

# Face Mask Detection Using CNN

SaiSupriya N<sup>a,1</sup>, Priyanka<sup>b</sup>, RashmiS<sup>b</sup> and Parthasarathy G<sup>b</sup>

<sup>a,1,b</sup>*School of Computing and IT, REVA University, Karnataka, India*

**Abstract:** The COVID-19 pandemic has become an evolving situation all around the globe and the spread is at an alarming rate. Adapting public health-informed hygiene practices can control the transmission of COVID-19, basic measure being wearing a mask. Face detection is the process of identifying faces in a given input image or video and indicating it by drawing a bounding box around the face. In this paper we introduce a deep learning computer vision model to recognise if a person visible through the camera is wearing a mask or not. The deep learning algorithm administered in our work is a transfer learning based Convolutional Neural Network (CNN) and the face detection task is done using the Viola-Jones algorithm approach.

**Keywords:** Face mask detection, Deep learning, Convolutional neural network, Face detection

## 1. Introduction

The COVID-19 pandemic has wreaked havoc all over the world. It has led to colossal loss of human life. The outbreak has had a very adverse impact on the entire economy and human lifestyle. This is caused by a group of viruses called coronaviruses that affect the respiratory tract. Major Symptoms observed are fever, cough, and shortness of breath. The disease is spread when a person touches surfaces contaminated with the virus and touching their face. The trend of wearing face masks in public is rising due to the COVID-19 pandemic across the globe. Scientists proved that wearing face masks helps curb the COVID-19 transmission immensely. People are mandated by laws to wear face masks in public. The main motive of the monitoring process is the detection of anyone who is not wearing a face mask. Face detection is a task concerned with the detection of facial features that appear at different scales, which assists in finding whether there are any faces in a given image approach being the Viola-Jones algorithm based detection using Haar cascade classifiers[4], [7]. Prior to the advent of deep learning based algorithms, feature extraction was performed followed by classification. Some of the various techniques used for feature extraction were SURF, SIFT, local binary patterns, histogram of gradients. Once these features were extracted the classification was performed by any machine learning algorithm like K-means clustering, Principal Component Analysis (PCA) [2],[5], random forests and Support Vector Machines (SVM) [2].

---

<sup>1</sup>Sai Supriya N, School of Computing and IT, REVA University, Bangalore-560064, Karnataka, India.  
E-mail: supriyanagendra12@gmail.com

CNNs extract features directly from images, and these extracted features are learnt while training the network on collected dataset of images and is not pre-trained. The automatic feature extraction aspect of deep learning models makes them highly accurate in computer vision [6]. Deep CNN architecture involves complex models with multiple layers each with an assigned function. They require large image datasets for higher accuracy [1],[2],[3]. The extensive research work done has motivated us to use CNN for the classification task in our work. The proposed CNN architecture capable of detecting masked and unmasked faces can be integrated with pre-installed CCTV cameras. This will aid in tracking safety violations, promote the use of face masks, and ensure a safe working environment.

## 2. Convolutional Neural Network

Convolutional Neural Networks are a class of most widely deep learning pattern recognition algorithms which are designed based on the human brain [2]. CNNs are composed of basic units called neurons. Neurons accept an input and fire an output similar to the neurons in the brain, hence the name. The neuron acts as a placeholder for an activation function, whose main purpose is to introduce non-linearity. Sigmoid, Rectified Linear Unit (ReLU), Tanh etc. are the most frequently used activation functions. The basic layers in CNNs are- Convolutional layer, Pooling layer, and Fully Connected layer [4]. The main feature that makes the CNNs different from other Neural Networks (NNs) is the special mathematical function called Convolution [3], performed in the convolutional layer. This layer consists of a set of filters with parameters which can be visualized as 2D matrices. These filters are slid over the input data across their height and width. The dot product of the values in the filter and corresponding part of the input is computed which outputs an activation map. This is repeated across the depth of the input (i.e., layers) and all the activation maps from each layer are stacked to get the feature map which gives us information about where a specific feature is situated in the input. This map acts as an input to the next layer that is the pooling layer. Pooling layers are used to reduce the spatial information of the feature maps. There are different pooling functions, but max pooling is most widely used. The output from this layer is flattened into a vector and passed to the Fully Connected layer. This layer helps classify the inputs into different categories by training.

## 3. Methodology

The approach we have implemented consists of three basic modules- training the CNN for classification, face detection using the Viola-Jones Algorithm and obtaining the input video stream to carry out the face mask detection task at hand. A flowchart of the methodology is depicted in Fig 1.

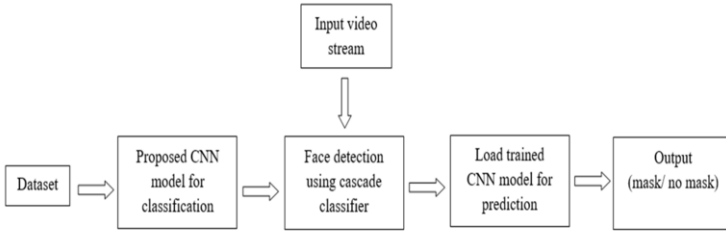


Fig 1: Face Mask Detection flowchart

**Figure 1.**Face mask detective flowchart

### 3.1. The dataset

The first crucial step is to acquire the image data to train our model for classification. The necessary masked and unmasked images for training the classifier were obtained from Kaggle, which is an online community owned by Google Inc. where various data scientists and machine learning enthusiasts interact and share data. This dataset consists of 3828 images. 1915 images belong to the class 'with\_mask' and 1913 images belong to the class 'without\_mask' as shown in Fig 2. The dataset was split into test and train modules using the inbuilt function in the sklearn library.

**Figure 2.**Masked and Unmasked face images

### 3.2. The proposed CNN model

For accurate classification we are using the transfer learning concept where we utilize a pre-trained model as it reduces the computation time and speed the entire process. The pre-trained model we have utilized is MobileNetV2. This model uses depth wise separable convolutions, width multiplier and linear bottlenecks and shortcut connections which was trained on the Imagenet dataset. MobileNetV2 is faster and aids to boost the accuracy of a traditional CNN model. Hence. The pre-trained model was customized, and the architecture of the proposed model is as shown in Fig 3. The base model consists of the MobileNetV2 without the top layers. Images each of (224 x 224 x 3) size are imputed to the model for training. The top layers removed are replaced with the following layers - an average pooling layer with kernel size (7 x 7), a flattening layer, dense layer with 128 neurons with ReLu activation, a dropout layer

with dropout rate 0.5 and lastly another dense layer with 2 neurons and softmax activation. All these layers constitute the head model and are placed on top of the base model set up.

### 3.3. Face Detection

An important task for face mask detection is detecting the faces in the input video stream for which we used the Viola-Jones algorithm for face detection. This algorithm makes use of Haar like features to recognize facial features and thus detects the location of faces[6] which makes it a simple and efficient approach to implement and works efficiently with impressive accuracy. The OpenCV library has pre-defined functions that can implement the face detection by just including the already available predefined haarcascade XML files to perform this task [7]. The face detection is done, and bounding boxes are depicted around the detected faces. The model detects multiple faces visible through the camera simultaneously. The prediction of whether the detected face is masked or not is made using the predict function available in the Keras models package. The bounding boxes appearing are colour coded where a red box with label 'No mask' appears in case the person is not wearing a mask or a green box with label 'Mask' appears when the person is wearing a mask.

### 3.4 Implementation

The proposed computer vision system was implemented using the Keras open-source, software library that is popularly used for implementation of deep learning models. Keras was run with TensorFlow as the backend for fast numerical computations, which makes it highly suitable for building and training this model. Various other python libraries like Numpy, scikit-learn, Matplotlib have been used. The model once built and compiled was trained using training images with a batch size of 32 over 20 epochs for each batch. The learning rate was set to 0.0001 and the loss function used was binary cross entropy. The input, video stream obtained, and face detection was implemented using the OpenCV library.

## 4. Experimental Results

The data considered here for training are a set of images of human faces with and without masks. The model was found to work exceptionally well with the implemented layers, and we could obtain an impressive training accuracy of 99.31%, validation accuracy of 99.35%, training loss of 24.5% and validation loss of about 26%. The following is depicted in the graphs in Fig 3.

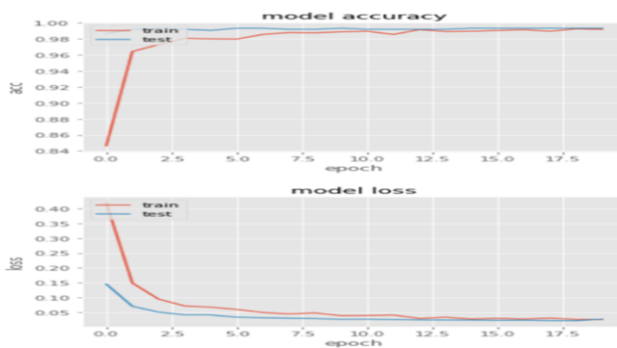


Figure 3. Accuracy and Loss plots

## 5. Conclusion

The COVID-19 pandemic has had a very adverse impact on all aspects of human life, first and foremost being health, economy, lifestyle and so on. Thus, making wearing a mask a crucial task. This brings up the need to overlook if everyone is wearing a mask when in public places. This task can be automated thus serving as an aid to control the spread of the virus. The deep learning model that we have proposed in this paper serves the purpose. This model accepts the live video from the camera as an input and monitors faces detecting whether they are wearing a mask or not.

## References

- [1] M. D. Putro and K. -H. Jo, Fast Face-CPU: A Real-time Fast Face Detector on CPU Using Deep Learning.2020 IEEE 29th International Symposium on Industrial Electronics (ISIE). Delft, Netherlands, 2020, pp. 55-60
- [2] Sanghvi, Kavish and Aralkar, Adwait and Sanghvi, Saurabh and Saha, Ishani .A Survey on Image Classification Techniques .November 25, 2020.
- [3] Dhillon, A., Verma, G.K. Convolutional neural network: a review of models. Methodologies and applications to object detection., ProgArtifIntell 9, 85–112 (2020).
- [4] M. S. Ejaz and M. R. Islam.Masked Face Recognition Using Convolutional Neural Network, 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2019, pp. 1-6.
- [5] S.Hema Kumar, J.UdayKiran, V.D.AKumar, G.Saranya, Ramalakshmi V.Effective Online Medical Appointment System. International Journal of Scientific & Technology Research, Volume 8, Issue 09, September 2019, Pages 803 – 805.
- [6] V.D.Ambeth Kumar, Dr.M.Ramakrishnan, V.D.Ashok Kumar and Dr.S.Malathi (2015) .Performance Improvement using an Automation System for Recognition of Multiple Parametric Features based on Human Footprint.for the International Journal of kuwait journal of science & engineering, Vol 42, No 1 (2015), pp:109-132.
- [7] V.D.AmbethKumar et.al .Facial Recognition System For Suspect Identification Using A Surveillance Camera.Pattern Recognition and Image Analysis (Springer), Volume 28, Issue 3, pp 410–420, 2018. (DOI: 10.1134/S1054661818030136)
- [8] V.D.Ambeth Kumar and M.Ramakrishan."Employment Of Footprint Recognition System", Indian Journal of Computer Science and Engineering (IJCSE) Vol. 3 No.6 Dec 2013.
- [9] Ramya,T.et.al.Personalized authentication procedure for restricted web service access in mobile phones. Applications of Digital Information and Web Technologies (ICADIWT).2014, Page(s):69 - 74, Bangalore, India (ISBN:978-1-4799-2258-1).