

Study on Face Mask Detector System in COVID-19 Era using Deep Learning

Palash Dutta Banik*, Sagarmoy Ganguly, Ropa Roy, Asoke Nath

Department of Computer Science, St. Xavier's College (Autonomous), Kolkata

ABSTRACT

Article Info

Volume 8, Issue 2

Page Number : 327-337

Publication Issue :

March-April-2022

Article History

Accepted: 10 April 2022

Published: 25 April 2022

Due to worldwide pandemic COVID-19, there arises a severe need of protection mechanismsto prevent man-to-man infection and face mask is one of the most important protection mechanisms. The basic aim of this study is to detect the presence of a face mask on human faces on live streaming video as well as on static images. The concept of Face Mask Detection System using Convolutional Neural Networks is to provide thousands of images of masked and non-masked individuals to a computer program and then train the computer program to recognize and distinguish the individuals in those images as masked or unmasked. In the present study the authors will use deep learning to develop face detector model. The proposed technique takes place in 2 phases. The first phase includes fine tuning a pre-trained classifier with our data set. The second phase includes applying the highly trained classifier to detect faces with masks and no masks. Alongside this, we shall use basic concepts of transfer learning in neural networks to finally output presence or absence of a face mask in an image or a video stream.

Keywords: Deep Learning, Machine Learning, Artificial Neural Network, AI Models, Face-Mask Detection System, CNN, MobileNetV2

I. INTRODUCTION

In the year 2020, emerged the Covid Pandemic caused by the Corona Virus. It had a severe negative impact on the communal health and global economy. Due to limited medical resources and the absence of effective antiviral, WHO recommended several measures to control infection rate and minimize the circulation rate. Wearing a face mask is one of many such measures to ensure public safety. The mask stops the primary source of SARS-CoV2 droplets expelled by an

infected victim from reaching and infecting a healthy individual. Thus, wearing a face mask was made mandatory everywhere quickly. For contributing towards public health, this project aims to generate a highly accurate, efficient, and real-time mask detection technique that can efficiently detect faces with no masks in public, enforcing them to wear masks. Initially, face detection takes place with the help of Object Detection algorithms and then it is reverse engineered. The primary research on face detection was done in 2001 using the design of handcraft feature and application of traditional

machine learning algorithms to train effective classifiers for detection and recognition [1]. Recently, face detection methods based on deep learning CNNs have been widely developed to improve accuracy and performance. Although numerous researchers have committed efforts in designing efficient algorithms for face detection and recognition but there exists an essential difference between ‘detection of the face under mask’ and ‘detection of mask over face’.

II. OBJECTIVE

The present study aims to contribute to public healthcare by developing an efficient technique that can accurately detect face masks in public areas to minimize the spread of Coronavirus. A binary classification algorithm-based model will be developed which will be able to accurately recognise whether a face is masked or unmasked.

III. PURPOSE

The primary and fundamental purpose of this study is to develop an application with the following features:

1. Ability to accurately and effectively distinguish whether a face is masked or unmasked.
2. Ability to differ between faces covered with masks and faces covered with other objects like scarfs etc.
3. Highly modular enabling easy future upgrades and optimizations.

IV. SCOPE

The present study to develop face mask detector that can distinguish between faces with masks and faces with no masks. A system is to be developed with which users with interact by providing input in the form of images or live video stream and the system will provide the output after recognising the face followed by evaluating whether the face is masked or unmasked.

V. APPLICABILITY

Once the model is properly trained and estimated accuracy is achieved, this application can be deployed in various public areas like Movie theatres, Malls, Banks, Academic Institutions etc to detect people with and without masks enforcing everyone to wear a face mask.

VI. SURVEY OF TECHNOLOGIES

VI.1. Deep Learning

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behaviour of the human brain allowing it to “learn” from large amounts of data [2]. It is part of a broader family of machine learning methods based on artificial neural networks with representation learning represented in Figure 1[3].

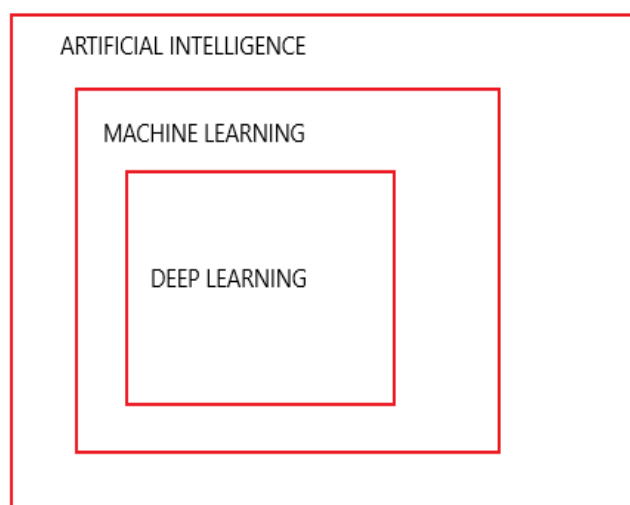


Fig 1: Deep Learning

VI.2. CNN

A **Convolutional Neural Network (CNN)** is a Deep Learning algorithm which can take an input image, assigns importance (learnable weights and biases) to various aspects/objects in the image and is able to differentiate one from the other [4]. The basic building blocks of CNN are layers and neurons with learnable

weights and biases. The layers enable the CNN to convert a 3D input volume into an output volume. Figure 2 shows a hand drawn representation of CNN layers.

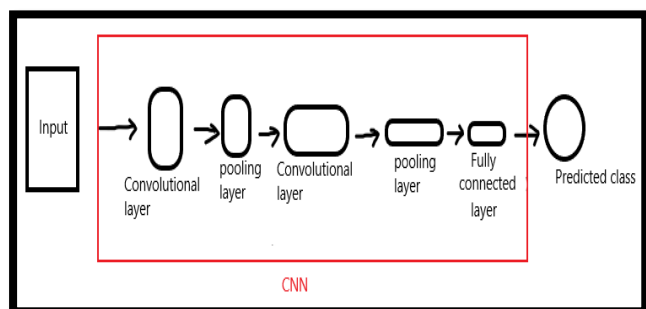


Fig 2: CNN Layers

VI.3. TensorFlow

TensorFlow is a cross- platform and open-source deep learning library exclusively designed for working with very high dimensional tensors and very equipped in parallel processing to optimally utilize modern graphics cards' parallel processing cores especially incorporated by NVidia through its CUDA cores [5]. TensorFlow was developed by Google and released for the general public in 2015.

VI.4. OpenCV

OpenCV is a library of functions aimed mainly at real time computer vision applications for faster integration of static or live camera feed and pre and post processing developed by Intel it is free to use and open source and supports cross platform integration and has support for languages like C++, Java and Python [6]. This project makes extensive use of OpenCV functions for image processing.

VI.5. Keras

Keras is an open-source neural-network library written in Python. It makes model transfer learning easier and more accessible over cross-platform applications. Many machine learning functions are pre-defined in Keras which reduces the lines of code and makes the program easier to understand [7].

VI.6. Transfer Learning

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems [8]. Figure 3 shows a Transfer Learning model.

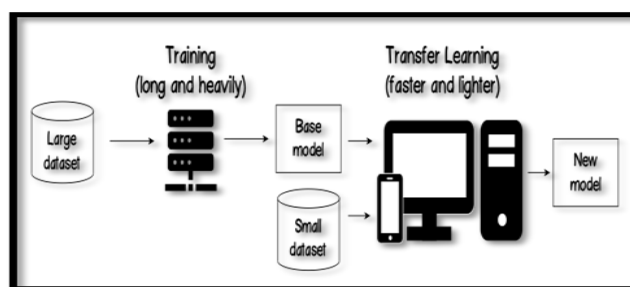


Fig 3: Transfer Learning Model

VI.7. MobileNetV2 Architecture

MobileNetV2 is a CNN architecture that seeks to perform well on mobile devices. It is based on an inverted residual structure where the residual connections are between the bottleneck layers [6]. The architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers [9]. Figure 4 shows a flowchart of MobileNetV2 architecture.

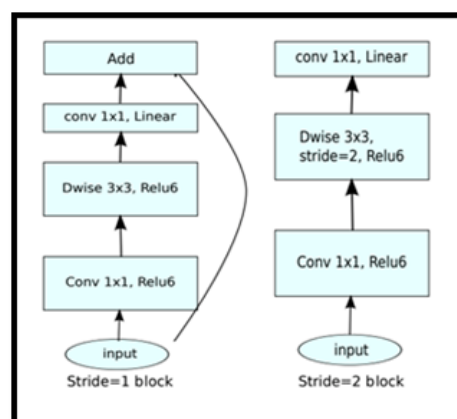


Fig 4: MobileNetV2 Flowchart

VI.8. Fine Tuning

This consists of unfreezing an entire model (or part of it), and re-training it on the new data with a very low learning rate. This can potentially achieve meaningful improvements, by incrementally adapting the pretrained features to the new data. Fine-tuning establishes a baseline model while saving considerable time [10]. Figure 5 shows a block diagram of the Fine-Tuning process.

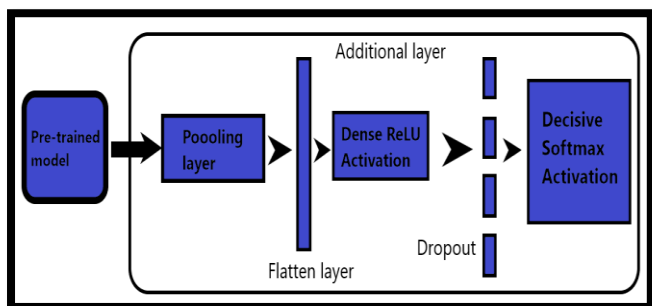


Fig 5: Fine Tuning

VI.9. Classifier

These are used for classification. Classification is the process of predicting the class of given data points, sometimes called as targets/ labels or categories. Classification predictive modelling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (Y) [11]. Figure 6 shows a graphical diagram of how a classifier works.

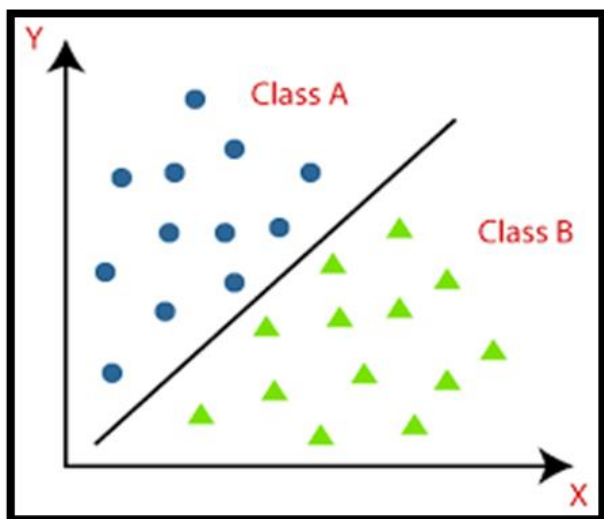


Fig 6: Classifier

VI.10. ImageNet

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. The project has been instrumental in advancing computer vision and deep learning research [12].

VI.11. CNN Layers

VI.11.1. Convolutional layer: The convolutional layer is sometimes called the feature extractor layer as features of the image get extracted within this layer. It works by placing a filter over an array of image pixels to perform convolution operation which creates a convolved feature map [13]. The output of this layer will work as input for the next layer and this layer also contains the ReLU activation to make all negative values to zero [13].

VI.11.2. Pooling Layer: The pooling layer reduces the sample size of a particular feature map, and it helps to process faster as it reduces the number of parameters that the network needs to process and it gives us a pooled feature map as an output [14]. There are two types of pooling methods - (i) **Max Pooling** - which takes maximum input of a particular convolved feature map, (ii) **Average Pooling** - which simply takes the average input of a particular convolved feature map.

VI.11.3. Fully connected Layer: Fully connected layer involves weights, biases, and neurons. It connects neurons in one layer to neurons in another layer and this fully connected layer allows us to perform classification on our dataset by training [15].

VI.11.4. Flattening Layer: This layer transforms the pooled feature map of input image into 1-d array for the computer to understand so that we can insert this data into an artificial neural network later. Thus, after the flattening layer we end up with a long vector of inputs data which can be passed through the artificial

neural network for further processing [16]. Figure 7 shows the input and output of flattening layer.

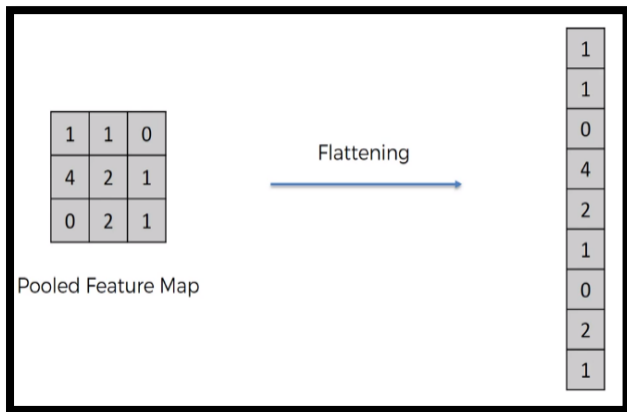


Fig 6: Flattening Layer

VI.11.5. Dropout Layer:

The Dropout layer is a mask that nullifies the contribution of some neurons towards the next layer and leaves unmodified all others. We can apply a Dropout layer to the input vector, in which case it nullifies some of its features; but we can also apply it to a hidden layer, in which case it nullifies some hidden neurons [17]. Dropout layers are important in training CNNs because they prevent overfitting on the training data. Figure 7 shows the position of dropout layer.

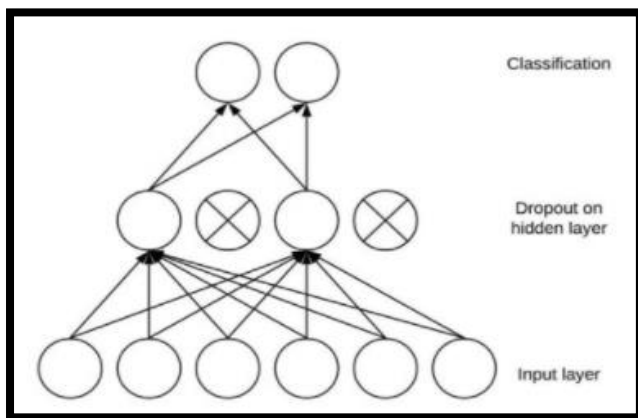


Fig 7: Dropout Layer

VII. ACTIVATION FUNCTIONS

VII.1. ReLU

The rectified linear activation function or ReLU, for short, is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output

zero [18]. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance.

VII.2. Softmax

The Softmax activation function is generally used for multiclass classification problems. The main advantage of using Softmax is the output probabilities range. The range will be 0 to 1, and the sum of all the probabilities will be equal to one. It returns the probabilities of each class, and the target class will have the high probability [19]. This function takes the results and classifies them into a binary ‘yes’ or ‘no’ value.

VIII. SYSTEM DESIGN

VIII.1. Basic Modules

Dataset: The face mask detection dataset will consist of “with mask” and “without mask” images. Thus, a two-class model of people wearing and not wearing masks will be trained to create the detector. The Training dataset has 1100 images consisting of both masked and non-masked faces in random order. The randomness in the dataset is much needed to prevent overfitting or underfitting of the model. It doesn’t allow the model to develop a particular ratio in while training. For example, if a strict order is followed in the Training dataset (2 images of unmasked face followed by one image of masked face) then the model will eventually start providing output according to this 2:1 ratio.

Fine-Tune: We fine-tune the MobileNetV2 architecture on our “Mask”/ “No Mask” dataset to obtain a classifier with high accuracy. Fine-tuning setup is a three-step process:

- Load MobileNetV2 with pre-trained weights.
- Append the newly constructed FC Head to the base in place of the old head.
- Freeze the base layers of the network. The head layer weights will be tuned during the process of backpropagation, whereas the

weights of these base layers will not be updated.

The Python script ‘train_mask_detector.py’ will accept our input dataset and fine-tune the MobileNetV2 upon it to create our ‘mask detector model’. A training history ‘plot.png’ will also be generated and it will contain the accuracy/loss curves.

Face Detection: This process is used to generate a bounding box location of the face in the image. Then we apply the facial landmarks to extract the ROI (Region of Interest). This in turn allows to localize the eyes, mouths, nose etc.

Face Mask Detection: Once the ROI is pre-processed, we perform mask detection using our highly accurate classifier to predict whether the image is “with mask” or “without mask”. Python script ‘detect_mask_image.py’ will be used to perform to mask detection in static images. Python script ‘detect_mask_video.py’ will use the webcam and apply facemask detection to every frame in the video stream.

Display: Based on the probabilities returned by the mask detector model we determine the ‘class’ label and assign an associated ‘colour’ for annotation. The ‘colour’ will be **Green** for “Mask” and **Red** for “No mask”.

VIII.2. Procedural Design

IX.3.1. Algorithm Design for face mask detector training script with Keras and TensorFlow (Training Phase)

1. import set of tensorflow.keras
2. import other necessary packages
3. construct the argument parser and parse the arguments
4. grab list of images in our dataset directory
5. initialize list of data and class images
6. pre-process the images
7. append pre-processed images and labels
8. perform one-hot encoding
9. load MobileNetV2
10. construct a new FC head
11. Freeze the base layers

12. Compile model
13. train the head of the network
14. plot the training loss and accuracy

VIII.3. Algorithm Design for implementing the face mask detector for images with OpenCV (Deployment Phase)

1. make necessary imports
2. construct argument parser and parse the arguments
3. load serialized face detector model
4. load face mask detector model
5. initialize video stream
6. detect faces and ROI in the video stream
7. detect mask or no mask
8. annotate the detection results
9. display the results

IX. IMPLEMENTATION AND TESTING

X.1. Implementation

In Phase 1 involves training the face mask classifier model and in Phase 2 involves face detection from images and live video streams and then classifying these faces as masked or non-masked by applying the trained face mask classifier model.

Phase 1 script/s and model/s

train_mask_detector.py (Python script for training the face mask classifier model)

Phase 2 Script/s and Model/s

detect_mask_image.py (script for classifying masked and non-masked faces from image)

Implementation of algorithm 2

X.2. Testing

For testing purposes, the testing was done with the face mask classifier model training script (train_mask_detector.py), the script which classifies faces from image (detect_mask_image.py), and the script which classifies faces from live video streams (detect_mask_video.py).

X.2.1. Face Mask Classifier Model Testing

Training the model with pre-defined dataset (obtained from online sources) and plotting the results.

Training another model with a custom dataset and plotting the results.

X.2.1.1. Image Testing

Given below are the test cases used.

- Image with a single, non-masked face. The facial region is clearly visible and doesn't have any kind of obstruction.
- Image with a single, masked face. The facial region is clearly visible and doesn't have any kind of obstruction except for a mask.
- Image with single non-masked face. The facial region has some kind of an obstruction but not a mask.
- Image with a single masked face. Apart from the mask itself, the facial region also has some other kind of obstruction.
- Image with multiple non-masked faces. Some faces are clearly visible while some of the faces have an obstruction but not a mask.
- Image with multiple faces. Apart from the mask itself, some faces also have other obstructions, and some faces are non-masked.

X.2.1.2. Video Testing

Given below are the test cases used.

- Video with a single subject. The video will start with a non-masked subject and with progression the subject shall wear a mask.
- Video with multiple subjects. The masked or non-masked state of the subject will be random and may or may not change with the video progression.

X. FACE MASK CLASSIFIER MODEL TRAINING

The Face Mask Classifier model was tested by training it with two datasets. The script to classify faces from image was tested with various images and the script to classify faces from live video streams was tested with

various video streams. All the yielded results are as follows.

XI.1. Face Mask Classifier Model Testing Results

The first model was trained with a dataset collected from online sources. The results were then plotted in the form of a graph given in figure 8.

The images in this dataset were pretty basic which resulted in a model that is accurate when it came to typical masked or non-masked images, but the classification wasn't that accurate when the faces in the images had some kind of obstruction like a scarf, colour, helmet etc.

To overcome the flaws of the first model, we trained a second model with a custom dataset. The dataset included the dataset used in case of the first model along with some other images which were individually collected. These custom images include images with obstructions in the facial region which was the point of weakness in the first model. The results were then plotted in a graph of similar kind shown in figure 9.

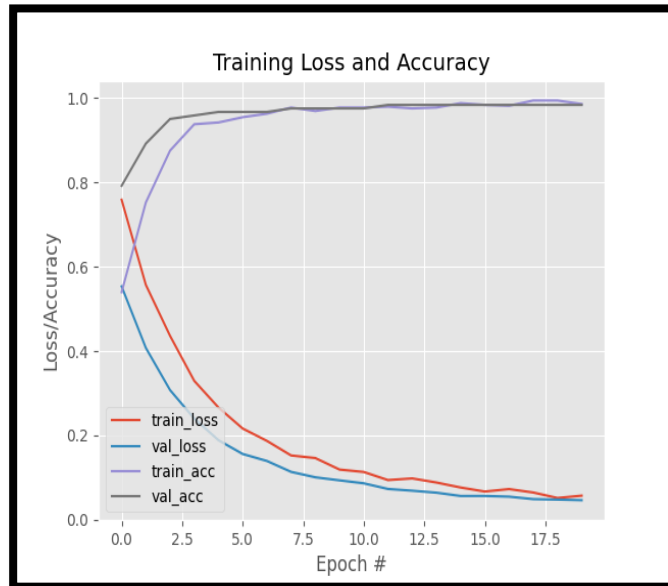


Fig 8: First Model Training results

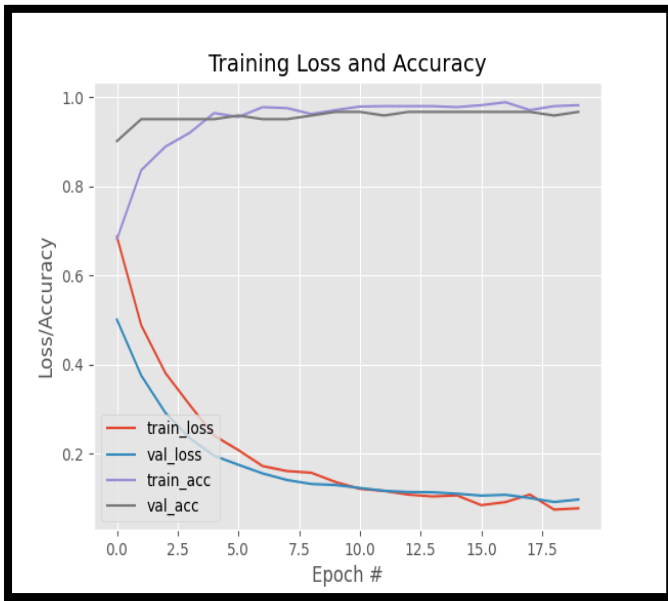


Fig 9: Second Model Training Results

XI. RESULT AND TESTING

Image Testing Results

Test case 1 has an image consists of a single non-masked face which is clearly visible and devoid of any kind of obstruction. As expected, the result is “No Mask: 99.99%” as shown in figure 10.

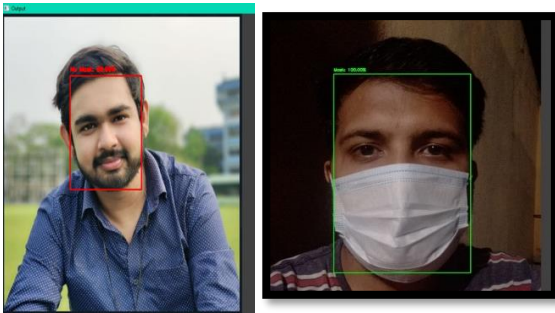


Fig 10: Test Case 1 output Fig 11: Test Case 2 output

Test case 2 has an image which consists of a single masked face. The facial region is devoid of any other kind of obstruction. As expected, the result is “Masked: 100.00%” as shown in figure 11.

Test case 3 has an image which consists of a single non-masked face, but the facial region is obstructed using a towel of some sort for testing the model. As expected, the result is “No mask: 90.46%” as shown in figure 12.

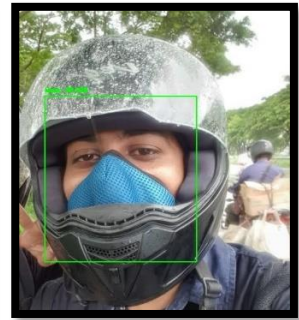


Fig 12: Test Case 3 output Fig 13: Test Case 4 output
Test case 4 has an image which consists of a single masked face along with another obstruction in the form of a helmet which is also partially covering the face. As expected, the result is “Masked 99.99%” as shown in figure 13.

Test case 5 has an image which consists of 5 unmasked faces which are covered with Holi colours. Our model detects only 4 faces as one of them is partially visible, thus missing half of the facial traits. Our model almost successfully classifies 3 faces as non-masked but fails to do so in case of one of the faces. This face is detected as masked which is an error as shown in figure 14.



Fig 14: Test Case 3 output Fig 15: Test Case 4 output

Test case 6 has an image which consists of several non-masked and masked faces. Some faces have some other obstructions like spectacles, beard, sunglasses etc. It is seen that; the results are more accurate in case of faces which are in the front. The only faulty result is seen in case of the face which is all the way in the back and has obstruction from sunglasses and beard. Also, the model fails to detect one face in the front as the facial traits aren't visible enough as shown in figure 15.

Video Testing Results

All the results are in the form of screen shots taken from live video stream. Screenshot 1, shown in figure

16, is taken at the beginning of the video stream of test case 1. The subject is not wearing any mask and as expected the result is “No Mask: 90.61%”. As the video progresses the subject wears a mask, and the result is “Mask: 99.99%” as shown in figure 17. Screenshot 2, shown in figure 17, is at the beginning of the video stream of test case 2. This video has two unmasked subjects, and the result is “No Mask” for both the faces.

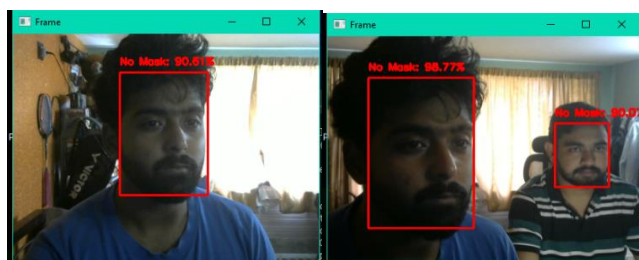


Fig 16: Screenshot 1 output Fig 17: Screenshot 2 output

XII. LIMITATIONS

No algorithm is perfect, and neither is ours. The present Face Mask Classifier model has some limitations when it comes to accuracy. There are a few observed reasons for the inaccurate results, and these are as follows:

XII.1. Facial Obstructions: When a face has some kind of an obstruction other than a mask like a scarf, beard, reflections, etc. the model has difficulty in classifying and the result is shown as masked.

XII.2. Multiple Faces: In case of an image with multiple faces, it is observed that the faces which are in the front have higher chances of having more accurate results than the faces which are in the back.

XIII. FUTURE SCOPE

As mentioned above, the model has some flaws and inaccuracies based on certain criterions. In the future, this model can be made to overcome these inaccuracies if it is trained with a far larger and diverse dataset. Also, several other supplementary features can be further added to the system. There is always room for improvement and this project serves as a pathway for

an extremely efficient and absolute system for face mask detection.

XIV. REFERENCES

- [1]. Shilpa Sethi, Mamta Kathuria, Trilok Kaushik, 2020, “Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread”, “Journal of Biomedical Informatics”, Volume 120, ISSN 1532-0464.
- [2]. M. Inamdar, N. Mehendale, 2020, “Real-Time Face Mask Identification Using Facemasknet Deep Learning Network”, “SSRN Electronic Journal.”, Volume 221, ISSN 134-0156
- [3]. Artem Oppermann, 2017, “What is Deep Learning and How Does It Work?”, Available: <https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac>, Accessed: 12/10/2021
- [4]. Sumit Saha, 2018, “A Comprehensive Guide to Convolutional Neural Networks”, Available: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>, Accessed: 12/10/2021
- [5]. Mohit Varikkuti, 2021, “What is Tensorflow and how does it work?”, Available: <https://towardsai.net/p/l/what-is-tensorflow-and-how-does-it-work>, Accessed: 12/10/2021
- [6]. Bala Venkatesh, 2020, “What is OPENCV and why do we need to know about it?”, Available: <https://www.topcoder.com/thrive/articles/what-is-the-opencv-library-and-why-do-we-need-to-know-about-it>, Accessed: 12/10/2021
- [7]. Martin Heller, 2019, “What is Keras? The deep neural network API explained”, Available: <https://www.infoworld.com/article/3336192/what-is-keras-the-deep-neural-network-api-explained.html>, Accessed: 12/10/2021
- [8]. Jason Brownlee, 2017 “A Gentle Introduction to Transfer Learning for Deep learning”, Available:

- <https://machinelearningmastery.com/transfer-learning-for-deep-learning/>, Accessed: 14/10/2021
- [9]. Sandler et al, 2018, "MobileNetV2: Inverted Residuals and Linear Bottlenecks", Available: <https://paperswithcode.com/paper/mobilenetv2-inverted-residuals-and-linear>, Accessed 14/10/2021
- [10]. Francois Chollet, 2020, "Transfer Learning and Fine Tuning", Available: https://keras.io/guides/transfer_learning/, Accessed: 15/10/2021.
- [11]. Sidath Asiri, "Machine Learning Classifiers", 2018, Available: <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>, Accessed: 15/10/2021
- [12]. Jason Brownlee, 2019, "A Gentle Introduction to the ImageNet Challenge", Available: <https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/>, Accessed: 15/10/2021
- [13]. VijaysinhLendave, 2021, "What is Convolutional Layer?", Available: <https://analyticsindiamag.com/what-is-a-convolutional-layer/>, Accessed: 15/10/2021
- [14]. Jason Brownlee, 2019, "A Gentle Introduction to Pooling Layers for Convolutional Neural Networks", Available: <https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/#:~:text=A%20pooling%20layer%20is%20a,Convolutional%20Layer>, Accessed: 15/10/2021
- [15]. Pooja Mahajan, 2020, "Fully connected vs Convolutional Layer", Available: <https://medium.com/swlh/fully-connected-vs-convolutional-neural-networks-813ca7bc6ee5>, Accessed: 15/10/2021
- [16]. JiwonJeong, 2019, "The Most Intuitive and Easiest Guide for Convolutional Neural Network", Available: <https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480#:~:text=Flattening%20is%20converting%20the%20data,called%20a%20fully%2Dconnected%20layer>, Accessed: 15/10/2021
- [17]. Amar Budhiraja, 2016, "Dropout in (Deep) Machine Learning", Available: <https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5>, Accessed: 16/10/2021
- [18]. Jason Brownlee, 2019, "A Gentle Introduction to the Rectified Linear Unit (ReLU)", Available: <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/>, Accessed: 16/10/2021
- [19]. Kiprono Elijah Koech, 2020, "Softmax Activation Function- How It Actually Works", Available: <https://towardsdatascience.com/softmax-activation-function-how-it-actually-works-d292d335bd78>, Accessed: 16/10/2021
- [20]. Keyur Rathod, 2018, "Face Detection", Available: <https://github.com/keyurr2/face-detection>, Accessed :18/04/2022.

AUTHOR'S PROFILE



Dr. Asoke Nath is working as Associate Professor in the Department of Computer Science, St. Xavier's College (Autonomous), Kolkata. He is engaged in research work in the field of Cryptography and Network Security, Steganography, Green Computing, Big data analytics, Li-Fi Technology, Mathematical modelling of Social Area Networks, MOOCs, Quantum Computing etc. He has published 257 research articles in different Journals and conference proceedings.



Mr. Palash Dutta Banik is a student of St. Xavier's College, currently pursuing M.Sc. in Computer Science. His interests lie in the field of Quantum Computing, Machine Learning, Coding, App Development, Cryptography and Network Security, UI Design, Cyber

Security, AI and real-world project implementation of these fields.



Mr. Sagarmoy Ganguly is currently a student at St. Xavier's College, pursuing his M.Sc. degree in Computer Science. His interests lie in the field of Quantum Computing, Machine Learning, Coding, Cyber Security, AI, Computer Hardware, and real-world

implementation of said fields.



Ms. Ropa Roy is a student of St. Xavier's College, currently pursuing M.Sc. in Computer Science. Her interests lie in the field of Quantum Computing, Machine Learning, Coding, AI, front-end development, and real-world project implementation of these

fields.

Cite this article as :

Palash Dutta Banik, Sagarmoy Ganguly, Ropa Roy, Asoke Nath, "Study on Face Mask Detector System in COVID-19 Era using Deep Learning ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 2, pp. 327-337, March-April 2022. Available at doi : <https://doi.org/10.32628/CSEIT228326>
Journal URL : <https://ijsrcseit.com/CSEIT228326>