An Accurate Real-Time Method for Face Mask Detection using CNN and SVM

Shili Hechmi *

University of Tabuk,
Tabuk 47512, Saudi Arabia
*asuhaili@ut.edu.sa **
** corresponding author*

ARTICLE INFO	ABSTRACT
	Infectious respiratory diseases, including COVID-19, pose a significant challenge to humanity and a potential threat to life due to their severity and rapid spread. Using a surgical mask is among the most significant safety precautions that can help keep this sort of pandemic from spreading, and manual monitoring of large crowds in public places for face masks is problematic. In this research, we suggest a real-time approach for face mask detection. First, we use a multi-scale deep neural network to extract features. As a result, the attributes are better suited for training the detection system. We employ SVM post-processing in the classification stage to make the face mask detection method more robust. According to the experimental findings, our strategy considerably decreased the percentage of false positives and undetected cases. *This is an open-access article under the CC BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/).*

I. Introduction

Respiratory infections have been the leading cause of mortality worldwide for many years. Deaths from pneumonia occur in all countries. Deaths are mainly among the elderly in high-income nations, whereas children are the primary casualties in low-income ones. At the same time, fatalities in both population categories are documented in most middle-income nations. Since the end of 2019, a novel beta-coronavirus has caused several viral pneumonia episodes in China in the Wuhan region [1][2][3] before spreading over the world, resulting in the worst contagious epidemic since the 1918 Spanish flu. This coronavirus, SARS-CoV-2 (Severe Acute Respiratory Syndrome CoronaVirus-2), is responsible for a clinical picture called COVID-19 by the WHO (for COronaVIrus Disease 2019), involving various organ assaults, but most notably an attack on the upper and lower airways. The COVID-19 epidemic has claimed the deaths of almost 5 million people worldwide since the first outbreak [4]. Before the COVID-19 pandemic, over 2.5 million adults and children perished from pneumonia every year: No other virus resulted in as many fatalities. After the emergence of COVID-19, matters became more complicated, and respiratory diseases became a significant challenge to global health, especially with their rapid spread and danger to all segments of society.

Studies on influenza, influenza-like illnesses, and human coronaviruses prove that a medical mask helps lessen the spread of contagious droplets from an infected individual and the possible contamination of the environment with these droplets [5]. These particles are expelled when a COVID-19 patient talks, sneezes, or coughs. People nearby may breathe in these contagious droplets through their mouths and noses and even be inhaled them into the lungs. The WHO highly recommends wearing masks to avoid infection [4]. Many nations have mandated that individuals wear medical masks in public spaces like squares to keep the sickness from spreading.

In contrast, most of the time, the verification process for wearing a medical mask is still done manually, which requires a lot of human capabilities and leads to many errors in identifying people who do not adhere to wearing a mask, especially in crowded places. In response to this need, tools for detecting the wearing of masks - without using facial recognition - are becoming increasingly common. Many recent types of research have been carried out to identify people without a medical mask [6][7][8][9][10], and the challenge remains to get the highest recognition rate.

Detecting individuals not using medical masks has posed a significant challenge to researchers since a person can wear a mask incorrectly and not completely cover the nose and mouth. Below we review the most critical research conducted in this field. Preeti Nagrath et al. [6] introduced SSDMNV2, a deep learning, TensorFlow, Keras, and OpenCV-based technique for recognizing face masks. The Single Shot Multibox Detector detects faces in this approach, and the MobilenetV2 algorithm acts as the classifier's framework. S. Sanjaya and S. Rakhmawan provide a machine-learning method for recognizing face masks in Indonesia, citing [7]. Authorities can plan for COVID-19 mitigation, evaluation, prevention, and response. The suggested model may be used with a security camera to halt the COVID-19 epidemic by recognizing individuals not using medical masks. The authors used the preprocessing step before training and testing the data. The regional result displays the percentage of persons in the cities with the highest and lowest percentages. Gui Ling Wu developed a masked face identification strategy focused on the process of attention [8] to enhance the efficacy of covered face image recognition. The covered face image is extracted first, followed by the face image component, utilizing the local restricted dictionary learning approach. The dilated convolution is then used to compensate for the resolution loss during the subsampling process. Finally, the attentive method neural network is used based on the relevant feature information in the face image to reduce information loss during the sub-sampling process and enhance the rate of facial identification.

Hariri, W. proposed an efficient Method for Masked Face Recognition [2]. The Author employs three pre-trained deep Convolutional Neural Networks (CNN), VGG-16, AlexNet, and ResNet-50 use them to extract deep features from the resulting regions. The Bag-of-features (BoF) is used for in-depth feature extraction and masked faces classification. Multilayer Perceptron (MLP) is applied for the classification process. In [9], G. Yang et al. suggested using a deep learning system based on YOLOV5 to replace manual testing. The technique is applied to look for face masks. The entire system was divided into four sections: facial mask image improvement, facial mask image segmentation, facial mask image identification, and interface interaction. GIOU Loss and Center Loss are combined to identify whether a face mask is used. Anirudh et al. proposed in [11] a face mask detection using image processing. This presented system has three phases: 1. Image preprocessing 2. Face recognition and cropping 3. Classifier for face masks. This technology can identify faces with and without masks and may be utilized with webcam cameras. This method will promote using face masks, help identify safety infractions, and provide a safe working environment. In [10], the authors proposed a new mixed method to automatically detect whether someone is protecting himself by wearing a mask. It combines convolutional neural network (CNN)-extracted visual characteristics with an image histogram that communicates information about the pixel intensity. The authors of this study present a few pre-trained models for creating feature extraction systems utilizing CNNs and several kinds of picture histograms.

In [12], the authors presented a system using convolutional neural networks to classify the detection of facial masks utilizing COVID-19 precautions in images and videos. A complete experiment on the dataset and an effectiveness assessment of the suggested method is presented. In addition, we have succeeded in preserving the inter- and intra-class facial mask detection variability using a symbolic approach. The authors explored different classifiers, including support vector machines and symbolic classifiers. This work is being prototyped to monitor temperature readings and find masks on individuals. The first technique uses a temperature sensor that records your body temperature, and the first approach employs a temperature sensor that measures your body temperature and immediately sprays disinfection. The second, your job is to provide people with safety systems to avoid COVID-19. The author proposed using deep learning concepts to monitor people's conditions continuously. Jiang, Mingjie, and Xinqi Fan introduced RetinaFaceMask [13], a single-stage detector that employs a feature pyramid network to combine very sophisticated semantic data with a novel context attention module focusing on facial mask recognition.

Furthermore, the authors present a novel cross-class object elimination method for denying hypotheses with a high union intersection and little confidence. We have reviewed several works related to masked and unmasked face identification. Through this study, we find a gradual and consistent increase in the accuracy of these systems. Nevertheless, most studies use face databases and not actual images, which does not allow them to be implemented directly into surveillance systems to identify individuals not using medical masks in common areas and in real time. Our method, detailed in the next section, tries to solve this problem.

This research introduces a novel medical mask detection model that uses a CNN to produce multi-scale in-depth features. Automatically, the deep network extracts the original image's multi-layer attributes. Usually, such a network extracts many characteristics that perform better than standard subjective attributes. In response, we used an image dataset to train a CNN-based feature extraction model. In this work, we refer to these derived characteristics as CNN features.

The following is a summary of our paper's primary contributions:
a. We propose a novel method for identifying medical masks that employ multi-scale CNN characteristics collected from layered windows using a deep neural network over several complete connection layers. An SVM classifier is trained using the finding rate from every detection window.
b. We present a complete medical mask detection method that is simple to use, efficient, and reliable.
c. We identify people not wearing medical masks from the actual images and not by using well-framed face images. Our automatic mask detection method can be easily implemented on an existing video surveillance system.

This article is organized as follows: Section 2 summarizes the most recent approaches to detect persons not wearing medical masks. Section 3 contains a complete overview of our suggested technique. The experiment's results are outlined in Section 4. We will conclude this study in Section 5 and suggest possible future steps.

II. Proposed Method

Many essential neurons are used in deep learning, each receiving the output of a lower-level neuron. Based on the nonlinear relation among outputs and inputs, the low-level characteristics are merged into a higher-level abstract concept that describes the scattered properties of the observed data. Bottom-up research is used to create a multi-layer abstract representation. Multi-layer feature learning is a completely automated technique that does not require any human participation. The inputs are transferred to several feature levels via the deep learning technique based on the learned network structure, and then classifies or identifies the top layer's output using a matching algorithm or classifier.

We propose a technique for detecting persons not wearing medical masks based on CNN. We can determine if someone uses a face mask when analyzing surveillance camera photos. First, the human body positioning module receives real-world photos containing multiple possible human body areas. The facial positioning module is then used to identify several potential face areas. The face mask detection module uses this data to identify several face mask detection zones. Finally, using an SVM classifier for post-processing, we determine the accurate face mask identification result. Figure 1 presents a structural representation of the system.
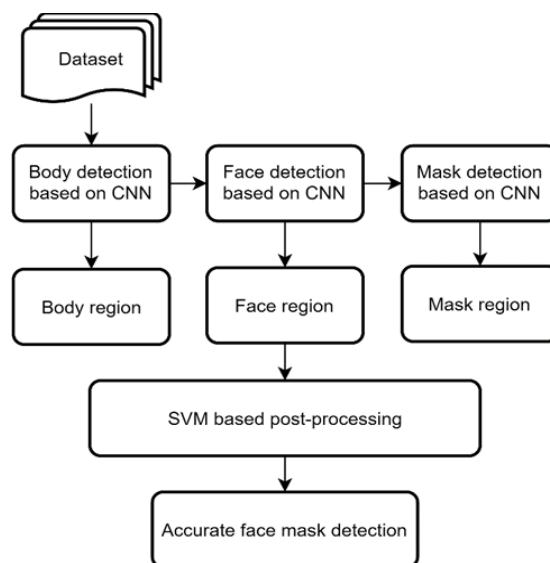

Fig. 1. Face mask identification method using a convolution neural network (CNN)

A. Dataset Description

There are several databases for the evaluation of facial mask detection methods. In order to provide a relevant evaluation of our proposed method, we wanted to use more than one database in the experimental tests but observed that each dataset's annotation format was distinct and that some data did not fit our needs. To solve this problem, we built a database with images collected from available databases with a new annotation and added images from the internet.

This constructed dataset contained 10229 images distributed as follows: 3250 images from WIDER Face [14], 4108 images from MAFA [15], 1521 images from RMFRD [16], and 1350 images obtained from the internet. We utilize 7412 images for training, 772 for verification, and 2045 for testing.

B. CNN Feature Extraction

Consider the *n*-layer structure $S = (S_1, ..., S_n)$. The system's input is *I*, and the output is *O*. It may be written as $I \rightarrow S_1 \rightarrow S_2 \rightarrow ... \rightarrow S_n \rightarrow O$. If the input *I* and the output *O* are equal, O has the same information as the initial input, indicating that no information was lost throughout the layering process (S_i). In other words, *O* is an alternative rendering of the input, whereas *I* represents the initial (the original information). The essential principle of deep learning is that the input and output are equal for each layer of an n-layer neural network. In an ideal world, no human assistance would be required during the learning process.

A CNN is a multi-layered neural network. Every layer consists of numerous two-dimensional planes, each with its own set of neurons. Simple (S-neurons) and complex (C-neurons) neurons comprise the network. The S-neurons create the S-plane, and the S-planes form the S-layer, the symbol we indicate. C-neuron, C-plane, and C-layer (U_c) are all equivalent. The *S* and *C* layers connect each intermediate network level. The input layer, on the other hand, comprises just one layer and direct access to the two-dimensional visual features. The techniques for extracting features from the sample are included in the CNN model's linked structure.

In a CNN, the input connections between S-neuron are flexible, whereas the rest are constant. The output of an S-neuron on the *l* level of the k_l S-plane is represented by $u_{sl}(k_l, n)$, whereas the output of a C-neuron on the k_l C-plane is represented by $u_{cl}(k_l, n)$. *n* is a two-dimensional coordinate representing the field's position in the input layer. The receptive field is limited initially, which rises in level *l*. The S-output neurons are in (1) to (2).

$$u_{sl}(k,n) = r_l(k)\Phi\left\{\frac{1+\sum_{k_l-1}^{k_l-1}\sum_{v \in A_l} a_l(v,k_{l-1},k)u_{cl-1}(k_{cl-1},n+v)}{1+\frac{r_l(k)}{r_l(k)+1}b_l(k)u_{vl}(n)} - 1\right\} \tag{1}$$

$$\Phi(x) = \{x \quad x \geq 0 \quad 0 \quad x < 0 \tag{2}$$

The connection coefficients of the excitatory and inhibitory inputs, respectively, are $a_l(v, k_{l-1}, k)$ and $b_l(k)$. $r_l(k)$ is a constant that regulates the feature extraction's selectivity. A higher number indicates a lower tolerance for noise and feature distinctions. The function $\Phi(x)$ is nonlinear. *v* is a vector that indicates the former neuron's relative location in the receptive field, *n*. The size of the *S* neuron's feature extraction reflecting the receptive field of n, is determined by A_l. As a result, the total of *v* includes all of the neurons in the defined region and the sum of $k_l - 1$ includes all of the sub-planes in the previous level. As a result, the numerator's sum term is frequently referred to as the excited term and is the sum of the product. Neurons are fed into the receptive field, multiplying their outputs by weights. $u_{vl}(n)$ is a supposed inhibitory neuron, *V* in the S-plane that may be used to illustrate the network's inhibitory impact. The V-output neurons are (3).

$$u_{vl}(n) = \left(\sum_{k_l-1}^{k_l-1}\sum_{v \in A_l} c_l(v)(u_{d-1}(k_{l-1},n+v))^2\right)^{\frac{1}{2}} \tag{3}$$

where $c_l(v)$ represents the weights of the V-neurons. The C-output neurons are (4) to (5).

$$u_{cl}(k_l, n) = \varphi \left[\frac{1 + \sum_{k_{l-1}=1}^{k_l} j_l(k_l, k_{l-1}) \sum_{v \in D_l} d_l(v) u_{sl}(k_l, n+v)}{1 + V_{sl}(n)} - 1 \right] \qquad (4)$$

$$\varphi(x) = \{ \frac{x}{\beta + x} \quad x \geq 0 \quad 0 \qquad x < 0 \qquad (5)$$

Where β is a fixed value. k_l is used to indicate how many sub-planes (S) are present at the l level. D_l is the C-receptive neuron's field. As a result, It is related to the size of the feature. The aforementioned fixed link's weight is $d_l(v)$, and it is a monotonically descending function of |v|. If the S neurons k_l is sub-plane got signals from the k_{l-1} sub-plane, $j_l(k_l, k_{l-1})$ *equals* 1, else it equals 0.

C. Multi-scale Detecting Method

The choice of a suitable observation scale is essential for identifying and comprehending targets because the characteristics of an object vary according to the scale. Since images contain objects of various sizes, choosing an ideal scale for picture analysis in advance is not feasible. As a result, the image's content must be considered at several scales.

We created a multi-scale feature extraction model with three CNNs. Each CNN model has eight layers, five of which are convolution layers and three total connection layers. Three layered and increasingly monumental rectangular panes automatically extract features from each image (the mask region, the face region, and the human body region). CNNs extract three features, which are then transmitted to two total connection layers, with the output of the second full connection delivered to the output layer. Finally, the linear SVM classifier is used to categorize all of the sub-blocks.

CNN is used to extract characteristics from each picture. We begin by selecting candidates for the mask region. We can see the mask information immediately in this place. Using the CNN model, we extract features for the face area. This vector is known as Feature A. We will have numerous incorrect detection regions if we identify the mask region by simply extracting its features. A second feature vector is extracted from a rectangle neighborhood to boost the accuracy. This neighborhood comprises the mask region and its immediate adjoining regions, namely the face region. This vector is known as Feature B. We locate the body region utilizing the observed face and mask portions. Feature C refers to the features recovered by the CNN from the human body area.

To train the body identification model, we employ Feature C. B and C is combined to produce W, a new feature utilized to train the face detection model. W is described as:

$$W = vB + \mu C \qquad (6)$$

where v and μ denote the confidences of B and C.

The properties of human observations suggest that different items attract different kinds of attention. For instance, we specify the object region as the target region to recognize an object in an image. The weight of a position decreases the further it is from the target zone. As a result, we put $v = 0.7$ *and* $\mu = 0.3$. We employ a mix of A, B, and C (Feature S) to train the mask identification model. S is defined as:

$$S = \alpha A + \beta B + \gamma C \qquad (7)$$

where $\alpha = 0.6, \beta = 0.3, \gamma = 0.1$ denote A, B, and C's confidences.

D. Post-processing using SVM

We discussed the CNN approach in the previous section for finding the coarse locations of the human body, face, and mask region. To create the feature vector, we compute their detection scores. Finally, we employ the SVM technique for post-processing to delete the inaccurate areas.

In this research, we apply three CNN detection algorithms (D_1, D_2, and D_3) to make three detections. D_1 represents the detection of a human body, D_2 represents the detection of a face, and D_3 represents the detection of a medical mask. *(B: s)* $\in D_i$ is a five-dimensional feature vector with $B = (x_1, y_1; x_2, y_2)$, where (x_1, y_1) denotes the position of the top left corner of the detection box and (x_2, y_2) denotes the position of the bottom right corner. Each part's feature score is represented by s. The dimensions of the candidate image are used to normalize Coordinate B., which satisfies $(x_1, y_1; x_2, y_2)$ $\in [0, 1]$. We create a 15-dimension feature vector *(D_1, D_2, D_3)* for the mask area. We generate a feature

vector in ten dimensions for the face region (D_1, D_2). A linear SVM is then used to classify the feature vectors of the mask, face, and human body regions. The training data are the labeled values (D1, D2, D3) discovered by the CNN algorithm.

E. Performance Evaluation Metrics

The following lists the many measures employed to evaluate the proposed model. The accuracy, precision, and recall equation is in (8) to (10).

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \tag{8}$$

$$Precision = \frac{Tp}{Tp+Fp} \tag{9}$$

$$Recall = \frac{Tp}{Tp+Fn} \tag{10}$$

T_p, T_n, F_p, and F_n denote true positives, true negatives, false positives, and false negatives, respectively. True positives are images appropriately categorized as positive, whereas images mistakenly classified as positive are known as false positives. True negatives are correctly predicted to belong in the negative class, while false negatives are wrongly classified.

III. Results and Discussion

In this part, we run a thorough benchmark of our technique and seven face mask detectors on five well-known datasets. This benchmark's main objective is to ascertain how these face detectors behave while identifying masked faces and in which instances they are likely to succeed or fail. We next undertake an in-depth discussion on how to create face detectors capable of handling faces obscured by various sorts of masks by evaluating the experimental findings of these face mask detectors.

All the experimental tests have been conducted using a laptop running on Windows 10 with the following specifications: AMD Ryzen 7-5700X processor with 32 GB. In this research, the PyCharm program with Python 3.9.12 has been selected for the creation and execution of numerous experimental tests using a variety of libraries, including OpenCV 3.0 and Darknet [17].

The suggested method was tested against various pre-existing models on the same datasets, and the findings are presented in this study. Five contemporary approaches [18][19][20][13][21] were chosen for this purpose. Table 1 compares several models with the suggested one using the accuracy metric.

From Table 1, the first four techniques (YOLO v4, R-CNN, ResNet50, and RetinaFaceMask) were published between 2020 and 2021, and each of them achieved an accuracy between 84% and 89%.

Table 1. Accuracy evaluation of various techniques

Technique	Year	Accuracy	%improvement
YOLO v4 [18]	2020	0.86	+0.08%
R-CNN [19]	2021	0.89	+0.05%
ResNet50 [20]	2021	0.84	+0.10%
RetinaFaceMask [13]	2021	0.88	+0.06%
CenterFace [21]	2022	0.91	+0.03%
CNN and SVM (ours)	2022	0.94	-

Each technique also improved the accuracy of the previous year's best-performing technique by a small percentage, ranging from 0.05% to 0.08%. The fifth technique, CenterFace, was published in 2022 and achieved the highest accuracy of 0.91, with an improvement of 0.03% compared to the previous year's best-performing technique. Finally, the last is our proposed technique, which combines CNN and SVM; the results exhibit the enhanced accuracy of our method over other recent methods. In particular, the suggested model attained a greater accuracy of 94% compared to previous techniques. The accuracy is improved by 3% [21], 5% [19], 6% [13], 8% [18], and 10% [20].

Table 2 compares several models with the suggested one using the precision metric. In Table 2, the precision of selected models has been analyzed. The precision scores range from 87% to 93%, with the CNN and SVM (ours) techniques achieving a precision score of 92%. The % improvement column shows the percentage improvement in precision compared to the previous year's best-performing technique. The results show that our model outperforms [19][20][13][21].

Table 2. Precision evaluation of various techniques

Technique	Year	Precision	%improvement
YOLO v4 [18]	2020	0.88	+0.04%
R-CNN [19]	2021	0.91	+0.01%
ResNet50 [20]	2021	0.87	+0.05%
RetinaFaceMask [13]	2021	0.91	+0.01%
CenterFace [21]	2022	0.93	+0.01%
CNN and SVM (ours)	2022	0.92	-

Table 3 compares several models with the suggested one using the recall metric. Results indicate that our strategy outperforms other recent approaches in terms of recall. Notably, the proposed model achieved a higher recall of 93%. The recall is improved by 1% [21], 2% [13], 3% [19], 5% [20], and 6% [18].

Table 3. Precision evaluation of various techniques

Technique	Year	Recall	%improvement
YOLO v4 [18]	2020	0.87	+0.06%
R-CNN [19]	2021	0.90	+0.03%
ResNet50 [20]	2021	0.88	+0.05%
RetinaFaceMask [13]	2021	0.91	+0.02%
CenterFace [21]	2022	0.92	+0.01%
CNN and SVM (ours)	2022	0.93	-

Overall, based on the information presented in Table 1, Table 2, and Table 3, we can conclude that various techniques for object detection, such as YOLO v4, R-CNN, ResNet50, RetinaFaceMask, CenterFace, and a combination of CNN and SVM (proposed), have been developed and evaluated. Therefore, we can conclude that our proposed CNN and SVM technique has achieved the highest accuracy, precision, and recall scores among the techniques evaluated in this study.

IV. Conclusion

Automatic identification of people not wearing face masks is a significant study issue. Using CNN and SVM, we propose an accurate real-time technique for detecting face masks. CNN enables the extraction of attributes better suited for training the detection model, whereas SVM is utilized for classification. The findings show that our approach significantly outperforms the other recent techniques utilized in the studies. In the future, we will address the issue of incorrectly worn face masks, making the identification system more intelligent.

Declarations

References

[1] C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," Lancet, vol. 395, no. 10223, pp. 497–506, Feb. 2020.

[2] W. Hariri, "Efficient masked face recognition method during the COVID-19 pandemic," Signal, Image Video Process., vol. 16, no. 3, pp. 605–612, Apr. 2022.

[3] N. Zhu et al., "A Novel Coronavirus from Patients with Pneumonia in China, 2019," N. Engl. J. Med., vol. 382, no. 8, pp. 727–733, Feb. 2020.

[4] WHO, "Infection prevention and control during health care when coronavirus disease (COVID-19) is suspected or confirmed," WHO, 2021. (Access on 29 July 2022)

[5] WHO, "Infection prevention and control of epidemic-and pandemic prone acute respiratory infections in health care," WHO, 2014. (Access on 29 July 2022)

[6] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," Sustain. Cities Soc., vol. 66, p. 102692, Mar. 2021.

[7] S. A. Sanjaya and S. Adi Rakhmawan, "Face Mask Detection Using MobileNetV2 in The Era of COVID-19 Pandemic," in 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Oct. 2020, pp. 1–5.

[8] G. Wu, "Masked Face Recognition Algorithm for a Contactless Distribution Cabinet," Math. Probl. Eng., vol. 2021, pp. 1–11, May 2021.

[9] G. Yang et al., "Face Mask Recognition System with YOLOV5 Based on Image Recognition," in 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Dec. 2020, pp. 1398–1404.

[10] E. Ryumina, D. Ryumin, D. Ivanko, and A. Karpov, "A Novel Method for Protective Face Mask Detection using Convolutional Neural Networks and Image Histogram," Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. XLIV-2/W1-, pp. 177–182, Apr. 2021.

[11] K. Anirudh, A. Ravi, V. S. Charan, and V. Chaurasiya, "Face Mask Detection Using Machine Learning," in 2022 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Feb. 2022, pp. 1–5.

[12] G. K. J. Hussain, R. Priya, S. Rajarajeswari, P. Prasanth, and N. Niyazuddeen, "The Face Mask Detection Technology for Image Analysis in the Covid-19 Surveillance System," J. Phys. Conf. Ser., vol. 1916, no. 1, p. 012084, May 2021.

[13] X. Fan and M. Jiang, "RetinaFaceMask: A Single Stage Face Mask Detector for Assisting Control of the COVID-19 Pandemic," Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern., pp. 832–837, 2021.

[14] S. Yang, P. Luo, C. C. Loy, and X. Tang, "Wider Face: A Face Detection Benchmark," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, pp. 5525–5533.

[15] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2682–2690.

[16] B. Huang et al., "Masked Face Recognition Datasets and Validation," in 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Oct. 2021, pp. 1487–1491.

[17] A. Farhadi, "Darknet: Open Source Neural Networks in C." (Access on 29 July 2022)

[18] K. Bhambani, T. Jain, and K. A. Sultanpure, "Real-time Face Mask and Social Distancing Violation Detection System using YOLO," in 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), Oct. 2020, pp. 1–6.

[19] J. Zhang, F. Han, Y. Chun, and W. Chen, "A Novel Detection Framework About Conditions of Wearing Face Mask for Helping Control the Spread of COVID-19," IEEE Access, vol. 9, pp. 42975–42984, 2021.

[20] S. Sethi, M. Kathuria, and T. Kaushik, "A Real-Time Integrated Face Mask Detector to Curtail Spread of Coronavirus," Comput. Model. Eng. Sci., vol. 127, no. 2, pp. 389–409, 2021.

[21] C. W. Yang, T. H. Phung, H. H. Shuai, and W. H. Cheng, "Mask or Non-Mask? Robust Face Mask Detector via Triplet-Consistency Representation Learning," ACM Trans. Multimed. Comput. Commun. Appl., vol. 18, no. 1s, pp. 1–19, 2022.