

Automatic Real-Time Medical Mask Detection Using Deep Learning to Fight COVID-19

Mohammad Khalid Imam Rahmani¹, Fahmina Taranum², Reshma Nikhat³, Md. Rashid Farooqi³ and Mohammed Arshad Khan^{4,*}

¹College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia

²Computer Science and Engineering Department, Muffakham Jah College of Engineering and Technology, Osmania University, Hyderabad, Telangana, India

³Department of Commerce and Business Management, Maulana Azad National Urdu University, Hyderabad, Telangana, India

⁴Department of Accountancy, College of Administrative and Financial Sciences, Saudi Electronic University, Riyadh, Saudi Arabia

*Corresponding Author: Mohammed Arshad Khan. Email: m.akhan@seu.edu.sa

Received: 24 July 2021; Accepted: 11 October 2021

Abstract: The COVID-19 pandemic is a virus that has disastrous effects on human lives globally; still spreading like wildfire causing huge losses to humanity and economies. There is a need to follow few constraints like social distancing norms, personal hygiene, and masking up to effectively control the virus spread. The proposal is to detect the face frame and confirm the faces are properly covered with masks. By applying the concepts of Deep learning, the results obtained for mask detection are found to be effective. The system is trained using 4500 images to accurately judge and justify its accuracy. The aim is to develop an algorithm to automatically detect a mask, but the approach does not facilitate the percentage of improper usage. Accuracy levels are as low as 50% if the mask is improperly covered and an alert is raised for improper placement. It can be used at traffic places and social gatherings for the prevention of virus transmission. It works by first locating the region of interest by creating a frame boundary, then facial points are picked up to detect and concentrate on specific features. The training on the input images is performed using different epochs until the artificial face mask detection dataset is created. The system is implemented using TensorFlow with OpenCV and Python using a Jupyter Notebook simulation environment. The training dataset used is collected from a set of diverse open-source datasets with filtered images available at Kaggle Medical Mask Dataset by Mikolaj Witkowski, Kera, and Prajna Bhandary. To simulate MobilNetV2 classifier is used to load and pre-process the image dataset for building a fully connected head. The objective is to assess the accuracy of the identification, measuring the efficiency and effectiveness of algorithms for precision, recall, and F1 score.

Keywords: COVID-19 pandemic; convolution neural network; deep learning; face detection; facemask detection; global health; machine learning



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

The COVID-19 pandemic has become a major threat to both global health and the economy. There were 168,040,871 confirmed cases of COVID-19 worldwide, including 3,494,758 deaths, as reported to WHO on 27 May 2021 [1]. The full participation of the population along with the Governments is required in fighting the war against this pandemic. Apart from maintaining physical distancing and washing hands frequently, the proper use of facemasks has now emerged as one of the pillars for preventing community transfer of the disease. The purpose of the mask is either to protect the transmission of the virus or to protect oneself from getting infected by the virus. A model to detect small objects, very effective in these verge submissions in surveillance platforms with good quality cameras especially in service industries was proposed by Roy et al. [2]. Chen et al. [3] have worked on stereo imagery 3D objects using a Markov random field in combination with CNN. Kai et al. [4] proposed two models using SEIR and ABM. Monte Carlo simulation was proposed for forecasting the effect of using face masks. A cogent impact at universal masking is obtained when at least 80% of the public with masks is compared with nominal influence (when only 50% or less face-covered people). Tian et al. [5] have proposed a 3D recognition of patterns on aerial images, used to know the variation, type of the novel illness to plan for accurate prevention. Tian et al. in [5] took an Epidemiological model with a concentration on communication for symptom inception and standardize increase rate through the initial outbreak of the virus for statistical analysis and justifications. Arif et al. [6] have explored and tried combining different parameters for affecting accuracy and validation for the statistical analysis.

Facemasks should be used as part of a comprehensive approaches as proposed by authors in [2–6]. Other measures to eradicate the virus includes physical distancing, avoiding crowd, proper ventilation, cleaning or sanitizing hands, taking steam to kill the virus in the initial stages before it enters the lungs, covering the face while sneezing, coughing, and maintaining hygiene [1,7]. During periodic infection outbreaks in Japan, it was observed that the use of facemasks by children in the age group of 9–12 years were effective [8]. WHO and UNICEF came up with a suggestion that facemasks act as a medium of protection and do no harm at places where physical distancing cannot be ensured for community transmission of SARS-CoV-2. Children in the age group of 6–12 years with austere perceptive or respiratory injuries facing difficulty in using facemask or for children having fear of facemasks are suggested for a face shield to protect from splashes of biological fluid viz. respiratory secretions, organic agents, and wreckage into the eyes. Children under 5 years of age may not have the desired skill to manage a mask; hence it may be avoided for this group of children. Howard et al. [9] explored the standard epidemiological measure of spread using a mask. Furthermore, the role of the government in putting constraints on to use of these masks for stopping the transmission of the disease is also emphasized. Fahmina et al. [10] have justified the role of machine learning and deep learning approaches to deal with today's realistic problems. Rao et al. [11] have proposed that the regulating authorities should take adequate action to control the spread of COVID-19 through proper monitoring and penalizing the offenders. Integrating with apps for identifying the erring persons and their address for further investigation or imposing penalty. Singhal et al. [12] say that viruses will not disappear soon as their different variants keep on emerging. In this scenario modified anatomical face mask (M-AFM) serves as an equally effective alternative to the N95 respirator, HME+ by filters are disposable and have a filtration efficiency of 99.99% [13,14]. Germany suggested a 40% reduction in the daily growth rate of COVID-19 cases with masks [15].

2 Motivation

COVID-19 is an extremely communicable virus. Moreover, its variants keep on changing which makes it difficult to control the virus and treat the patients. Before understanding the importance of masks to counter the spread of this viral infection it is important to know the transmission dynamics of COVID-19 infection.

This is transmitted via aerosols or droplets formed out of speaking, coughing, or sneezing of infected people. A healthy person may be infected with the virus if s(he) comes within short-range (typically 1 meter) to an infected person and the aerosols or droplets containing the virus are taken in through inhaling or come into contact directly with the mouth, nose, or eyes. The prevention of the spread of the virus by use of masks can be understood in the following two steps:

- a) Wearing masks by an infected person: The aerosols or droplets containing the virus will not be allowed to get into the air and infecting others.
- b) Wearing masks by a healthy person: Even if some aerosols or droplets containing the virus are in the air (from an infected person), the mask will prevent inhaling those aerosols or droplets and hence protect a person from getting infected.

In the case of COVID-19 infection, sometimes no apparent symptoms appear even if the person is infected with the virus which is known as asymptomatic infection. Therefore, the policy of wearing a mask must be enforced by the national or local administration. However, a healthy person may get infected with the virus from other means of infection like coming in contact of the aerosols or droplets containing the virus directly and then touching her/his mouth, nose, or eyes etc. Researchers are working to understand the need for facemask detection during today's pandemic situation with the spur to understand the significant effects of using facemasks. It will help the police department to easily locate the people who are violating the rules made for putting on the facemasks. The idea explored is to design a strategy to automatically detect a person not wearing a facemask, to make the manual detection work of face masks easier and its wearing can be ensured in social gatherings automatically. Facemask detection by the camera may be introduced in public places like malls, parks, and exhibitions to abide by the rule imposed by the government to control the spread of the virus.

3 Literature Survey

Masking is a non-pharmacological intervention and a randomized trial in Denmark has shown at least a 50% reduction in the risk for people using surgical masks as per Brooks et al. [16]. Description of the benefits of wearing a mask is in Tab. 1. The authors of the reference [17] discuss the experimental analysis on using a mask to justify the reduction of virus spread. Mask reduces the production of virus-laden droplets from a pre-symptomatic or asymptomatic patient, that estimates to spread more than 50% of the disease. Multi-layered fabric masks usage helps to block the discharge of exhaled respiratory enzymes or microorganisms in the air. Fabric mask helps to reduce 50% to 70% of virus spread [18].

With human experimentation, it is revealed that more than 80% of the blockage of respiratory droplets is with fabric masks when compared to surgical masks. Author Ueki et al. in [19] provided technical knowledge and healthcare policies to the poor, that enable them to escape from infectious diseases. The figures of causalities show that lack of access to basic facilities will increase the risk of communicable diseases like COVID-19 and others [20]. Reference [21] by Chavda et al. proposed a two-stage mask detector architecture. Dlib and MTCNN are used to compare the performance in the first stage, and NASNet Mobile-based model is used for categorizing classes as masked and unmasked for localizing and recognizing people's identity in the second stage. Mahin et al. [22] have researched the usage of the machine learning approach to test the validity of accuracy of the results generated from the simulation by importing these values to multiple classifiers techniques. The authors have created their dataset from the Qualnet simulator and have verified the results generated from the simulator using machine learning techniques. Xiong et al. [23] have worked on the concept of machine learning and have shown its importance to validate the self-generated dataset from multiple scenarios, in converses the effective use of the information to assist in abnormal object detection based on the Mask R-CNN approach. The aim was to achieve the initial instance of the segmentation model through traditional R-CNN and extract the overlapping ratio of the results. Models' results are combined to detect and verify the actual logistics of monitoring images.

Table 1: Effects of covering face with a mask versus infection risk

Source	Reviewed	Outcome
Hendrix et al.	Hair Salon at Missouri, out of 139 customers with masks, only 2 cases of infection were found.	Abided with the mask-wearing norm and no infections among 67 people made available for follow-ups.
Payne et al.	Roosevelt, Guam	The infection was reduced by 70% for 382 US navy service members.
Wang Y et al.	Household Beijing China	The risk of secondary infection was reduced to 79% for 124 households of diagnosed cases comprising 335 people.
Gallaway et al.	Arizona	Temporal association between an institution of state population shows a decline in new cases for those strictly wearing masks.
Rader et al.	United States	A web-based survey on a crowd of 374021 shows that a 10% increase in wearing masks would appropriately stop transmission three folds.
Lyu and Wehby	Washington DC	An estimated overall initial daily decline of 0.9% to 2% was reported after following mask mandates for about 21 days.
Karaivanov et al.	Canada	An estimated weekly decline of 0.25% to 40% was reported after following mask obligations.
Case Study	Thailand	During investigations on contact tracing among a crowd of 1000 masked people exposed to a high risk of diffusion, it is noted that the probability of acquiring infection is less than 30%.

Novel Coronavirus (2019-nCoV) is a disease identified as the cause of an eruption of breathing illness as explained by Hui et al. in [24]. Reference [25] by Lin et al. have used a G-Mask strategy for segmentation and detection of faces. The process adopted in the strategy includes features extraction using ResNet-101, generating ROIs by RPN, and then using ROIAlign to fix or preserve the precise spatial location and finally generating binary masks using a fully convolutional network. By following this approach, the face is detected with better accuracy by precisely segmenting each image. The results were tested with available and generated datasets and it has been interpreted that G-Mask produces better performance and highlighted limitations are expected to increase the speed of simulation. Khan et al. in [26] have surveyed about experiences of learners in online classes during the COVID-19 outburst and analyzed the satisfaction level of online classes. Loey et al. in [27] have presented a unique model for covered face recognition, focusing on medical mask objects to prevent COVID-19 from transmission. For image detection, the YOLO v2 based ResNet-50 model was used to yield high-performance results. Ge et al. in [28] have worked on MAFA and LLE-CNN's datasets for mask detection using 30-occlusions, 811-Internet images, and 35,806-masked faces. With these datasets, the approach has remarkably outperformed giving a novelty of 15.5% to the pre-existing architectures like R-CNN and MT-CNN. Mahin et al. have extended a paper [29] that is a case study to discuss the classification techniques of machine learning language and its applicability to justify the results on different testbeds for various classifiers concerning multiple performance metrics. Uma Dulhare et al. in [30] have proposed a framework of distributed Linear Algebra in Apache Software Foundation to demonstrate an installation,

execution of multiple algorithms of clustering and classification on different environments. Dulhare et al. in [31] have presented a proposal to highlight the purpose of capturing improper images and opting for image fusion to overcome these defects. Dulhare et al. in [32] have given a proposal about the integration of Taj Shenwi images using a series of partially focused images from the same scene. Vinitha et al. in [33] have proposed an approach of deep learning framework to monitor social distancing using object tracking and detection approaches. The experiments were conducted on a group of people using the spatial location of a camera. The aim was to check the closeness among group members using Bird's eye view strategy. Bottou et al. in [34] have proposed an approach to train Multilayer Neural Networks for easy processing using backpropagation. Gradient-based learning strategy assisted to manufacture a multifaceted decision surface that could effortlessly categorize high dimensional difficult patterns with negligible pre-processing. Researchers in their proposal elucidate that the spontaneous pattern identification role of CNN helps to invariably improve the overall recognition of the system to upgrade its performance. Yuan in [35] argued that Occlusion clues to missed discoveries for precision in detecting a visual field. The problem is simplified into a high-level meaningful component for improving analytics of network, location, and scale of the face, through scaling techniques. These analytical features are predicted by the activation map setting to avoid additional parameter locales. Simulation results show that the proposed model acts better in terms of precision of occlusion face detection and recognition. The entire world is stressed with multiple inventions but cannot eradicate the virus. The effect of the second wave has created a disaster because of insufficient infrastructure and scarcity of resources. Even though the newly invented vaccines as defined in [36] are new hope but have experienced an acute shortage in supply. The available dataset in [37] and [38] has daily based information on the cases, viz. several affected cases, deaths, and recovery rates. Dataset is organized in a period of timestamp sequential template and hence the interpretation of cases is cumulative in number. Data from Kaggle can be applied to train machine learning or deep learning approaches. The attributes used in the dataset are low in counts as it is based on realistic data and the pandemic is new. Johns Hopkins University has made an excellent dataset [1] repository available for academia and investigators using the affected instance for research. Kaggle dataset for COVID is designed in 2020 can be used to train machine learning or deep learning approaches. The attributes used in the dataset are less in number as it is based on realistic data and the pandemic is not that old [38]. Various websites with datasets and their related information for the virus are listed out and hyperlinked. The websites [39–44] that help to extract information related to the pandemic and its relevant datasets can be found.

4 The Proposed Model

In this section, a brief description is explained about the tools and other components used along with the description of the proposed work.

4.1 Dataset

Many Datasets are available that can be used to identify the purpose of the study like Kaggle and Kera. The Kaggle has a COVID and Face Detection Dataset. Kaggle COVID dataset consists of almost 2,80,000 values with attributes of cities, the status of recovered, and unrecovered people concerning different places of the world. Kaggle Face detection dataset contains 853 plus images for discriminating people with masks, unmasked and not properly masked as its classes. It contains images of people wearing medical masks and XML files containing their descriptions, annotation, and masks as defined in [38] and [39]. Kera in [37] is an open-source sequential API with a Python interface for the implementation of Artificial Neural Networks. More artificial mask dataset was collected from 'Prajna Bhandary' using PyImageSearch. The dataset includes 1,376 images separated into two classes of 690 pictures with masks and 686 pictures without masks. The artificial dataset created by Prajna

Bhandary has worked with standard face images using facial landmarks technique as explained in Section 3.2.

4.2 Models Used for Deployment and Training

4.2.1 Caffe Model

Caffe (Convolutional Architecture for Fast Feature Embedding) is a deep learning framework library that helps to build a learning system with a sophisticated set of multi-layered configurations. It is well suited for machine vision learning tasks like CNN. Caffe vectorizes input data representation to an array for synchronization and quick data analysis. Caffe creates a repository of the deep net called a Caffe model or deployment. prototxt file. Deployment. prototxt contains the evidence concerning the input size, model definition and is easily compatible with any input image. The prototxt file is created for multiple configuration layers for the input parameter data size for filters, kernel_size, strides (to select the moves), padding, batch normalization with Mean or Variance, elementwise sum, average pooling, dilation, bias filler, or weight filler. The deployment. prototxt file is used only to deploy the model and the Caffe model is used for training to generate a binary *.caffemodel file. Using the pre-trained weights from a *.caffemodel file, Caffe is instantiated with the Net object using the model definition available in the repository of deployment. prototxt file.

4.2.2 Convolution Neural Network (CNN)

A Convolutional Neural Network is a model designed exclusively for Deep learning. It works on an input image, disperses weights to various facets in the image for differentiating and to compare its output image. The three architectural ideas used in Convolutional networks are a degree of shift, scale, and distortion invariance as defined by Bottou et al. in [34]. The three layers of CNN used for filtering are: The first layer is used for edge filtering; the second layer is used for filtering corners and contours; the third layer is used for differentiating the part of the object and then finally the object is identified. The corrections are made through a backpropagation technique to optimize tuning. If the Model does not respond correctly, then modifications are made to the weights of its neurons. CNN is declared as under-fitted if it performs badly on training data and good on testing. On comparing the classification algorithms, the pre-processing requisite is less in CNN. Primitive approaches are hard engineered with good training to facilitate the filters for generating correct results with closed interpretations. The architecture of a CNN is similar to that of the configuration connections of Neurons placement in the human brain and it tries to mimic the learning behavior. The mathematical model of CNN is shown in Fig. 1.

Mathematical logic is applied to the process for generating output for input. The approach used is defined as follows: CNN layers help to process and recognize the image. Object images are fed as numbers in a matrix form, which represents the intensity of pixels in the entity. Mathematical operations are performed by the neurons in the hidden layer on these numbers. The result is sent to the output layer for generating the predictions and further tested with the actual value. The image (q) is transformed using the filter matrix (p) using the formula (1) to generate a feature map matrix with the indices i and j representing the row and the column respectively. The 1D array is generated using formula (1) is further passed to a fully connected layer.

$$C[i, j] = (p*q)[i, j] = \sum_a \sum_b q[a, b]p[i - a, j - b] \quad (1)$$

The Rectified linear unit (ReLU) is used to process small amounts of data by firing neurons, which uses an atom activation feature. The pooling layer is used to fetch the data like a map-reduce function. The output layer is a fully connected layer used to identify the object i.e., once the feature map is extracted from layer 1, the next step is to pass the results to ReLU layer 2, which performs element-wise operations by setting all

negative values to zero and introducing non-linearity to the network for producing a rectified feature map as its output. This output is passed to the pooling layer to do down-sampling and to reduce the dimensions of the feature map. This resultant is passed for flattening to reduce the 2D array to a single long continuous linear vector at layer 3. The flattened resultant is fed to a fully connected layer to classify an object as its successful completion of recognition. Specific neuron responds to stimuli within a constrained expanse of the visual arena referred to as the Receptive Field. Collection of receptive fields are used to cover the visual area. Little to no accuracy with an average precision score is noted for the prediction of classes for binary images as explained in [13]. The architecture performs an improved fitting as the selected weight is characterized on minimum parameters to pattern recognize an image.

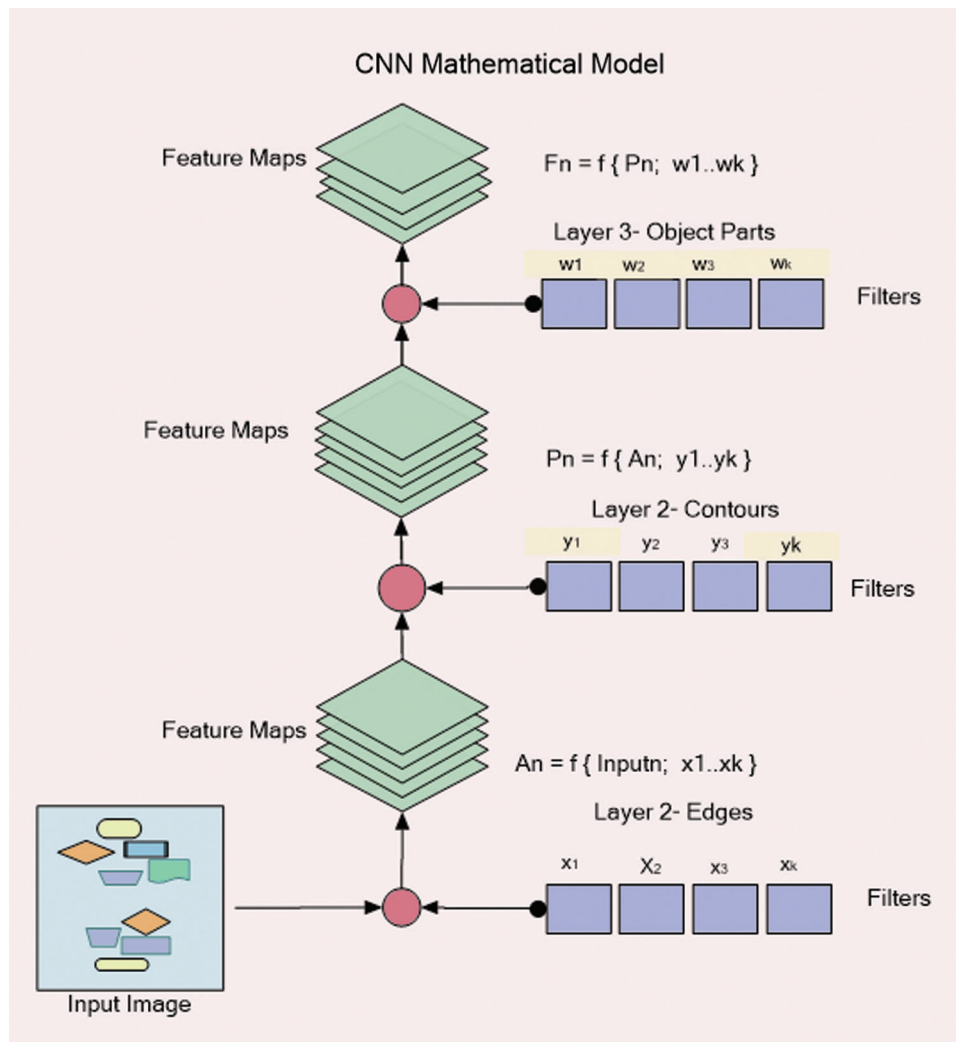


Figure 1: Mathematical model of CNN

CNN helps in reducing the complexity of the images to a simple form, thereby making easy processing with no data loss that results in good prediction. The multiple epochs train the system in such a way that the model predicts the input by using the learned patterns but not from the beginning as defined in [18]. This helps the model to generate more accurate results in less time as concluded by the author in [26]. Based on the training data CNN automatically extracts features that may be used later for object classification.

The approach trains a model by extracting all the available data before testing an application. The aim of the authors Chen et al. are to represent an image in the form of 0's and 1's and replace all negative values with zeros [3]. Hence it works on the approach of extracting as much as possible and then using the extracted knowledge to train and predict the system to make a decision of its own as explored by authors in papers [18,19] and [38].

4.2.3 Other Components

MobilenetV2 is a family of universally persistent neural networks used for computer vision designed with a network of mobile devices to support proper phases applied to discriminate an image like classification, filtering, and detection. A unique and accurate approach to the classification of an image is done using the MobilenetV2 classifier.

4.3 Issues and Challenges in Existing System

Mask is detected with characteristics like diversified types [11,33]; multiple degrees of obstructions [35] and varied accommodations in [11] and as explored in [40–44], which increase the complexity of identification [28] in the existing systems making it an extremely challenging task. It is more interesting to work on the issues and generate an extraordinary solution with exceptional results by providing proper training. Even on generating a solution for a challenge, there is still a possibility of high rising scrutiny in the advancements and developments of face detections [13]. Selecting a high-quality camera with a good aspect ratio [31], event analysis, and video surveillance are also considered as challenges. Several reasons were found for the poor achievement of existing face mask detection models as compared to the normal ones. Two of them are:

- i) Lack of suitable datasets with properly masked faces and facial recognition [33,35,39].
- ii) The presence of masks on the face brings a certain kind of noise [19], which further deteriorates the detection process.

4.4 The Approach Used in the Model

The model works on ascertaining a person's identity on a video stream, with a covered face with the assistance of computer vision and deep learning, collaborated with OpenCV, Tensor flow, and MobileNetV2. The approach is to apply face detection with OpenCV to compute the bounding box location of the face in the image by measuring confidence for a greater threshold to filter out weak detection and consider the strong ones. Append the pre-processed object and its label to an array of data, which is encoded to binary using a one-hot encoding technique. Further data augmentation, pooling, flattening, and activation are applied as defined in Fig. 2, to construct the base model for training.

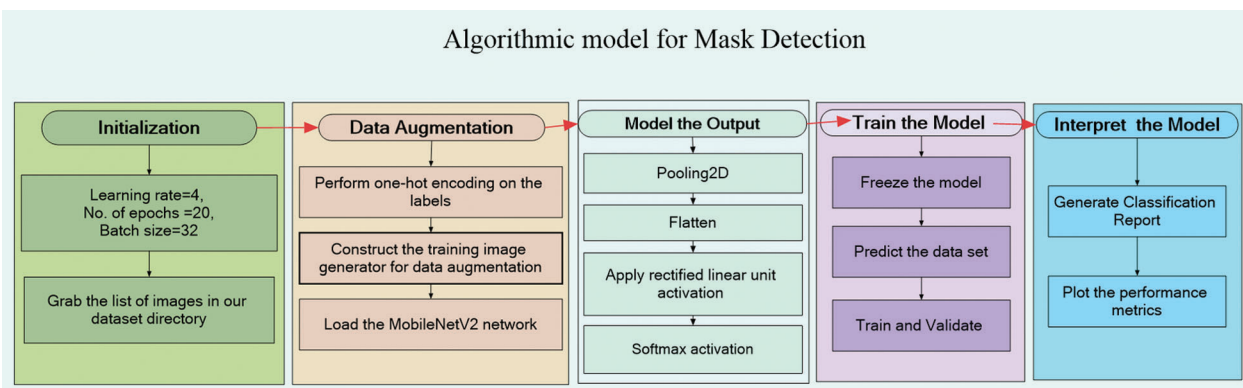


Figure 2: Algorithmic model for mask detection

Some of the techniques used in designing the model include one-hot encoding, Data augmentation, Pooling, Flattening, Activation and Cleaning. Normally, the images are grayscale with the value of pixel ranging from 0 to 255 and 28*28 dimension, which needs to be preprocessed before the feed [45]. Conversion of each float type image with a dimension of 28*28 in a matrix form to a 1D array of dimension 28*28*1 is required before the image is processed using one-hot encoding.

One-hot is an encoding mechanism used to represent categorical variables into a group of binary numbers or vector of numbers with a coset leader of weight one for the representing class, which is used to make a better prediction. One hot encoding uses binary data and works using fit transforms in which only one index bit of an array is hot (have a value 1). It uses a row vector for each image with a 1*10 dimension. As machine learning algorithms understand only non-categorical data a boolean column is required for each category of class. The resultant obtained is fed for image generation. Data augmentation is used to exaggeratedly grow the magnitude of a training dataset using varied versions of images. The attributes used for image generation include rotation range, zoom range, width and height shift range, shear range, and type of flip. Flattening is used to represent data in simple readable form, like converting data to a 1D Array. The model is constructed by using approaches on layers defined in CNN mathematical model as depicted in Fig. 3. MobileNetV2 network is used for fine-tuning and establishing a baseline model in less time for easy interpretation as it has 53 deep layers, which can train millions of images and helps to load a pre-educated version from the ImageNet dataset.



Figure 3: Realistic scenario with a mask

4.5 The Algorithm

4.5.1 Definition of the Algorithm

The algorithm is defined in five phases as depicted in Fig. 2.

Step I: The initialization phase consists of assigning values to learning rate (α), epochs, and batch size and fetching the images into a local directory, α is a hyper-parameter castoff to regulate the rate at which the estimates are updated to learns the values closest to the predictions. The extent of change to the prototype during the individual step of search is 10e-4 with Adam's Optimizer.

Step II: Data Augmentation is used to increase the amount of training data with cropping, flipping, and padding, etc. To initiate Augmentation, the image data generator class is called. The parameters and initialized values to this class include rotation range:=20°C for random rotations, zoom range:=0.15 for random zoom, width_shift_range:=0.2 fraction of total width, height_shift_range:=0.2 of total height, shear range:=0.15°C slant intensity in the anticlockwise direction, horizontal flip:=True: used to flip input

randomly in the horizontal direction, fill mode:="nearest" is used to choose the closest pixel value is repeated it for all the empty values. Other options of fill mode include wrap, reflect, and constant.

Step III: It works by passing fetched data to the pooling layers for reducing the number of parameters, network computation, and spatial size of representation with Backpropagation as it operates independently of the feature map. A pool size of 7*7 is used. The next processor is flattening used to convert pooled collective data to a single column with binary data. Further, it is passed to the fully connected layer for rectified linear and softmax activation with 2 neurons in the output layer. Softmax converts a vector of values to a probability distribution calculated based on confidence, i.e., the boundary of detection is explored if the percentage of confidence is greater than 50%. ReLU is a linear function used to output the input directly if it is positive, otherwise, it returns zero. It is also used to fix the vanishing gradients problem.

Step IV: Step four is about training and validating the model. The encoded binaries are used for testing and training by applying a split function on data, labels, test size, stratify and random state of 42 are used. Training is done using function Aug. flow (train X=data, train Y=percentage of test size, batch_size = 32), testing using model. Predict (test X=label, batch size=BS) and validation_data (testX, testY). The model is simulated over varying test sizes of 20%, 30%, and 50%.

Step V: Generate the report for measuring the accuracy of validity of detection and activation of the image to train the system. Print the classification report generated using test and predictions followed by serializing model to local disk with a file Saving option of ".h5" i.e., (build a serialized dataset, such as HDF5 Hierarchical Data Format 5) and finally plot the graphs to interpret the statistics on a huge dataset.

4.5.2 Working Principle of the Algorithm

CNN accepts input image data as an array and converts it into pixel values by repeatedly convolving and pooling to get the output in vector form. The resultant output is passed to a Feed forward neural network and the Back propagation approach [19] is applied at each iteration of the training phase. Caffe model is used to repeatedly train the images till all the entities of the dataset are completed. After a series of epochs, the model is said to work accurately and can discriminate the dominating and non-dominating features of an image [10]. The configuration of layering in convolution is done in the sequence: Input->BN layer-> Scale ->CL -> RELU activation function layer -> Pool -> Expand ->Eltwise ->fully connected layer.

- a) The input image is converted into a suitable form of size 224 x 224 pixel of Multilevel Perceptron in [13] for image transformation. The input data is passed to layer type: batch normalization. The lr_mult and decay_mult are set to 0 indicating that the parameter Mean, sliding coefficient, and variance of batch normal. The next layer type is Scale used to standardize mean and reduce variance to 1, this helps to avoid gradient explosion and dispersion.
- b) The parameter initialized in the next convolution layer: 64 number of convolution kernels, pad: original image is padded with 3 pixels, kernel matrix is 2*2, stride or slide with two pixels on the input matrix, and for initialization of kernel weight_filler is used.
- c) The parameters configuration is used for deployment and the Caffe model is used for training the convolutional network. A convolution layer num_output is set as 64, it represents the number of convolution kernels used, pad: 3 means that the size of 3 pixels is expanded based on the original input so that the convolution kernel can slide integers on the input matrix views; kernel_size:7 indicates the size of the convolution kernel currently used is 7 × 7; a stride of 2 means a convolution kernel slides two pixels each on the input, kernel matrix is initialized with Gaussian distribution with $\mu=0$ and $\sigma = 2$. These applied setting helps to train RELU in a conducive manner. RELU is used to reduce the effects of gradient disappearance and explosion.

- d) The pooling layer is used to increase the number of convolution kernels thereby removing redundant information and reducing the amount of calculation. `kernel_size: 3` means that the range of each pooling is 3×3 with a step size of 2.
- e) Expansion is performed for `numout = 128` convolution layer and the size of the convolution kernel of the expansion layer is set to 1. Flattening is applied using `mbox_conf_softmax` and `mbox_conf_flatten` along axis 1 and 2 till the image is reshaped using order (0, 1, 2, 3) of permutation. The concept of flattening is as defined in Section 4.6.1 and Fig. 2.
- f) Eltwise layer is used to integrate two or more layers of maps to generate a new one. Example: Add `layer_128_1_conv2` to `layer_128_1_conv_expand` yields `layer_128_1_sum`.
- g) The Softmax activation is further applied to correctly classify the discriminated result as explained in [33]. Softmax parameters are used as the triggering function for multi-class taxonomy problems where class affiliation is applied on the two-class labels {Masked, Un-Masked}. These encoded labeled variables are then sent for one-hot encoding, which gives the probabilistic softmax output representation.
- h) The activation in configuration generates one value for each node in the output layer used for detection. The output values represent the probabilities that sum to one. The parameter used to detect output are set as {No. of classes: 2, `share_location`: enable, `non_max_supression` threshold is set to 0.45, and confidence thresholds is set to 0.5}.

4.6 Performance Metrics

To perform the analysis of the model, metrics used to measure the performance include Accuracy, Precision, Recall, and F1 Score using Eq. (2) to (9) as shown in Tab. 2.

Table 2: Relationships of performance metrics with parameters

Metrics Formulae	Eq. No.
$Accuracy = (TP + TN)/(TP + TN + FP + FN)$	(2)
$Precision = TP/(TP + FP)$	(3)
$Recall = TP/(TP + FN)$	(4)
$F1\ Score = (2 * Precision * Recall)/(Precision + Recall)$	(5)
True positive rate (TPR)= $TP/Actual\ YES$	(6)
False positive rate (FPR)= $FP/Actual\ NO$	(7)
True negative rate (TNR)= $TN/Actual\ NO$	(8)
False negative rate (FNR)= $FN/Actual\ YES$	(9)

These metrics used are listed as follows:

- a) **Accuracy:** The supreme instinctive performance measure is accuracy and it is calculated as the ratio of correctly predicted observations to the total number of observations considered [46,47].

- b) **Precision:** It is calculated as the ratio of correctly predicted positive observations to the total number of observations that are predicted positively.
- c) **Recall:** It is the ratio of correct positive predictions observed to the complete observations in the actual class—yes or positive is Recall.
- d) **F1 score:** The weighted average score of Precision and Recall is the F1 Score [46]. This score is based on both false positive and false negative predictions. F1 score is more beneficial than accuracy, especially in the cases when there is a usage of jagged class distribution.

5 Analysis

5.1 Simulation Parameter

The batch size of 32, is initialized to select the pieces of data that will move over CNN. Every single track of the intact training dataset is called an epoch. A total of 20 epochs were used in the training.

5.2 Strategy

After each epoch, the ability of neurons increases towards enhancement of the ability to classify the training images thereby making an upgradation in prediction. After training the CNN model, the test dataset is used to verify object accuracy. The model has a high-level accuracy in detection and is low resource heaviness.

5.2.1 Capturing of Face

OpenCV is used for simulation to detect a face in a live video stream captured either from the dataset or by a webcam. Each video is created from frames of still images. The aim is to perform the face detection for each frame in a video, by grabbing the frame from the threaded video stream and resize it to have a maximum width of defined pixels. Given a grayscale image, the algorithm guides at smaller sub boundaries and attempts to locate a face by observing for precise features in that region. The detection is performed for boundary, face, and mask. A boundary is created using the pixel coordinates (for top-left and bottom-right corner) and thickness of the rectangle.

5.2.2 Face Detection

Face detection aims to recognize a face in an image and highlight it with a bounding box and return its location and landmarks. Human faces are complex to design as it is based on many characteristics like expression, contours, geometry, rotation, scaling, and brightness along with occlusions like sunglasses, mask, and headcovers (cap or scarf). The Feature-based approaches are used for Face Detection by using OpenCV. The edge or contours of the features helps to discriminate an object from a face.

5.2.3 Differentiating a Masked from Unmasked Face

To differentiate between faces with masks and without masks, a classifier is used. Detection of faces in the frame is dependent on the category. Classes are defined with two categories {Masked, Unmasked} and are placed in two separate folders. This path is selected to load all the images and allocate them to a list called a label. Pre-process the image and resize it to 224 x 224 dimensions and then convert the labels through one-hot encoding. Then split according to the training and testing percentage and do the analytics. Then apply data augmentation using cropping, scaling, shearing, rotation, and flipping to train the system.

5.3 Results and Discussions

After training, the results generated are displayed in Figs. 3 to 11. Fig. 3 depicts a scenario of detecting a frame boundary to identify a face with a mask as explained in Section 4.6. The highlighted numeric value is the accuracy of correct recognition. Fig. 4 is a case of detecting a no mask using the concept of algorithm 4.7 and further, the numeric value demonstrates the correctness of its prediction and accuracy.


```

Epoch 8/20
95/95 [=====] - 263s 3s/step - loss: 0.0480 - accuracy: 0.9866 - val_loss: 0.0317 - val_accuracy: 0.9909
Epoch 9/20
95/95 [=====] - 302s 3s/step - loss: 0.0473 - accuracy: 0.9843 - val_loss: 0.0319 - val_accuracy: 0.9883
Epoch 10/20
95/95 [=====] - 282s 3s/step - loss: 0.0377 - accuracy: 0.9889 - val_loss: 0.0333 - val_accuracy: 0.9896
Epoch 11/20
95/95 [=====] - 335s 4s/step - loss: 0.0394 - accuracy: 0.9900 - val_loss: 0.0305 - val_accuracy: 0.9909
Epoch 12/20
95/95 [=====] - 314s 3s/step - loss: 0.0276 - accuracy: 0.9921 - val_loss: 0.0310 - val_accuracy: 0.9896
Epoch 13/20
95/95 [=====] - 266s 3s/step - loss: 0.0274 - accuracy: 0.9951 - val_loss: 0.0338 - val_accuracy: 0.9896
Epoch 14/20
95/95 [=====] - 249s 3s/step - loss: 0.0232 - accuracy: 0.9958 - val_loss: 0.0262 - val_accuracy: 0.9935
Epoch 15/20
95/95 [=====] - 237s 2s/step - loss: 0.0239 - accuracy: 0.9936 - val_loss: 0.0330 - val_accuracy: 0.9909
Epoch 16/20
95/95 [=====] - 231s 2s/step - loss: 0.0263 - accuracy: 0.9914 - val_loss: 0.0243 - val_accuracy: 0.9935
Epoch 17/20
95/95 [=====] - 227s 2s/step - loss: 0.0326 - accuracy: 0.9894 - val_loss: 0.0269 - val_accuracy: 0.9935
Epoch 18/20
95/95 [=====] - 227s 2s/step - loss: 0.0296 - accuracy: 0.9883 - val_loss: 0.0254 - val_accuracy: 0.9935
Epoch 19/20
95/95 [=====] - 225s 2s/step - loss: 0.0258 - accuracy: 0.9920 - val_loss: 0.0262 - val_accuracy: 0.9909
Epoch 20/20
95/95 [=====] - 226s 2s/step - loss: 0.0193 - accuracy: 0.9956 - val_loss: 0.0281 - val_accuracy: 0.9909
[INFO] evaluating network...
      precision    recall  f1-score   support
with_mask      0.98      1.00      0.99       383
without_mask   1.00      0.98      0.99       384

   accuracy
macro avg   0.99      0.99      0.99       767
weighted avg 0.99      0.99      0.99       767

[INFO] saving mask detector model...
(base) PS C:\Mask Detection\CODE\Face-Mask-Detection-master> sasdsdfg
sasdsdfg : The term 'sasdsdfg' is not recognized as the name of a cmdlet, function, script file, or operable program.
Check the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ ~~~~~
+ sasdsdfg
+ ~~~~~
+ CategoryInfo          : ObjectNotFound: (sasdsdfg:String) [], CommandNotFoundException
    
```

Figure 6: Snippet showing the epochs

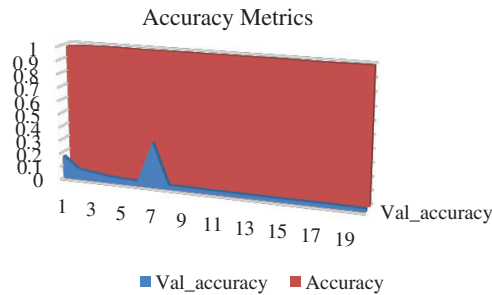


Figure 7: Accuracy vs. val-accuracy

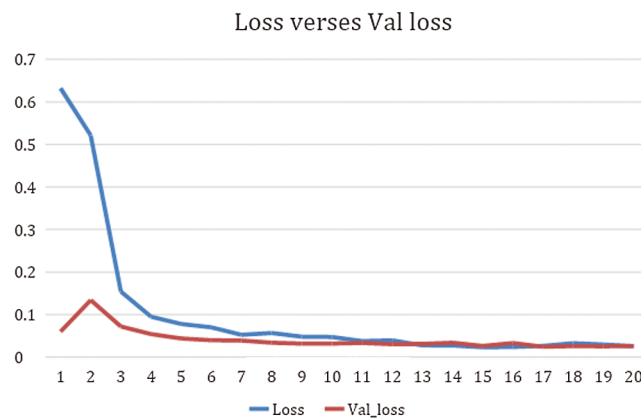


Figure 8: Loss vs. val-loss

Accuracy deals with the reckoning of predictions to ensure fairness of predicted value with true value. It is monitored through epochs in the training phase to check the accuracy of the final model. Accuracy is easier to predict and interpret than loss as it is measured in percentage. Loss is constructed on the probability of predicting uncertainty, which lists the summation of errors made against the training set or validation set. The interpretation in [Tab. 3](#) shows that the accuracy increases with multiple runs, hence some time constraints must be set to retrain the system.

Table 3: Performance analysis post second training

Prediction	Precision	Recall	F1 Score	Support
Mask	.99	.99	.99	383
No Mask	.99	.99	.99	384

The statistics generated using [Eq. \(2\)](#) to [\(9\)](#) are depicted in [Fig. 9](#) in the first run of the epoch.

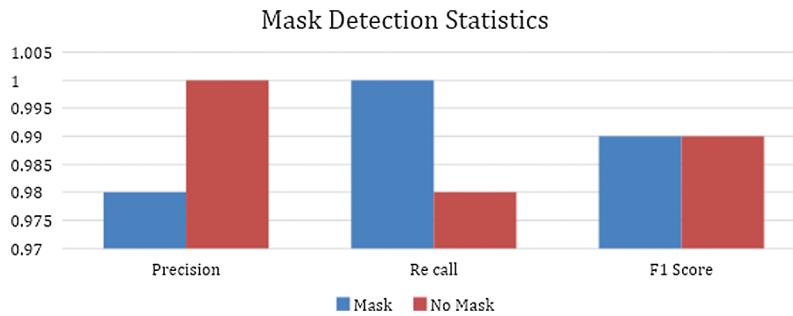


Figure 9: Performance analysis post first training

After training the model for 20 epochs, the captured snippet shows the loss and accuracy generated for prediction on loading an image at different instances of time. When the system is trained with a 50% dataset, the analytics obtained is shown in [Fig. 10](#), stating that the Value-Accuracy and Value Loss generated is zero and accuracy for all epochs are more than loss, which is near to 100%. The results of metrics obtained are shown in [Fig. 11](#) obtained as a snippet from the command prompt.

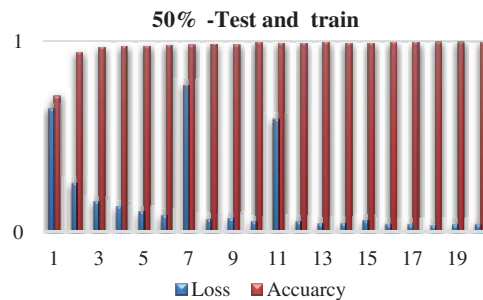


Figure 10: Accuracy on 50% test and train

```
[INFO] evaluating network...
```

	precision	recall	f1-score	support
with_mask	0.98	0.99	0.99	961
without_mask	0.99	0.98	0.99	963
accuracy			0.99	1924
macro avg	0.99	0.99	0.99	1924
weighted avg	0.99	0.99	0.99	1924

Figure 11: Metrics with 50% training

6 Conclusion

The communicable disease including COVID-19 can be prevented to a greater extent by wearing a mask or covering the face, therefore, protecting the users and also the others around. The use of face masks in social gatherings plays a vital role in controlling the spread of the disease. This is a sustainable and affordable remedy. Covering of faces or wearing masks provides a barrier to transmission to and from the users. This is a win-win strategy for all of us. Different studies have proved the efficiency of covering masks with different percentages. However, covering the mouth & nose etc. reduces the ability to communicate. The main objective of the proposed model is to control the transmission of the virus, by imposing a constraint on covering a place from where the transmission instigates. The face mask detection model is implemented for both the training and development of the image dataset, which was divided into categories of people having masks and no masks. The technique of OpenCV used in the model helps to generate accurate results by using training and predicting provided if the mask is covering the mouth and nose area properly, else it gives less accuracy. The idea of Deep learning concepts applied is to use the prediction and validate the results using the training provided. This will remove manual checking of masks and gives a better solution to the problem of mask detection. The proposed model gives 99% accurate results in identifying the captured and loaded images. The major finding of the proposal is the correct prediction of two classes: masked and unmasked. Our proposal has not considered the third class of wearing improper masks.

Considering the third class of improper wearing of masks, the future extension of the proposal is its utilization in the malls to generate a huge and realistic dataset. Further, the system can be trained on different datasets and do comparative analysis over metrics like noise distortions, distance, clarity of image, and improper usage of a mask.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] "Realistic repository to track coronavirus cases," J. Hopkins University, 2021: <https://www.jhu.edu>.
- [2] B. Roy, S. Nandy, D. Ghosh, D. Dutta, P. Biswas *et al.*, "Moxa: A deep learning based unmanned approach for real time monitoring of people wearing medical masks," *Transactions of the Indian National Academy of Engineering*, vol. 5, pp. 509–518, 2020.
- [3] X. Chen, K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma *et al.*, "3D object proposals for accurate object class detection," in *Advances in Neural Information Processing Systems*, vol. 28, pp. 424–432, 2015.
- [4] D. Kai, G. P. Goldstein, A. Morgunov, V. Nangalia and Rotkirch, "Universal masking is urgent in the covid-19 pandemic: SEIR and agent based models, empirical validation, policy recommendations," *Physics and Society, ArXiv Preprint*, arXiv:2004.13553, 2020.

- [5] L. Tian, X. Li, Fei Qi, Q. Y. Tang, V. Tang *et al.*, “Calibrated intervention and containment of the Covid-19 pandemic,” *Physics and Society, Nature Communications*, vol. 1, no. 6, pp. 1-54, 2020. arXiv: 2003.07353.
- [6] F. Arif and U. N. Dulhare, “A machine learning based approach for opinion mining on social network data,” *In Computer Communication, Networking and Internet Security. Lecture Notes in Networks and Systems*, vol. 5. Springer, Singapore. pp. 135–147, 2017.
- [7] L. Peng, “Defense to fight against coronavirus for the people of China,” 2020: http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml.
- [8] WHOMOI, “Child health care and advice on the use of masks to prevent covid,” 2020: <http://www.salute.gov.it/nuovocoronavirus>.
- [9] J. Howard, A. Huangc, L. Zhiyuan, Z. Tufekcie, V. Zdimalf *et al.*, “An evidence review of face masks against covid-19,” *in Proceeding National Academy of Sciences of the USA*, vol. 118, no. 4, pp. 1–12, 2021.
- [10] T. Fahmina, Z. Abid and K. U. R. Khan, “Legitimate path formation for aodv under black hole attack in manets,” *in in Proceeding 4th Int. Conf. of Electronics, Communication and Aerospace Technology*, Coimbatore, India, pp. 1489–1497, 2020.
- [11] S. M. Rao T, S. A. Devi and S. Ram, “A novel approach to detect face mask to control covid using deep learning,” *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 6, pp. 658–668, 2020.
- [12] T. Singhal, “A review of coronavirus disease-2019 (Covid-19),” *National Center for Biotechnology Information, U.S. National Library of Medicine, Journal List, Nature Public Health Emergency Collection*, vol. 84, no. 4, pp. 281–286, 2020.
- [13] R. Panda, P. Kundra, S. Saigal, D. Hirolli and P. Padhihari, “Covid-19 mask: A modified anatomical face mask,” *Indian Journal of Anesthesia*, vol. 64, no. 14, pp. 144–145, 2020.
- [14] S. K. Reddy and M. R. Mathur, “Pandemics & one health: India's evolving response,” *Public Health Foundation of India, Indian Journal of Medical Resessssarch, VIEWPOINT*, vol. 153, no. 3, pp. 245–248, 2021.
- [15] T. Mitze, R. Kosfeld, J. Rode and K. Wälde, “Face masks considerably reduce Covid-19 cases in Germany,” *In Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 51, pp. 32293–32301, 2020.
- [16] J. T. Brooks and J. C. Butler, “Effectiveness of mask wearing to control community spread of sars-cov-2,” *American Medical Association, JAMA Insight*, vol. 325, no. 10, pp. 998–999, 2021.
- [17] “Science brief: Community use of cloth masks to control the spread of sars-cov-2,” Centers for Disease Control and Prevention, 2021: <https://www.cdc.gov/>.
- [18] W. G. Lindsley, F. M. Blachere, B. F. Law, D. H. Beezhold and J. D. Noti, “Efficacy of face masks, neck gaiters and face shields for reducing the expulsion of simulated cough-generated aerosols,” *Aerosol Science and Technology, Taylor and Francis*, vol. 55, no. 4, pp. 449–457, 2021.
- [19] H. Ueki, Y. Furusawa, K. Iwatsuki, Horimoto, M. Imai *et al.*, “Effectiveness of face masks in preventing airborne transmission of sars-cov-2,” *American Society of Microbiology, MSphere*, vol. 5, no. 5, pp. 1–4, 2020.
- [20] M. K. Anser, Z. Yousaf, M. A. Khan, A. Abdelmohsen, Nassani *et al.*, “Does communicable diseases (including covid-19) may increase global poverty risk? a cloud on the horizon,” *Environmental Research*, vol. 187, pp. 1–13, 2020.
- [21] A. Chavda, J. D'souza and A. Damani, “Multi-stage CNN architecture for face mask detection,” *in Proc. of 6th Int. Conf. for Convergence in Technology (I2CT)*, Maharashtra, India, 2021.
- [22] S. H. Mahin, T. Fahmina and K. U. Rahman K., “Detection and interception of black-hole attack with justification using anomaly based intrusion detection system in manets,” *International Journal of Recent Technology and Engineering*, vol. 8, no. 2S11, pp. 2393–2398, 2019.
- [23] H. Xiong, J. Wu, Q. Liu and Y. Cai, “Research on abnormal object detection in specific regions based on mask R-cnn,” *International Journal of Advanced Robotic Systems*, vol. 17, no. 3, pp. 1–9, 2020.
- [24] D. S. Hui, E. I. Azhar, T. A. Madani, F. Ntoumi, R. Kock *et al.*, “The continuing 2019“-Ncov epidemic threat of novel coronaviruses to global health- the latest 2019 novel coronavirus outbreak in wuhan, China,” *International Journal of Infectious Diseases*, vol. 91, pp. 264–266, 2020.

- [25] K. Lin, H. Zhao, J. Lv, C. Li, X. Liu *et al.*, “Face detection and segmentation based on improved mask R-CNN,” *discrete dynamics in nature and society*, Hindawi, vol. 2020, Article ID 9242917, pp. 1–11, 2020.
- [26] M. A. Khan, T. Kamal, A. Illiyan and M. Asif, “School students’ perception and challenges towards online classes during covid-19 pandemic in India: An econometric analysis,” *Journal of Sustainability*, vol. 13, no. 4786, pp. 1–15, 2021.
- [27] M. Loey, M. Gunasekaran, M. Hamed, N. Taha, N. Eldeen *et al.*, “Fighting against covid-19: A novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection,” *Elsevier Public Health Emergency Collection Sustainable Cities Society Elsevier Ltd*, vol. 20, pp. 1–8, 2021.
- [28] G. Shiming Ge, Jia, Y. Qiting and L. Zhao, “Detecting masked face in the wild with Llcnn,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Corpus*, Honolulu, HI, USA, pp. 2682–2691, 2015.
- [29] S. H. Mahin, N. Reshma and T. Fahmina, “Intrusion detection system using machine learning,” *wiley machine learning and Big data: Concepts, tools and applications*, Chapter, vol. 16, Scrivener Publication LLC, pp. 455–486, 2020.
- [30] U. N. Dulhare and H. Sheikh, “Hands on Mahout learning tool,” *In Machine Learning and Big Data: Concepts, Algorithms, Tools and Applications*, John Wiley & Sons, Sebastopol, CA 95472 USA, pp. 363–422, 2020.
- [31] U. N. Dulhare, A. M. Khaled and M. H. Ali, “A review on diversified mechanisms for multi focus image fusion,” in *Proc. of Int. Conf. on Communication and Information Processing*, 2019.
- [32] U. N. Dulhare and A. Khaleed, “Taj-shanvi framework for image fusion using guided filters,” in *Data Management, Analytics and Innovation, Proc. of ICDMAI*, Springer, Singapore, vol. 2, pp. 419–427, 2019.
- [33] V. A. Valentina, “Covid-19 face mask detection with deep learning and computer vision,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 8, pp. 3127–3132, 2020.
- [34] Y. L. L. Bottou, Y. Bengio, and Patrick, “Gradient based learning applied to document recognition,” in *Proc. of the IEEE*, vo. 88, no. 11, pp. 2278–2324, 1998.
- [35] Z. Yuan, “Face detection and recognition based on visual attention mechanism guidance model in unrestricted posture,” *Hindawi Scientific Programming*, vol. 2020, Article ID 8861987, pp. 1–10, 2020.
- [36] “Global analysis by doctors on the coronavirus disease,” BNO-Glo, 2020: <http://3g.dxy.cn/newh5/view/pneumonia>.
- [37] J. Brownlee, “Understanding the approaches of machine learning,” 2020: <https://machinelearningmastery.com/>.
- [38] “Repository to detect people with mask,” Kaggle-Dataset, 2020: <https://www.kaggle.com/andrewmvd/face-mask-detection>.
- [39] “Mastering deep learning using Opencv for face detection,” Kera-Dataset, 2021: <https://www.pyimagesearch.com>.
- [40] “Approaches to model biological systems,” “MOBs-Lab,” 2020: https://www.mobs-lab.org/2019_ncov.html.
- [41] “Specialized agents to monitor world's public health,” WHOPH, 2021: <https://www.who.int/>.
- [42] “Worldwide tracking of coronavirus,” BNO, 2021: <https://bnonews.com/index.php>.
- [43] “Tracking the epidemic in China,” 2021: <http://weekly.chinacdc.cn/news>.
- [44] “Update on covid-19,” Singapore’s Ministry of Health, 2021: <https://www.moh.gov.sg/covid-19>
- [45] M. K. I. Rahmani, N. Pal and K. Arora, “Clustering of image data using k-means and fuzzy k-means,” *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 7, pp. 160–163, 2014.
- [46] M. Tahir and A. Idris, “Md-lbp: An efficient computational model for protein subcellular localization from hela cell lines using svm,” in *Current Bioinformatics*, vol. 15, no. 3, pp. 204–211, 2020.
- [47] M. Tahir, F. Khan, M. K. I. Rahmani and V. T. Hoang, “Discrimination of Golgi proteins through efficient exploitation of hybrid feature spaces coupled with smote and ensemble of support vector machine,” *IEEE Access*, vol. 8, pp. 206028–206038, 2020.