Face Mask Detection

Aarav Gupta   
Computer Engineering (COE-20)  
Thapar UniversityJammu ,India  
agupta61\_be20@thapar.edu

Ravi Singh  
Computer Science and Engineering (CSE-11)  
Thapar UniversityPunjab,India  
Hrithik Kakkar  
Computer Engineering (COE-19)  
Thapar UniversityUttar Pradesh,India

Lakshay Kadam  
Computer Science and Engineering(CSE-2)  
Thapar UniversityDelhi,India  
lkadam60\_be20@thapar.edu

*Abstract*—COVID-19 has made wearing face masks a part of everyone's daily lives. Making sure the people wear masks inside stores and public spaces has become a priority. As well, being able to determine if someone is wearing a mask is important for contact tracing and the transmission of COVID. This was the motivation to create a face mask detection model that can detect face masks in real-time.

Keywords- Covid 19, Deep Learning , Tenserflow,Dataset ,DCNN,MobileNet

# Introduction

Prior to the coronavirus disease 2019 (covid-19) pandemic, there was no concrete evidence supporting the use of community masks to reduce the spread of respiratory infections. Masks are primarily intended to prevent the wearer from spreading the viral droplets (source control). Covid-19 and other respiratory infections spread primarily through inhalation of respiratory aerosols .

The virus propagates and migrates down the respiratory tract and may lead to pneumonia, acute respiratory distress syndrome (ARDS) and even death. Scientists have attempted to build automated facial mask recognition systems in public locations to ensure the use of face masks in common areas. Following the COVID-19 epidemic, other researchers developed their own techniques for monitoring face masks in common areas. Employing image processing algorithms, Surveillance systems are utilized for monitoring public spaces in order to guarantee that no one’s face is visible in crowded locations . Deep learning-based approaches for object identification and imagery analytics have been increasingly popular over the years. The majority of the past research has been conducted using convolutional neural network models. There are two instances in which current face mask detection algorithms are unable to reliably identify the masks. When there is a large number of people in a single image or video frame, it is difficult to precisely identify all of the faces “with mask and without mask.” In our nation, ladies wear half-faced veils that serve the same purpose as face masks, but the current methods do not identify them as face masks.

How to construct a more efficient and accurate classification approach is a key aspect for the implementation of facial mask detection techniques in a mobile environment. However, several deep learning models are costly and time-consuming in their evaluation steps, making them unsuitable for mask detection in the facial image paradigm in a mobile environment. In order to overcome the shortcomings of the existing approach, the suggested method makes use of Depthwise Separable Convolutions with MobileNet for mask detection in facial images . Depthwise separable convolution  was first proposed in  and is now widely used in image processing for classification tasks .

# Related work

Face mask detection is a subset of object recognition that uses image processing algorithms. Digital image processing may be divided into two broad categories: classical image processing and deep learning-based image analysis. As opposed to classical image analysis, which uses complex formulas to recognize and interpret pictures, deep learning-based approaches utilize  models that mimic the workings of the human brain. Deep Learning models have been used in the majority of past research. After correctly recognizing the face in the picture or video, the CNN based approach evaluates if the face has been disguised. It is also capable of identifying a moving face and a mask in a video as a surveillance job performance. Accuracy is great with this method. An algorithm called YOLO-v3 was developed by Bhuiyan   to identify face masks in public spaces. They trained the YOLO-v3 model on their own custom dataset of photos with people labelled as “mask and non-mask.”  It is necessary to create a CNN model that can distinguish between ROIs with and those without a face mask in order to extract the facial area as a ROI. The trained model’s outcome allows for implementation on low-power devices, making the mask detection method’s inclusion faster than previous strategies. To recognize people who were not wearing face masks in government workplaces, Balaji  utilized a VGG-16 CNN model developed in Keras/TensorFlow and Open-CV to detect people who were not wearing face masks. To compensate for the model’s light weight, Fan  offered two additional methods. A unique residual contextual awareness module for crucial face mask regions Two-stage synthetic Gaussian heat map regression is used to identify better mask discrimination features. Ablation research has found that these strategies can improve feature engineering and, as a result, the effectiveness of numeric identification. For AIZOO and Moxa3K, the suggested model outperforms prior models.

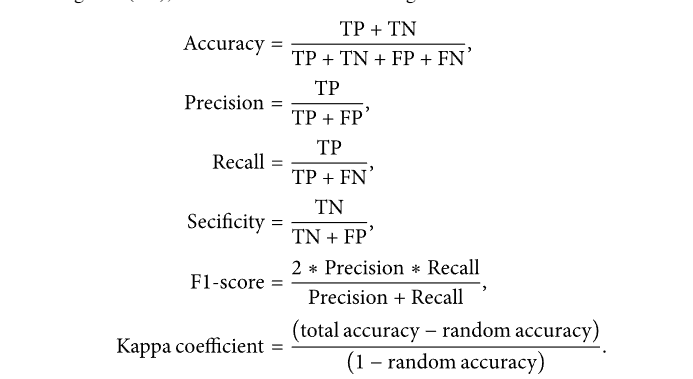
In this study, a Depth wise Separable Convolution Neural Network-based MobileNet for the detection of face masks by classifying facial images is developed in this study in an effort to answer the shortfalls of previous research in this area . Our technique improves the work performed  by replacing the conventional convolution with a depth-wise separable convolution in the neural network . Table 2shows a tabular summary of selected earlier works.

|  |  |  |  |
| --- | --- | --- | --- |
| Study | Technique | Results | Limitations |
| Toppo et al. (1) | Mobile NetV2 | 88% | Revised parameter settings can improve the system performance |
| Kaur et al. (2) | CNN-based approach | 86% | Light weight DWS-based CNN can provide more  efficient results |
| Fan et al. (3) | Residual contextual awareness module | 91% (Acc.) | Due to the constraints of the datasets, more processing is necessary to generate visualizations. |
| Bhuiyan et al. (11) | YOLO-v3 model | 86% (Acc) | YOLOv4 needs to be compared using the proposed model. |
| Mata (12) | CNN model | 60 % (Acc) | More effective techniques required for improved results |
| Balaji et al. (13) | VGG-16 CNN | N/A | DWS solution can provide better results |
|  |  |  |  |

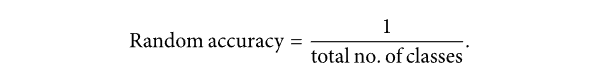
## Equations

##### Evaluation Metrics

The performance of the classification models on testing data was evaluated using the accuracy (Equation ([1](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), precision (Equation ([2](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), recall (Equation ([3](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), specificity (Equation ([4](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), F1-score (Equation ([5](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), and kappa coefficient (Equation ([6](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))). The F1-score is the harmonic mean of recall and precision. Recall, precision, and accuracy are computed using the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which can be calculated using the confusion metric.



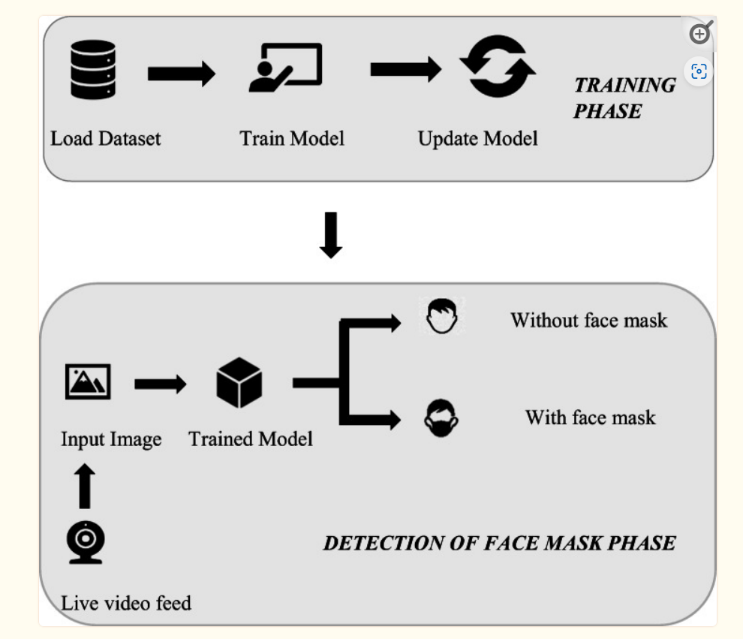
Kappa coefficient is the measure of agreement between predicted and true values in testing datasets. The value of kappa can be 0 to 1. If the value of kappa is 0, there is no agreement between the predicted and actual image, and if the value of kappa is 1, then the predicted and actual image are identical. Thus, the higher the value of kappa, the more accurate the classification. Moreover, the random accuracy for binary classification can be calculated as



# method

## General flow chart

## The implementation of the face mask detector system could be executed in two phases

****

The first phase is the training phase. This stage is initiated with the collection of the dataset. One of the most crucial steps is to have a good quantity and quality of data. One can prepare the dataset or use already existing datasets from the various available sources. If preparing yourself, the size of data could be increased by using techniques like data augmentation. Also, the data has to be cleaned before use because it plays a significant role in building a model. Various Steps involved in data cleaning are shown in Fig. [​](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9069221/figure/Fig5/)

 After obtaining a good quality dataset, the model is selected under the system’s demands and trained on the chosen dataset. Multiple techniques could be used to accomplish the target.

The live video feed could be obtained using a mobile phone, a camera, or a surveillance camera and hence could vary in format, i.e., H.265, H.264, etc.

**NOISE**There are several cases where the video frame cannot capture the images as desired. There is a possibility of the video recorded being blurred or having noise, etc. In scenarios like these, image pre-processing comes to the rescue. Further, there are several methods in OpenCV that could be used to enhance the quality of the image. For instance, blurriness could be reduced using the filter2D function of OpenCV, which enhances the sharpness of the picture. Also, image denoising techniques of the same library are helpful to deal with noisy images. Various transforms or histograms could be used for the same. Additionally, object tracking could also be considered to detect faces.

DATASET

Dataset 1 consists of 4095 images in which 2165 images with people wearing face masks and the other 1930 images with people who do not wear face masks mostly contain a front face pose with a single face in the frame and with the same type of mask having white color only.

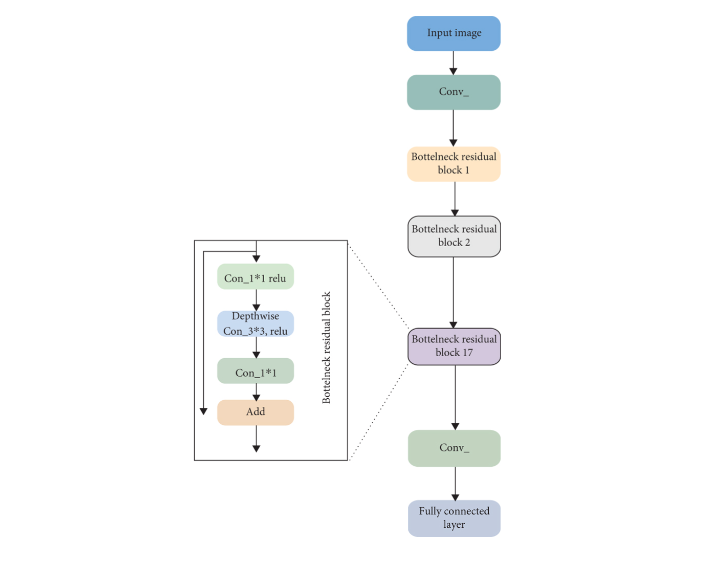


## Proposed Method

In this study, we are considering one DL architecture namely transfer learning-based MobileNetV2. For convenience, the dataset was named dataset-1 respectively. Dataset-1 contains with and without mask images (refer to Figure) for a few samples of images taken from dataset-1), respectively. Each dataset is split into two groups, one for training the models, while, the other for testing the models. In the case of training MobileNetV2 architecture, 80% data of each dataset was used, whereas, the remaining 20% data was used for testing the model. Data augmentation technique was used to increase the amount of data by making slight changes like resizing, zooming, and rotating the images. This technique helps to reduce the problem of overfitting during training the model. We resized images to, rotated images to  degrees, and zoomed images using a  zoom-in factor. A schematic diagram of DCNN and MobileNetV2 for face mask detection has been presented.

MobileNet-v2 is a convolutional neural network that is 53 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.

It is a Google-based developed architecture that is pertained on 1.4 million images of 1000 classes. It is an advanced DCNN architecture that performs well on mobile devices. In MobileNetV2, we do not have to train the model from scratch, we only change the last output layers according to our domain. The architecture of MobileNetV2 is based on its previous version (i.e., MobilenetV1). To preserve the information, it introduced a new structure named “inverted residual.” The problem of information destroying in convolution blocks by a nonlinear layer applies the technique of Depthwise Separable Convolution (DSC) by using a linear bottleneck layer.

****

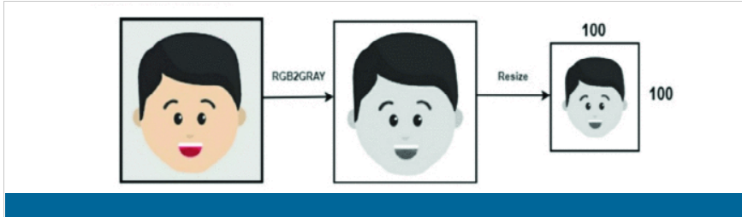
### Data preprocessing involves conversion of data from a given format to much more user friendly, desired and meaningful format. It can be in any form like tables, images, videos, graphs, etc

Data visualization is the process of transforming abstract data to meaningful representations using knowledge communication and insight discovery through encodings

The total number of images in the dataset is visualized in both categories – ‘with mask’ and ‘without mask

#### Conversion of RGB image to Gray image

Modern descriptor-based image recognition systems regularly work on grayscale images, without elaborating the method used to convert from color-to-grayscale. This is because the color-to-grayscale method is of little consequence when using robust descriptors. Introducing nonessential information could increase the size of training data required to achieve good performance



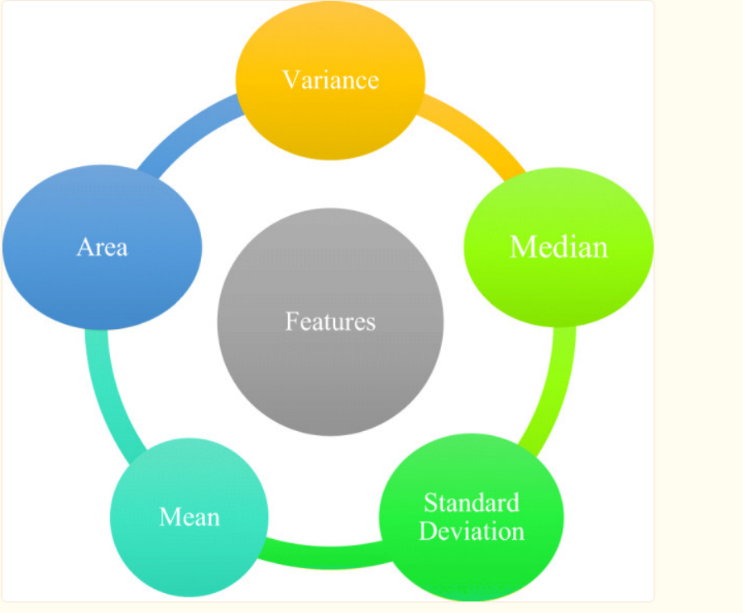
#### Image Reshaping

The input during relegation of an image is a three-dimensional tensor, where each channel has a prominent unique pixel. All the images must have identically tantamount size corresponding to 3D feature tensor.

The images are normalized to converge the pixel range between 0 and 1. Then they are converted to 4 dimensional arrays using *data=np.reshape(data,(data.shape[0], img\_size,img\_size,1)*) where 1 indicates the Grayscale image. As, the final layer of the neural network has 2 outputs – with mask and without mask i.e. it has categorical representation, the data is converted to categorical labels.

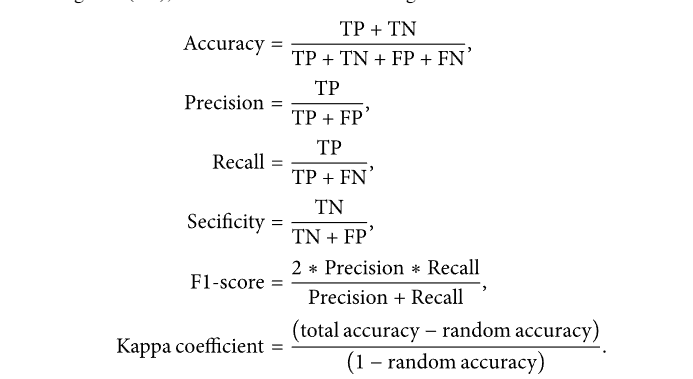
 Feature extraction

 Extraction of features is a way to get rid of unnecessary information from the data, thereby reducing the computational cost and still having imperative and relevant data reserved. Also, the reduced data helps increase the model’s learning rate. Moreover, real-time face mask detection leverages machine learning and deep learning techniques for feature extraction. In deep learning, neural networks themselves facilitate extracting features without human intervention. The input data is passed to the feature extraction network, with different backbone architectures, including MobileNetv2 and Xception. Subsequently, the result is forwarded to the classifier network categorizing a person with or without a mask. On the other hand, algorithms, like histogram of oriented gradients (HOG) and Principal Component Analysis (PCA), could be utilized to obtain features in the machine learning model.​

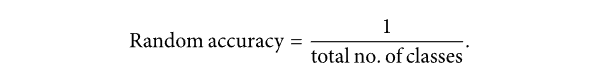


##### **Evaluation Metrics**

The performance of the classification models on testing data was evaluated using the accuracy (Equation ([1](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), precision (Equation ([2](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), recall (Equation ([3](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), specificity (Equation ([4](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), F1-score (Equation ([5](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))), and kappa coefficient (Equation ([6](https://www.hindawi.com/journals/wcmc/2022/1536318/#EEq1))). The F1-score is the harmonic mean of recall and precision. Recall, precision, and accuracy are computed using the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which can be calculated using the confusion metric.



 Kappa coefficient is the measure of agreement between predicted and true values in testing datasets. The value of kappa can be 0 to 1. If the value of kappa is 0, there is no agreement between the predicted and actual image, and if the value of kappa is 1, then the predicted and actual image are identical. Thus, the higher the value of kappa, the more accurate the classification. Moreover, the random accuracy for binary classification can be calculated as



#### Spliting  the Data

## After the pre-processing phase, the data is split into two batches, which are training data namely 80 percent, and the rest is testing data. Each batch is containing both of with-mask and without-mask images.

#### Training the Model

## The next phase is building the model. There are six steps in building the model which are constructing the training image generator for augmentation, the base model with MobileNetV2, adding model parameters, compiling the model, training the model, and the last is saving the model for the future prediction process.

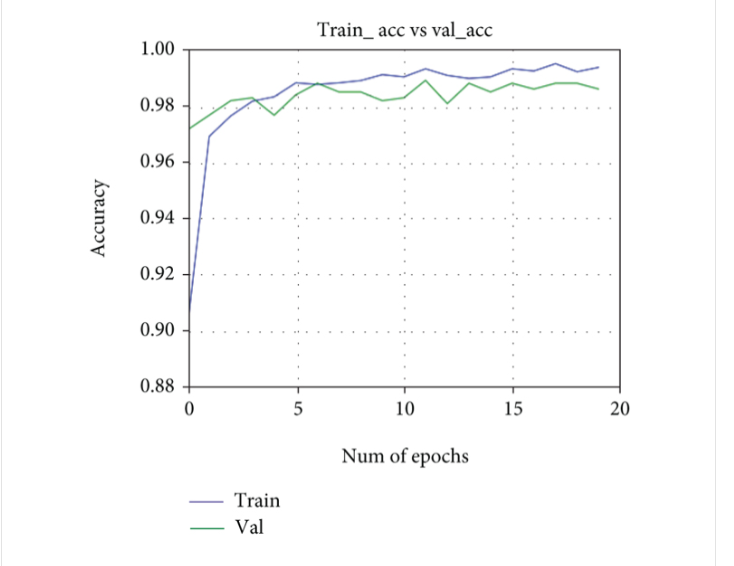
#### Implementing the model

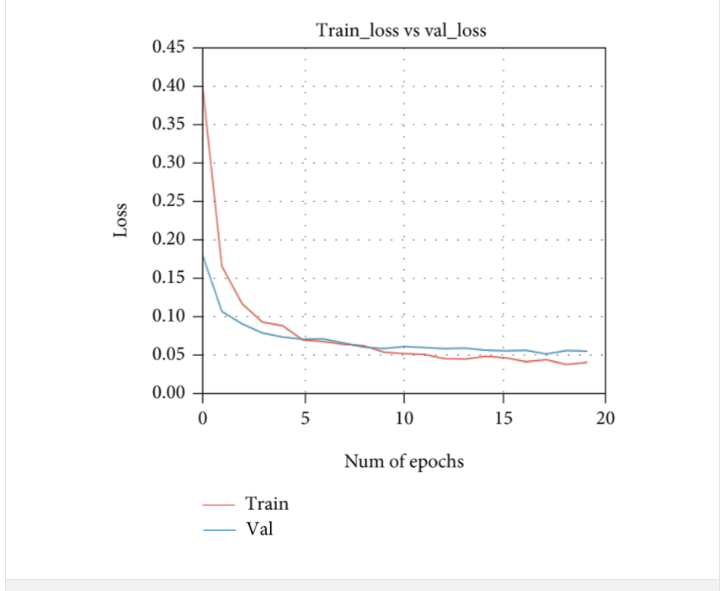
The model implemented in the video. The video read from frame to frame, then the face detection algorithm works. If a face is detected, it proceeds to the next process. From detected frames containing faces, reprocessing will be carried out including resizing the image size, converting totharray, pre-processing input using MobileNetV2.

The next step is predicting input data from the saved model. Predict the input image that has been processed using a previously built model. Besides, the video frame will also be labeled that the person is wearing a mask or not along with the predictive percentage.

##### RESULT AND ANALYSIS

On dataset-1, shows the MobileNetV2 model’s training and validation curves. Figure 1 shows the MobileNetV2 model’s training and validation curves. It shows that over 20 epochs, the training and validation accuracy achieved by MobileNetV2 are 99% and 98%, respectively. Figure 2 represents the training and validation loss curves of the MobileNetV2 model on dataset-1 which shows that over 20 epochs, the training and validation losses are 5%.



 Figures 2  further specify that there is less gap in training and validation loss curves,

which indicates that the employed model is well converged on the datasets and

there was no problem of overfitting occurring during training and validation.

CONCLUSION

COVID-19 is one of the fast-spreading viruses that have been threatening human health, world trade, and the economy. Its high mutation and spreading rate made the situation difficult to be under control.

In this paper, we briefly explained the motivation of the work at first. Then, we illustrated the learning and performance task of the model. Using basic ML tools and simplified techniques the method has achieved reasonably high accuracy. It can be used for a variety of applications. Wearing a mask may be obligatory in the near future, considering the Covid-19 crisis. Many public service providers will ask the customers to wear masks correctly to avail of their services. The deployed model will contribute immensely to the public health care system. In future it can be extended to detect if a person is wearing the mask properly or not. The model can be further improved to detect if the mask is virus prone or not i.e. the type of the mask is surgical, N95 or not.Taking precautionary measures may reduce the spreading of this virus, and one of the most important measures is to wear a face mask in public places. Therefore, in this study, a deep learning-based approach has been applied to detect the face mask automatically. The learning model MobileNetV2 transferred learning-based model, have been evaluated. The datasets consist of our own collected dataset containing 4095 images of individuals with 2165 images with mask and 1930 images without masks (dataset-1). The comparative results show that MobileNetV2achieved  and  classification accuracy on dataset-1.

REFERENCES

Paper-1 :-[Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/document/9342585)

Paper-2:-[Face Mask Detection Using Deep Convolutional Neural Network and MobileNetV2-Based Transfer Learning (hindawi.com)](https://www.hindawi.com/journals/wcmc/2022/1536318/)

Paper-3:-[Face mask detection in COVID-19: a strategic review - PMC (nih.gov)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9069221/)

Paper-4:-<https://www.researchgate.net/publication/358586028_Face_Mask_Detection_Methods_and_Techniques_A_Review>

Paper-5:- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9069221/>