








# A Hybrid Deep Learning System to Detect Face-Mask and Monitor Social Distance

Lutfun Nahar<sup>1</sup>, Nanziba Basnin<sup>1</sup>, Sirazum Nadia Hoque<sup>1</sup>,  
Farzana Tasnim<sup>1</sup>, Mohammad Shahadat Hossain<sup>2</sup>,  
and Karl Andersson<sup>3</sup>

<sup>1</sup> International Islamic University, Chittagong, Bangladesh  
farzanatasnim34@gmail.com

<sup>2</sup> University of Chittagong, Chittagong, Bangladesh

<sup>3</sup> Lulea University of Technology, Skelleftea, Sweden

**Abstract.** Coronavirus Disease 2019 (COVID-19) emerged towards the end of 2019, and it is still causing havoc on the lives and businesses of millions of people in 2022. As the globe recovers from the epidemic and intends to return to normalcy, there is a spike of anxiety among those who expect to resume their everyday routines in person. The biggest difficulty is that no effective therapeutics have yet been reported. According to the World Health Organization (WHO), wearing a face mask and keeping a social distance of at least 2 m can limit viral transmission from person to person. In this paper, a deep learning-based hybrid system for face mask identification and social distance monitoring is developed. In the OpenCV environment, MobileNetV2 is utilized to identify face masks, while YoLoV3 is used for social distance monitoring. The proposed system achieved an accuracy of 0.99.

**Keywords:** COVID-19 · MobileNetV2 · Convolutional Neural Network (CNN) · YoLoV3 · Deep learning

## 1 Introduction

SARS-CoV-2 is a highly infectious respiratory illness caused by the virus. COVID-19 was originally identified in China in December 2019 [2]. In China animals exhibited COVID-19 from where it proliferated globally corresponding to a pandemic [4]. When an infected individual coughs or sneezes, saliva droplets or discharge from the nose are the main ways the virus is disseminated. In those who are afflicted, Covid 19 mainly affects the lungs, and in extreme cases, it can lead to ARDS and pneumonia, which can be fatal. 80% of the time, relatively moderate symptoms will be present, while 14% will develop pneumonia, 5% will experience septic shock and organ failure (most commonly respiratory failure), and 2% will experience fatalities. The main signs of Covid 19 infection include fever, light headedness, dyspnea, headaches, dry coughs that eventually produce

phlegm, and in rare cases, loss of taste and smell. Diarrhoea and weariness have also been mentioned in a few occasions. The most important rule for preventing the spread is to maintain social distance and wear a mask when outside. Avoiding crowded places and huge gatherings of people helps to limit the transmission of COVID-19. Furthermore, wearing a mask keeps the coronavirus from spreading. According to the Centers for Disease Control and Prevention (CDC), masks are one of our most useful tactics. However, it is challenging in crowded areas and marginal communities (i.e. among Bangladeshi garment worker communities) to maintain social distancing and facemasks. As a result, technological intervention for controlling the outspread of the infection is necessary [3]. In this research, a deep learning approach is taken into account to detect facemask as well as measure the distance between two people. In order to, build the face-mask detection model MoblieNet V2 is employed in training the dataset collected from Kaggle. This dataset is further testes using testing dataset as well as real-time dataset. To monitor social distancing YOLOV3-tiny is used because of its high speed processing ability of video image datasets. To distinguish the performance of the proposed models, it is compared with CNN, CNN-LSTM and other CNN state-of-the-art modelsutilized in previous studies.

## 2 Literature Review

In [7] three significant changes has been made to the Viola-Jones (VJ) framework used for object detection. Multi-dimensional (SURF) features, logistic regression and AUC are the changes added to the prior VJ framework. As a result, the convergence rate is faster than that in the previous model. However, this solution cannot beat OpenCV detectors which work in real-time.

In [8] a face mask detection model combining both deep learning methods and classical machine learning is proposed. A CNN pre-trained model ResNet50 is used for feature extraction while decision tree and Support Vector Machine (SVM) are used for the purpose of classification. Inspite of the higher accuracy achieved through machine learning models, the computational speed was slower.

In [5], a deep learning model which uses VGG-16 is proposed for detection of facial emotion and recognition. Using VGG-16 an accuracy of 88% is achieved. However, VGG16 is more than 533MB due to its depth and quantity of completely connected nodes, as a result, deployment of VGG-16 is a challenging task.

[9] uses a SSDMNV2 technique, which is very light and can even be utilized in embedded devices (like the NVIDIA Jetson Nano and Raspberry Pi) to conduct real-time mask detection, employs the Single Shot Multibox Detector as a face detector and MobilenetV2 architecture as a framework for the classifier. The accuracy derived from the method is 0.9264.

In [12] implements the model on a Raspberry Pi 4 to monitor activities and detect violations using camera. When a breach is discovered, the raspberry pi4 notifies the state police control center and the general public of the situation. However, installing the set up can be expensive, moreover the processing power of raspberry pi 4 is much slower.

[10] proposed a system called Covid Vision to assist individuals in relying less on employees while adhering to COVID-19 standards and constraints. Covid Vision is developed using convolutional neural networks (CNNs) for a face mask detector, a social distance tracker, and a face recognition model. The accuracy of the system was found to be 96.49%. The YOLO V3 model performs better than prior models, although good video is crucial.

[11] developed a Face Mask and Social Distancing Detection model as an embedded vision system to help combat the Covid-19 pandemic with the benefits of social distancing and face masks. The pretrained models such as MobileNet, ResNet Classifier, and VGG were employed. Following the implementation and deployment of the models, the chosen one received a confidence score of 100%.

## 3 Methodology

### 3.1 System Architecture

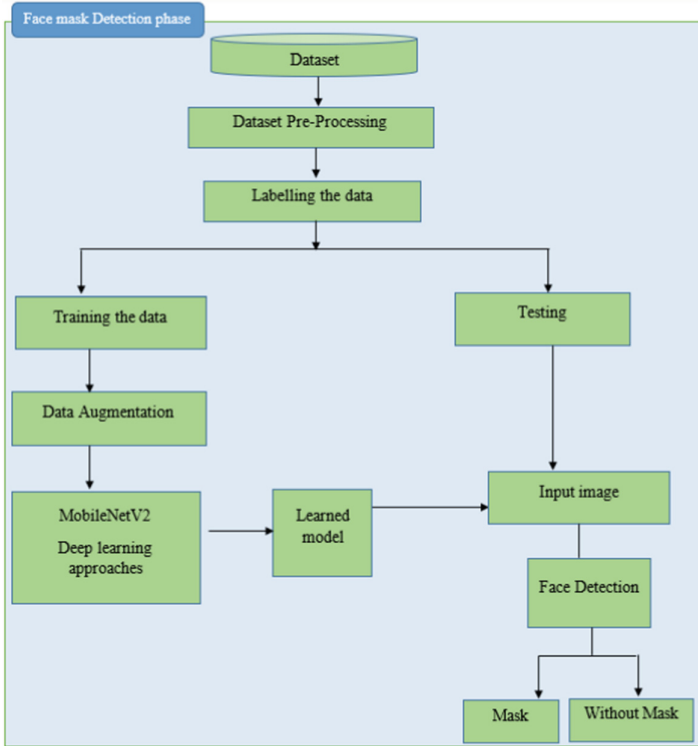
Since this research focuses on building a Hybrid Deep Learning system, two different system architectures are used. Figure 1 demonstrates the system architecture for face-mask detection. Initially, the dataset is pre-processed and labeled into categories namely, “no-mask” and “mask”. The dataset is then divided into two parts: testing and training. The training set is augmented to expand the size of the image set. This training set is then trained using the Mobile Net V2 in order to develop a learning model. This learning model is then tested to classify whether the images are masked or unmasked against the test dataset. Figure 2 depicts the system architecture which is used to recognize social distancing. To achieve this the input video is trained by applying the YOLOV3 model. The distance between centroids is measured to calculate the distance between two people standing side by side.

### 3.2 Data Collection

An open-source platform named kaggle is used to acquire the dataset. This dataset [1].

### 3.3 Data Pre-processing

Grayscale conversion, normalization, resize and data augmentation is performed to pre-process the dataset. Grayscale is merely the conversion of colorful images to black and white. It is commonly used in deep learning techniques to minimize computing complexity. Normalization is the process of projecting image data pixels (intensity) to a preset range (typically (0, 1) or (-1, 1), also known as data re-scaling. This is widely used on many data formats, and you want to normalize them all so that you may apply the same techniques to them. The images are resized into a size of 224 by 224. Afterward, data augmentation is performed in order to make small changes to current data to promote variety



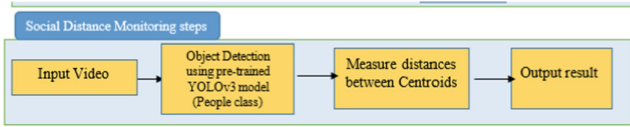
**Fig. 1.** System architecture of face mask detection

without gathering new data. It is a strategy for increasing the size of a dataset. Horizontal and vertical flipping, rotation, cropping, shearing, and other data augmentation techniques are common which are applied to the dataset.

### 3.4 System Architecture

#### 3.4.1 MobileNet Architecture for Face-Mask Detection

Depthwise separable convolutions are an essential component of many efficient neural network architecture . An efficient neural network can be developed by using Depthwise Seperable Convolutions as they are an fruitful replacement for standard convolutions. The basic concept is to replace a factorized version of a complete convolutional operator with one that separates convolution into two different layers. In the first layer, a single convolutional filter, known as a depthwise convolution, is applied to each input channel to perform light filtering.



**Fig. 2.** System architecture of social distancing

The second layer, a  $1 \times 1$  convolution known as a pointwise convolution, creates additional features by computing linear combinations of the input channels. Empirically Depthwise Seperable Convolutions work similarly to those regular convolutions, however, the cost is only  $h_i \cdot w_i \cdot d_i(k^2 + d_j)$  (1) This cost is recognised as the dephwise and  $1 \times 1$  pointwise convolutions sums. In mobile net the value of  $k = 3$  ( $3 \times 3$  depthwise separable convolutions), corresponding to a s 8 to 9 times smaller computational cost and reduction accuracy than that of standard convolutions.

**3.4.2 YoLoV3 System Architecture**

The compact version of YOLOv3 is called YOLOv3-tiny. Compared to YOLOv3, it has fewer layers, making it more deployable on devices with constrained resources. Precision in object detection is sacrificed in favor of processing overhead and inference time reduction due to the fewer layers. YOLOv3-tiny takes in RGB picture with a resolution of  $416 \times 416$  as input. YOLOv3-tiny only predicts bounding boxes at two different scales, in contrast to YOLOv3. The input image is splited into  $13 \times 13$  grids by the first scale, and  $26 \times 26$  grids by the second scale. In each grid, the framework generates three bounding boxes. The network generates a 3D tensor with class predictions, object class confidence, and bounding box information. The network is made up of five different types of layer categories: convoluted layers, routes, maxpooling, upsamples, and YOLO layer. Up sampling is used in route layers, which are responsible for establishing various flows in the network, to provide a variety of detection scales. The Yolo layer is in charge of generating the output vector.

**3.4.3 Measurement of Distance**

The Euclidean distance is used to calculate the distance between two centroids (Figs. 3 and 4).



Fig. 3. Sample of images with and without Face mask

$$E(X, Y) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$d(P, Q) = \min(E(p_1, q_1), E(p_1, q_2), E(p_2, q_1), E(p_2, q_2))$$

where,  $P = [p_1, p_2]$  and  $Q = [q_1, q_2]$  are ground point sets. Every  $D_{ij}$  distance matrix D is computed.

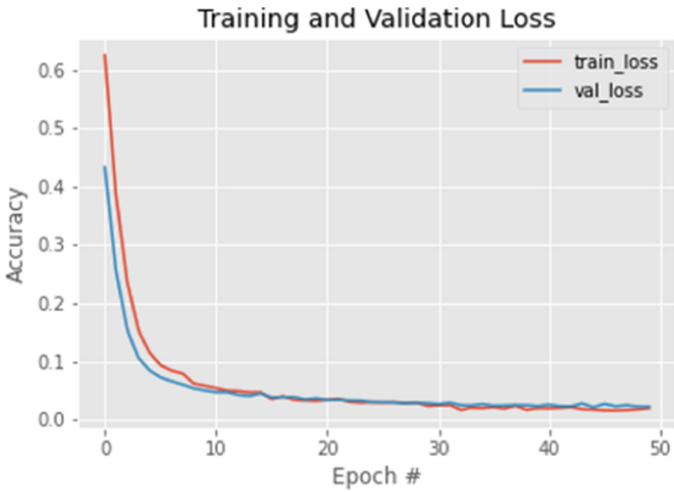


Fig. 4. The flow of data in YOLOV3

### 3.5 System Implementation

The training module is built using Google Collab. The Goggle drive is used as an external storage device in Collab to retrieve files. This platform enables real-time application execution as well as deep learning libraries (Tensor Processing Unit) by offering access to a powerful GPU and TPU. The darknet GitHub repository is used to clone the whole repository to the root address. Some of the libraries utilized to create models in this work are OpenCV, SKlearn, Tensor Keras, and Matplotlib. The model is ready to train after separating the data into training and testing portions of 80% and 20%, respectively.

## 4 Result and Desicussion

### 4.1 Training the Dataset Using MobileNet Model

Figure 5 shows the Training and Validation Accuracy curve of the pre-trained MobileNetV2 model used in the detection of the face-mask. The red line represents the accuracy, while the blue line represents the validation accuracy. There is little to no difference in accuracy between training and testing. Furthermore, it can be deduced that both the line dramatically rises from approximately 5 epochs, after which it remains constant with little change. For evaluation of the performance, there is need some performance metrics.to evaluate the performance for the models MobileNet and YOLOV3, performance matricees namely, precision, recall, F1-score are used [8].

$$\text{Accuracy} = \frac{TP+TN}{((TP+FP)+(TN+FN))}$$

$$\text{Precision} = \frac{TP}{((TP+FP) )}$$

$$\text{Recall} = \frac{TP}{((TP+FN))}$$

$$\text{F1Score} = 2 \frac{(\text{PrecisionRecall})}{(\text{Precision}+\text{Recall})}$$

According to Table 1 it is observed that the Precision, Recall and f1-Score is 0.99.

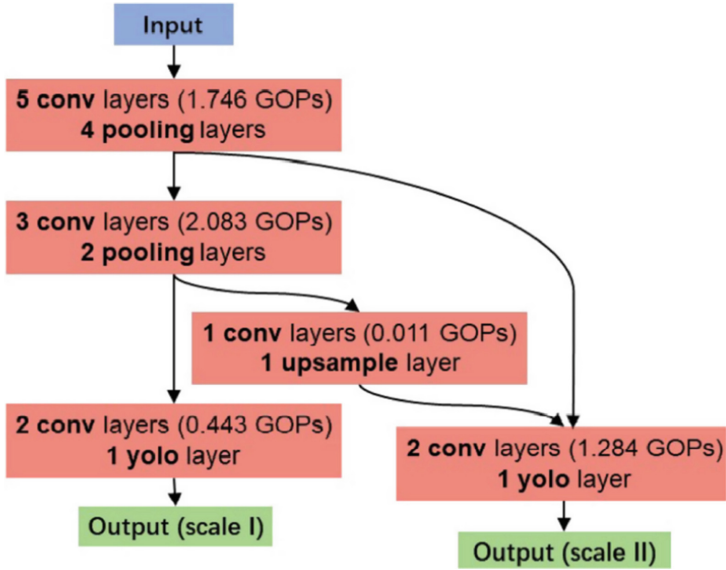


Fig. 5. Training and validation accuracy

Table 1. The proposed model's classification report

	Precision	Recall	F1-score	Support
With mask	0.99	0.99	0.99	383
Without mask	0.99	0.99	0.99	384
Macro avg	0.99	0.99	0.99	767
Weighted avg	0.99	0.99	0.99	767

## 4.2 Social Monitoring System

The YOLO V3 models are used to detect social distance between objects. Figure 6 illustrates not only pedestrians.

## 4.3 Face Mask Recognition in Real Time

The dataset is trained using MobileNetV2. The aforementioned model simultaneously detects multiple individuals in real-time. Figure 8 shows the real time detection generated by the model (Fig. 7).





Fig. 6. Training and validation accuracy

#### 4.4 The Comparison of Performance for Face-Mask Detection

As shown in Table 2, the proposed model is compared to other existing work. The proposed model MobileNetV2 outperforms the models used in other research works. CNN has the lowest performance of any of the models.

Table 2. Comparison of performance for face mask detection

Reference	Method	Accuracy
Li et al. 2013 [7]	Yolo Algorithm	93%
Keniya and Mehendale, 2020 [6]	ResNet50, ResNet18, Yolo	92.8%
Hussain and Al Balushi, 2020 [5]	VGG-16	88%
Proposed method	CNN	20%
Proposed method	MobileNetV2	99%

#### 4.5 The Comparison of Performance for Monitoring Social Distancing

As shown in Table 3, the proposed model outperforms the YOLO-V4 Model by a difference of 9%.

**Table 3.** Comparison of performance for monitoring social distance

Method	Accuracy
Yolo-V4 Model	90%
Proposed Model	99%

**Fig. 7.** Face mask recognition in real time



Fig. 8. Social distance monitoring from real time video

## 5 Conclusion

This research focuses on providing a cost efficient and computationally cheap deep learning based system which will not only detect face mask but also monitor social distancing among people. Here two systems are developed, the face-mask dataset is trained by using MOBILENetV2 model, resulting in an accuracy of

0.99%. Real-Time testing is also carried out, through which the presence of face-mask is detected. Further, YOLOV3-tiny is used to monitor the social distancing among individuals from video footage. The distance among two individuals are measured using the Euclidean distance formula. Further, the Mobilenet is compared with other pre-trained models used in previous studies and CNN models developed solely for the purpose of comparison in this research. It is observed that MobileNet V2 outperforms YOLO algorithm, ResNet 50, ResNet 80, VGG-16 as well as proposed CNN and CNN-LSTM models. In addition, to distinguish the performance of YOLOV3-tiny it is compared to that of YOLOV4. In terms of speed and accuracy YOLOV3-tiny exhibited better performance.

As for future work, this model can be embedded with android application which utilizing mobile cameras. Moreover, we aim to extend our study to include and experiment with people detection by utilizing 3-D dimensions with three parameters (x, y, and z), in which we can feel uniform distribution distance throughout the entire image and do away with the perspective effect.

## References

1. <https://www.kaggle.com/datasets/prithwirajmitra/covid-face-mask-detection-dataset>
2. COVID, C., et al.: Evidence for limited early spread of covid-19 within the united states, january-february 2020. *Morbidity and Mortality Weekly Report*. **69**(22), 680 (2020)
3. Dancer, S.J.: Controlling hospital-acquired infection: focus on the role of the environment and new technologies for decontamination. *Clinical Microbiology Reviews*. **27**(4), 665–690 (2014)
4. Goel, S., Dayal, R.: Novel corona-virus disease 2019 (covid-19): a perilous life-threatening epidemic. *Coronaviruses* **2**(2), 215–222 (2021)
5. Hussain, S.A., Al Balushi, A.S.A.: A real time face emotion classification and recognition using deep learning model. In: *Journal of Physics: Conference Series*. vol. 1432, p. 012087. IOP Publishing (2020)
6. Keniya, R., Mehendale, N.: Real-time social distancing detector using socialdistancingnet-19 deep learning network. SSRN 3669311 (2020)
7. Li, J., Zhang, Y.: Learning surf cascade for fast and accurate object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3468–3475 (2013)
8. Loey, M., Manogaran, G., Taha, M.H.N., Khalifa, N.E.M.: A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic. *Measurement* **167**, 108288 (2021)
9. Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., Hemanth, J.: Ssdmnv2: a real time dnn-based face mask detection system using single shot multibox detector and mobilenetv2. *Sustainable Cities and Society*. **66**, 102692 (2021)
10. Prasad, J., Jain, A., Velho, D., Ks, S.K.: Covid vision: an integrated face mask detector and social distancing tracker. *International Journal of Cognitive Computing Engineering*. **3**, 106–113 (2022)
11. Tebulbi, S., Messaoud, S., Hajjaji, M.A., Mtibaa, A.: Real-time implementation of ai-based face mask detection and social distancing measuring system for covid-19 prevention. *Scientific Program*. **2021** (2021)
12. Yadav, S.: Deep learning based safe social distancing and face mask detection in public areas for covid-19 safety guidelines adherence. *International Journal of Research in Applied Science and Engineering Technology*. **8**(7), 1368–1375 (2020)