this article seeks to predict Covid-19 for a person. Rather than using the readings from the wearable device, a dataset with various values for each parameter is used. This dataset was generated at random from existing research on Covid and non-Covid patients' health parameters. This research seeks to provide several anti-covid measures, such as face mask detection and social distancing, in addition to predicting Covid for a person. The following is how the rest of the article is organised: Section II goes through the related Works, whereas Section III goes over the system model. The methodology is described in Section IV. The simulation results and discussions are presented in Section V. Finally; Section VI concludes the proposed work.

II. RELATED WORKS

Zhao et al. [5] used the Backpropagation NN Model to anticipate Covid-19 trends in a country and predict the number of cases over a time period such as daily, weekly, or bimonthly. Khan et al. [6] have demonstrated "The use of deep neural networks and transfer learning for the detection of COVID-19 infectees through the radio-graphs of chest X-rays". They have achieved more than 90% accuracy on the ResNet34 and DenseNet-121 models. Existing technology for mask detection uses deep learning models such as Mask-R-CNN with ResNet as the backbone, which is not ideal. Hence, Liu and Zen [7] have proposed a new approach by using the 'YOLO' model. Hou et al. [8] have used the YOLOv3, an object detection algorithm, for social distancing using deep learning. Negi et al. [9] have proposed CNN and VGG-16 based Deep Learning models, which have achieved more than 95% accuracy but on a small dataset of around 1500 images for both train and test sets combined.

As a result, the focus of this research is on face mask identification using VGG-19 on the Face mask dataset consisting of around 12,000 images for both the train and test sets combined. The wearable devices now on the market for monitoring health parameters are not being used to forecast Covid-19 for an individual, and this research attempts to use these parameters to do so.

III. SYSTEM MODEL

The image classification neural network and the object detection classifier (Haar cascade classifier) are discussed in this section. Using data from health parameters, a Covid-19 Prediction model for an individual is created.

978-1-6654-8316-2/22/$31.00 ©2022 IEEE

The proposed models are as follows:

1. A Covid-19 prediction model for an individual using health parameters data.

2. Mask detection and social distancing using image classification neural network and Haar cascade classifier.

Various health parameter values have been used in the first model, including Oxygen Saturation (SpO2 in %), Heart Rate (HR in beats/min), Respiration Rate (RR in breaths/min), Body Temperature (BT in °F), Blood Pressure (BP in mm/Hg) - systolic and diastolic, Blood Sugar levels (in mg/dL). The age and vaccination status of these patients are also considered. Then, by applying the data to different classification algorithms in Machine Learning (ML), Covid-19 for an individual may be detected.

In the second model, the existing object detection classifier – "Haar cascade classifier" along with different image classification neural networks such as "VGG-19, VGG-16, ResNet50, Xception" [10]-[13] have been used. With the help of transfer learning, one of these neural networks is used in the final model. The VGG-19 [10] model is discussed, which is something that we're particularly interested in. The "Haar cascade classifier" to detect faces in an image is used. Then the image data is applied to the model, which integrates the concept of mask detection and social distancing, which provides the output.

A. VGG-19 Architecture

"The VGG-19 is a variant of the VGG model, which consists of 19 layers (16 convolution layers, 3 fully connected layers, 5 MaxPool layers, and 1 SoftMax layer) [10]. VGG is a deep convolutional neural network (CNN) that is used to classify pictures", and its architecture is shown in Fig.1. For the VGG-19 model, 20,024,384 parameters (99.91%) are non-trainable.

B. Haar Cascade Classifier

"Object detection using Haar cascades classifiers is a practical object detection method, and it is an ML-based approach where a cascade function is trained from many positive and negative images. It is then used to detect objects in the other images" [14]. A Haar cascade model [14] is used to detect faces to obtain bounding box coordinates of a face in an image.

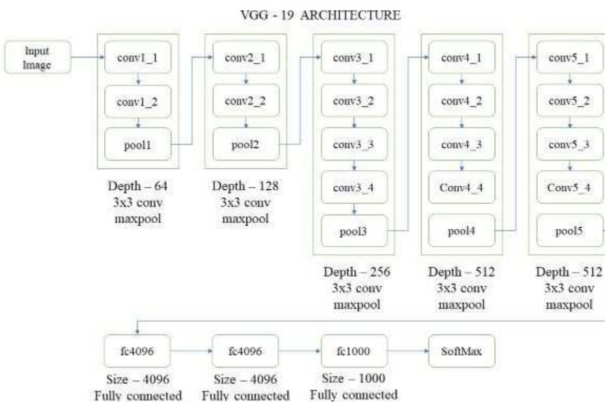


Fig. 1. VGG-19 architecture

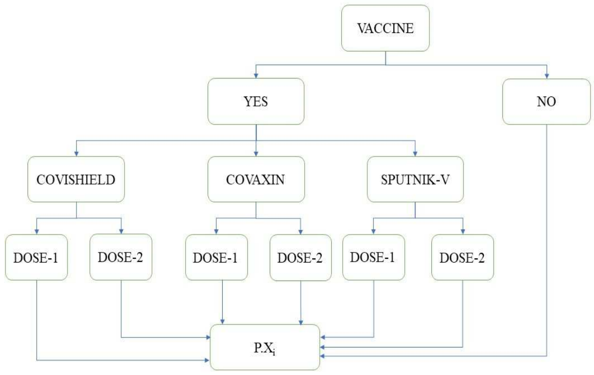


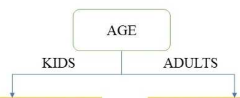
Fig. 2. Probability estimation through vaccination data

IV. Methodology

A. Covid-19 Prediction Methodology

It must be noted that the proposed Covid-19 prediction model is meant for research purposes, and it does not intend to produce the actual result. This model utilises the health parameter data to predict Covid-19 and suggests a necessary test. A random dataset has been generated for this model, which consists of two different classes in the dataset: "Covid +ve", and "Covid -ve". A dataset with a total of 1,00,000 samples is used. In the dataset, Covid +ve and Covid -ve ratio is 35:65. The dataset has been split with a test size of 20%. In the Covid +ve samples, approximately 50% are un-vaccinated, and 50% are vaccinated. Here, a majority have taken only one dose of vaccination in the vaccinated people. During the second covid-19 wave, people who took only one vaccination dose were also prone to the Covid-19 infection in India. The vaccines with one dose were assumed to have an efficacy of <60%. The effectiveness of the vaccines is subject to change against different variants and the data available. The suggested model has a workflow that examines whether or not the individual has received a vaccine. It then computes a probability value based on the vaccine dosages and is shown in Fig.2. This probability value indicates how safe a person is from Covid-19 infection. It then divides the person into one of the two categories based on their age: children (10-18 years) and adults (18-65 years). The age ranges are used for research purposes. After classifying the individual based on their age, it records the health parameter values such as SpO2, HR, RR, BT, BP - Systolic and Diastolic, Blood sugar levels. Fig.3 illustrates the health parameter values of a Covid infected person and a normal person. The readings from Fig.3 are a few accurate readings obtained from different studies [15] on the health data for a Covid infected individual and a non-infected individual.

Fig. 3. Health parammeters



PARAMETERS	NORMAL	COVID	NORMAL	COVID
SpO2 (%)	96-100	<94	94-100	<85
HR (Beats/min)	70-110	>110	60-100	>100
RR (Breaths/min)	12-20	>22	12-18	>22
BT (°F)	98-99	>100	98-100	>101
BP (mm in Hg)	110/70	>130/90	120/80	>150/100
SUGAR (mg/dL)	110-200	>220	70-110	>150

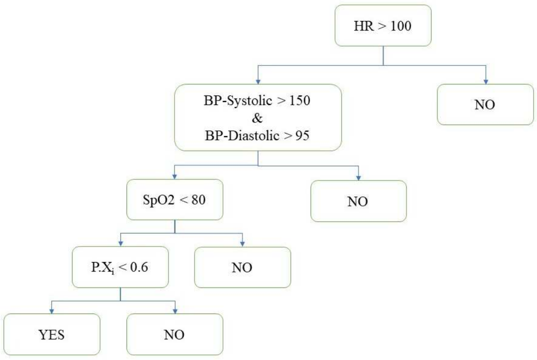


Fig. 4. Hierarchial order

Based on these values and the results of the study on Covid infected patients, four essential parameters are employed to predict if a person is infected with Covid or not. The probability value calculated from the vaccination is also considered because it impacts the likelihood of becoming infected with Covid-19. Here, if the probability value is <0.6, a Covid +ve is assumed since the vaccines with one dose have an efficacy <60%. The parameters used to classify a person into Covid-19 positive are HR, BP (Systolic and Diastolic), SpO2, and the Probability value. Fig.4 depicts the Hierarchical Order.

B. Mask Detection and Social Distancing

This paper uses the CNN and the classifier to develop the face mask detection and social distancing model, as described in Section III. This model can be fed 2D crowd photos as input and will attempt to recognise faces. Maintaining Covid standards and safety norms in public spaces can be beneficial. When a 2D image is provided as input, the model first attempts to detect the people in the image. The algorithm can then determine whether or not a person has worn a mask, indicating whether or not social distance standards have been violated. The goal of this research is to combine Mask Detection with Social Distancing, which will be effective and useful in public spaces, particularly in congested areas and densely populated countries like India, where manual supervision is difficult [16]. Here, the OpenCV [17] tool is used, which is quite useful for modifying the image data [18].

The model, as shown in Fig.5 employs the Haar Cascade "Frontal Face" classifier to detect people in a given input image. The VGG-19 transfer learning model is trained on a face mask dataset [19] comprising about 12,000 photos. The two classes in this dataset are “Mask” and “No Mask”. Roughly 10,000 photos are used to train the model for image classification, accounting for about 80% of the dataset. This model will then distinguish between people who use masks and those who don't. In the social distancing section, when the Haar cascade classifier is applied to a 2D image and a person is detected, it constructs bounding boxes around the person's face. The Euclidean distance between each bounding box is calculated if the image contains more than one person. If the distance between the centroids of these boxes is smaller than the threshold value (120 px), the Social Distancing rules are being violated. To distinguish between the social distance infractions, the colours red and green have been chosen. The social distancing and mask detection outputs are then combined within the same image. Bounding boxes in red or green surround the recognised faces in the output image, with the labels “Mask” or “No Mask”.

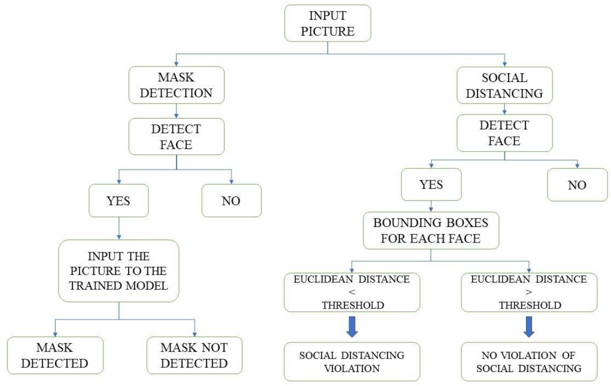


Fig. 5. Mask detection and social distancing model

V. Results And Discussion

A dataset of 1,00,000 samples is used, and multiple Classification Algorithms like Naive Bayes, Logistic Regression, Support Vector Machine (SVM), Decision Tree, and Random Forest are employed on the Covid-19 prediction model. There are 20,000 samples in the test data. Two different classes have been considered in the Dataset: “Covid +ve” and “Covid -ve”. Table1 displays a few performance metrics for the “Covid +ve” class, including recall, precision, and F1-Scores, as well as the test accuracies for different algorithms.

Random Forest has achieved a high accuracy compared to the other algorithms. Some of the reasons are less overfitting, less bias, and better handling of data imbalance than other algorithms on this data. The randomness factor of this algorithm ensures low correlation among the decision trees, which outperforms the other algorithms, including the normal decision tree algorithm. The high computational time of this technique, compared to other algorithms, is a disadvantage.

The mask detection model uses Stochastic Gradient Descent (SGD) as the optimizer and Categorical Cross-Entropy as the Loss function. Different transfer learning models were utilised, including VGG-19, VGG-16, ResNet50 and Xception [10]-[13], and their performance metrics are shown in Table 2. VGG-16 and VGG-19 are members of the same group, although the number of convolutional layers in each differs.

TABLE I. COVID-19 PREDICTION MODEL RESULTS

Algorithm	Precision	Recall	F1-Score	Test Accuracy
SGD	0.82	0.97	0.88	90.97
Naïve Bayes	0.68	1.00	0.81	82.85
Logistic Regression	0.86	0.89	0.88	90.98
SVM	0.89	0.93	0.91	93.50
KNN	0.87	0.94	0.90	92.74
Decision Tree	0.97	0.96	0.97	97.63
Random Forest	0.97	1.00	0.99	99.03

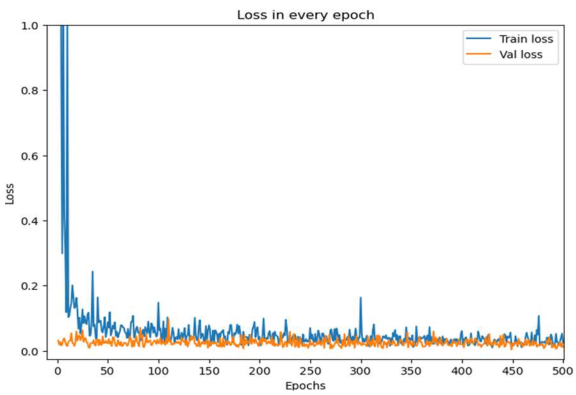


Fig. 6. Loss vs. epochs

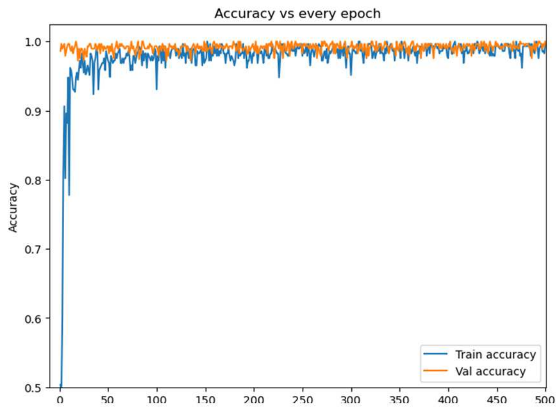


Fig. 7. Accuracy vs. epochs

For the VGG-19 model, Fig.6-7 show the Loss vs. Epochs graph and the Accuracy vs. Epochs graph, respectively. The Validation loss of the VGG-19 model is 0.0217 for 500 epochs, and it achieves an accuracy of 99.31% on the test data. Few findings from the integrated social distancing and mask detection model are included. In the photos in Fig.8-9 , a bounding box can be observed surrounding each face identified. A text label indicates whether the person has worn a mask on the top of each box. The box's colour shows the severity of the violation of social distancing. If the box is red, it means there has been a breach of social distancing. There is no breach of social distance if the box is green.

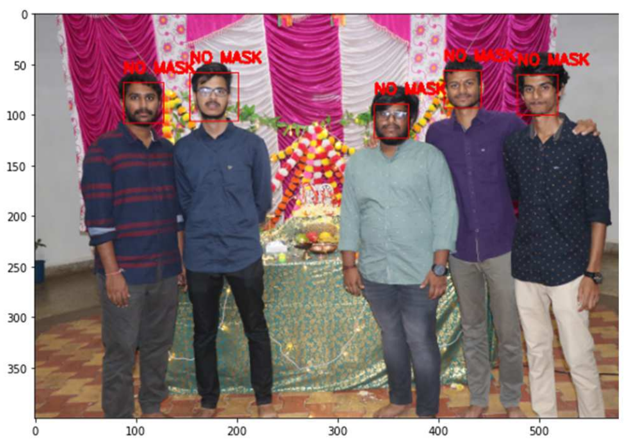


Fig. 8. Result-1

TABLE II. MASK DETECTION MODEL RESULTS

Neural Network	Validation Loss	Test Accuracy
ResNet-50	5.5920	86.37
Xception	0.0024	98.97
VGG-16	0.0219	99.25
VGG-19	0.0217	99.31

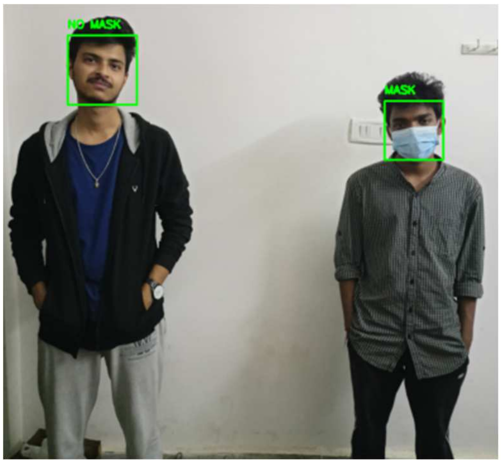


Fig. 9. Result-2

VI. CONCLUSION AND FUTURE WORK

The Covid-19 prediction model can predict with higher than 90% accuracy when using most Classification Algorithms. With a 99.03% accuracy, Random Forest is the most accurate of all the Classification algorithms examined. Mask Detection Model detects masked faces with high accuracy of 99.31% with the VGG-19 model. For the Mask detection model the VGG-16, VGG-19, and Xception models have almost identical accuracies but VGG-19 had less no. of trainable parameters among the models used. As scope for future research, one can develop a Crowd Counting Model that captures the number of individuals in a given image and integrates it with Mask Detection and Social Distancing models.

REFERENCES

[1] M. F. Baghizadeh, "What dentists need to know about covid-19,"Oral Oncology, vol. 105, p. 104741, Apr. 2020.

[2] A. Zuccarini, "Covid-19 variants," [Accessed on 03-01-2022]. [Online]. Available: https://epicentre.org.za/2021/12/14/diff-covid-variants-expl/

[3] A. Department of Health, "Covid-19 Symptoms," [Accessed on 25-02-2022]. [Online]. Available : https://www.health.gov.au/health-alerts/covid-19/symptoms-and-variants

[4] R. Stojanovi ́c, A.ˇSkraba, and B. Lutovac, "A headset like wearable device to track covid-19 symptoms," in 2020 9th Mediterranean Conference on Embedded Computing (MECO), Jun. 2020, pp. 1–4.

[5] H. Zhao, Y. Li, S. Chu, S. Zhao, and C. Liu, "A covid-19 prediction optimization algorithm based on real-time neural network training—taking italy as an example," in 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Apr. 2021, pp. 345–348.

[6] A. Khan, S. Younis, and H. Algethami, "Covid-19 identification using deep neural networks," in 2021 International Conference of Women in Data Science at Taif University (WiDSTaif), Mar. 2021, pp. 1–7.

[7] R. Liu and Z. Ren, “Application of yolo on mask detection task,” in 2021 IEEE 13th International Conference on Computer Research and Development (ICCRD), Jan. 2021, pp. 130–136.

[8] Y. C. Hou, M. Z. Baharuddin, S. Yussof, and S. Dzulkifly, “Social distancing detection with deep learning model,” in 2020 8th International Conference on Information Technology and Multimedia (ICIMU), Aug.2020, pp. 334–338.

[9] A. Negi, K. Kumar, P. Chauhan, and R. Rajput, “Deep neural architecture for face mask detection on simulated masked face dataset against covid-19 pandemic,” in 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Feb. 2021, pp. 595–600.

[10] A. Kaushik, “Vgg-19,” [Accessed on 12-01-2022]. [Online]. Available: https://iq.opengenus.org/vgg19-architecture/

[11] A. K. Dash, “Vgg-16,” [Accessed on 24-02-2022]. [Online]. Available: https://iq.opengenus.org/vgg16/

[12] A. Kaushik, “Resnet50,” [Accessed on 24-02-2022]. [Online]. Available: https://iq.opengenus.org/resnet50-architecture

[13] Z. Akhtar, “Xception,” [Accessed on 24-02-2022]. [Online]. Available: https://iq.opengenus.org/xception-model/

[14] OpenCV, “Object detection,” [Accessed on 12-01-2022]. [Online]. Available: https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html

[15] L. Lonini, N. Shawen, O. Botonis, M. Fanton, C. Jayaraman, C. K.Mummidisetty, S. Y. Shin, C. Rushin, S. Jenz, S. Xu, J. A. Rogers, and A. Jayaraman, “Rapid screening of physiological changes associated with covid-19 using soft-wearables and structured activities: A pilot study,” IEEE Journal of Translational Engineering in Health and Medicine, vol. 9, pp. 1–11, Feb. 2021.

[16] P. Shukla, R. Kundu, A. Arivarasi, G. Alagiri, and J. shiney, “A social distance monitoring system to ensure social distancing in public areas,” in 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Mar. 2021, pp. 96–101.

[17] OpenCV, “Opencv tool,” [Accessed on 12-01-2022]. [Online]. Available: https://opencv.org/

[18] M. Sharma, “Open-cv social distancing intelligent system,” in 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Dec. 2020, pp. 972–975.

[19] A. Jangra, “Face mask dataset,” [Accessed on 21-02-2022]. [Online]. Available : https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset/metadata