Journal of Engineering Science and Technology
Special Issue on ICITE2021, October (2021) 1-10
© School of Engineering, Taylor's University

REAL-TIME FACE DETECTORS AND CLASSIFICATION PERFORMANCE IMPLEMENTING RESNET-50 MODEL FOR FACE MASK-WEARING CONDITIONS

ANNA LIZA A. RAMOS*, JHINO S. CERBITO*, SOPHIA MIGUELA B. HADI*, TRICIA ELIS A. BLANCA, KIER MIGUEL O. DELORIA, KYLE SPENCER F. GO

Institute of Computer Studies, Saint Michael's College of Laguna, Philippines,
*Corresponding Author: annaliza.ramos@smcl.edu.ph

Abstract

The World Health Organization requires the community to wear a face mask to avoid transmission of COVID-19. The study investigates the performance of face detectors and evaluates the classification performance based on face mask-wearing conditions. The study built a total of 13,806 datasets that recorded an overall classification performance of 98%. The findings show that Multi-task Cascade Convolutional Neural Networks outperformed the other face detectors with an average score of 70% in accordance to distance, angles, occlusions, and multiple detections across given set conditions. Furthermore, the model recorded an accuracy performance of 83% for "correct wearing of face mask", 91% for "incorrect wearing of face mask", and 95% for "no face mask". However, despite the promising performance rates, the identified best face detector decreases when the given conditions are set to a higher level. To further improve and optimize the face mask-wearing conditions, the study highly recommends employing both statistical and mathematical analysis.

Keywords: Face detection, Face mask, Face mask-wearing conditions, MTCNN, Rest-Net50.

1. Introduction

The World Health Organization requires people to wear a face mask to prevent the spread of transmittable diseases and reduce the risk of infection caused by COVID-19 [1, 2]. Different face masks were used such as surgical masks, cloth masks, N95, and the like [3, 4]. However, the process of monitoring is becoming more difficult because of human limitations. With the advancement of technology, the emergence of face detection schemes and studies provide relevant solutions through focusing on identifying the face landmarks [5], analyzing the facial features [6], human face recognition [7, 8], and predicting both the overall face and landmarks using MTCNN [9, 10]. However, face mask-wearing conditions still encounter new challenges in detecting different orientations, degrees of occlusion, face detectors' performance [11], and the limitations on available datasets that lead to the construction of more types of datasets.

In relation to this, a particular study built its MAFA datasets that provide face mask occluded datasets with different orientations resulting in remarkable results [12]. Another study investigated publicly available datasets by implementing RestNet50 and SVM, which recorded a much higher accuracy [13]. A study examining resolution issues through implementing SRCnet recorded a 98.70% accuracy classification performance [14]. The application of Principal Component Analysis to detect masked and unmasked resulted in 70% recognition accuracy for the faces wearing a mask [15]. A face detection without a mask with alarm features applied in the operating room recorded a rate of 95% accuracy [16]. Moreover, there has also been a study that employed a model that detects social distancing and face masks using computer vision and MobileNet V2 architecture [17]. In comparison, calculating the distance of a person from the camera is more robust and accurate. [18, 19]. A study that tested datasets with 13,359 images - 7,067 with mask and 6,292 without a mask, used R-CNN, SSDMNV2, MobileNetV2 algorithms returned an accuracy performance of 99.96% for without face mask and 98% for the ones with a face mask on [20].

Therefore, this study aims to further investigate the performance of the face detectors currently available that were used in the existing studies by taking a closer look at those studies that used various face detectors with unique results and methods of investigation. This study also constructs its datasets that include various orientations, distances, angles, occlusions, and multiple detections to address the limitations on the types of datasets and to better recommend the top face detectors when it comes to face mask detection. The results of this study will help us learn a lot about how face detectors work and can be used to improve the algorithms that are currently being used to build smart, flexible detectors that can recognize people even when they are wearing face masks.

2. Methodology

Figure 1 shows the conceptual framework for the real-time face detector. It would convert the input image from RGB to grayscale and then use a median filter to reduce noise. The image pyramid method would be used to resize each image and create multiple copies of each. In order to generate a bounding box and five landmark points for each detected face. The images would be fed into Multi-task Cascade Convolutional Neural Networks (MTCNN) with 12x12 input size images, MTCNN operates and outperforms the P-net, producing bounding boxes with lower and higher confidence values using the non-maximum suppression method. The higher confidence value is then fed into the R-net, which generates more precise bounding boxes with a 24×24 input size. These bounding boxes are then fed into the O-net, which has a 48×48 input size. The face mask-wearing conditions would then be categorized using the O-net output and the ResNet-50 object classification model.

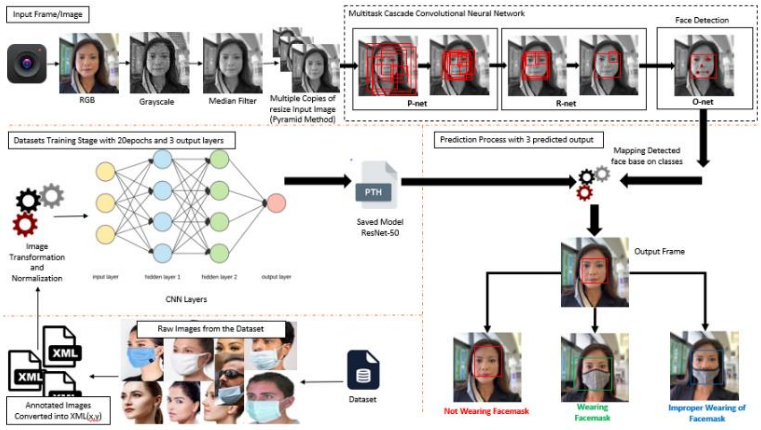

Input Frame/Image
RGB
Grayscale
Median Filter
Multiple Copies of resize Input Image (Pyramid Method)
Multitask Cascade Convolutional Neural Network
P-net
R-net
Face Detection
O-net
Datasets Training Stage with 20epochs and 3 output layers
Image Transformation and Normalization
input layer
hidden layer 1
hidden layer 2
output layer
CNN Layers
PTH
Saved Model ResNet-50
Prediction Process with 3 predicted output
Mapping Detected face base on classes
Output Frame
Raw Images from the Dataset
XML
XML
Annotated Images Converted into XML(x,y)
Dataset
Not Wearing Facemask
Wearing Facemask
Improper Wearing of Facemask

Fig. 1. Conceptual framework.

2.1. Building datasets

Figure 2 shows the process of building the datasets captured in an uncontrolled environment in terms of lighting conditions such as direct sunlight, indirect sunlight, and even darker areas. Likewise, face positions are captured in slightly-sided left and right, or even full-sided left and right, slightly looking upward and downward. Also, foreign objects such as hair, face shields, and sunglasses can cause half-face occlusion. The overall total datasets collected were 13,806 with 8,155 images of "correct wearing of face mask (CWFM)", 4,122 images of "improper wearing of face mask (IWFM)", and 1,529 images of "no face mask (NFM)." These images were saved through an XML file and annotated by dragging the rectangular bounding box to the region of interest and saving them to its XML file.

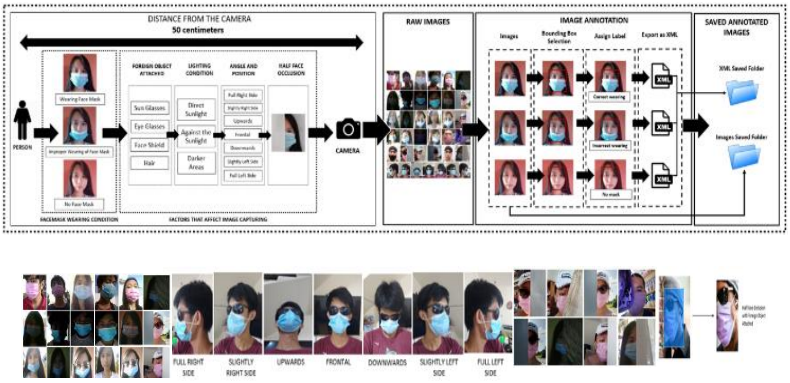


Fig. 2. Datasets collection.

2.2. Image processing/preprocessing stage

Figure 3 shows the processing of the image. Each image has variations in the face dimension depending on the bounding box per image. All images were resized to 224 x 224 to be uniform in size and then converted from RGB to Grayscale to make MTCNN face detection more adaptable to different situations and reduce the computing cost. For the RESNET-50 model, the images were converted from RGB to BGR for training, testing, and validating the images.

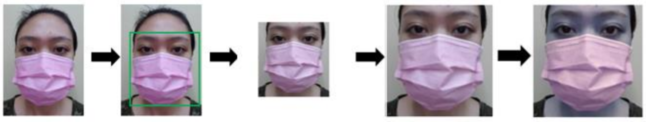


Fig. 3. Image processing process.

2.3. Building model

The transformed data was processed to train the model. The partition of the datasets applied was: (80%) training data, (10%) validation data, and (10%) testing data after the preprocessing stage, summing up to a total of 13,806 datasets. The training and validation data in fully connected layers were used to train and test ResNet-50.

The employed ResNet-50 pre-trained model with fully linked layers for feature extraction and classification generates an efficient classification. The robust feature extractor produced efficient classification in small datasets. The flattened layer with ResNet-50 final layer, dense layers with 1024, and rectified Linear unit activator were utilized as input and output layers for the model. The model's next output was a dense layer with a value of 3 and a softmax activator. Consequently, the study used the Adam optimizer for a higher prediction rate and set the epoch count to 20. The model was saved in the.h5 format.

2.4. Experimental setup

2.4.1. Object classifiers

The applied convolutional neural network object classifier model according to its size, accuracy, and CPU inference step. The model size was narrowed, but the accuracy must be around 90% with enough CPU inference steps for a more efficient training time. ResNet-50, InceptionV3, MobileNetv2, and pure CNN were utilized as the CNN model classifiers chosen from the Keras Application Library that fit the criteria. The experiment examines a model classifier's accuracy to determine the right percentage of detection, recognize the precision-recall and f1-score, and see which efficient attributes to use.

2.4.2. Face detection

The study investigated the performance of face detection algorithms such as the HAAR Cascade, DLIB, FaceNet, and MTCNN to find the best detector that suits the study. Likewise, the distance between the web camera and the person was between 1 and 2 meters in order to figure out how accurate and precise the face detector algorithm was in real time. It also tested each distance at three different angles, which were 0°, 45°, and 90°. While testing the algorithms, two main categories were created: partial and half face occlusions. The partial occlusion, which covers 25% of the face, could occasionally but not entirely detect the face, while the half occlusion, which covers 50% of the face, had difficulties in detecting it. It also included several foreign items to test the algorithms and multiple face detection with various lighting conditions.

2.5. Evaluation procedure

2.5.1. Experimental subject

The subjects of the study were the proponents, with a total of four (4) respondents. For each session, the study set each type of class that needed to be examined in terms of distance, angles, occlusion, and multiple object detection of the face, using a laptop camera with 1280x720 resolution. The study initialized the duration to 60 seconds per session and counted each change for each respondent every second. Each session takes 240 seconds to finish testing for distance, angles, and occlusions. Testing for multiple detections and overall performance takes 180 seconds.

2.5.2. Performance accuracy

The study applied the classification accuracy to the model's real-time performance based on the identified categories in this experiment.

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} * 100 \tag{1}$$

$$Precision = \frac{TP}{(TP+FP)} * 100 \tag{2}$$

3. Result

The study evaluated the performance of the object model classifiers and face detectors based on the following parameters (Table 1):

Table 1. CNN object classifiers results

Object Classifiers		Precision	Recall	F1-score	Accuracy
CNN	Correct Wearing	0.25	0.01	0.01	
	Incorrect Wearing	0.60	0.98	0.75	0.61
	No Face mask	0.79	0.16	0.26	
ResNet-50	Correct Wearing	0.98	0.99	0.98	
	Incorrect Wearing	0.99	0.98	0.98	0.98
	No Face mask	0.96	0.95	0.96	
InceptionV3	Correct Wearing	0.98	0.96	0.95	
	Incorrect Wearing	0.98	0.99	0.99	0.98
	No Face mask	0.94	0.96	0.95	
MobileNetV2	Correct Wearing	0.94	0.97	0.87	
	Incorrect Wearing	0.98	0.95	0.96	0.95
	No Face mask	0.85	0.90	0.87	

Table 1 shows the results of the object classifiers. The CNN-based object classifier acquired a poor accuracy result of 61%, indicating that it has low precision, recall, and true positive prediction rates in each class, while the f1-score indicated that the prediction rates were not balanced and had a relativity deficiency in each class. Next, the MobileNetv2 got an accuracy result of 95%, implying that each class had sufficient precision and recall percentage findings. However, the f1-score shows low relativity in each class because the IWFM class got a 96% f1-score compared to the other related class, which only got 85%. MobileNetv2 could become biased in predicting each class.

The study then trained datasets using InceptionV3, which produced a different accuracy result compared to the first two models. The accuracy result of InceptionV3 got a 98% prediction rate and got a high accuracy result in its precision and recall. The f1-score of InceptionV3 demonstrated the relativity strength prediction accuracy rate in each class.

The ResNet-50 also demonstrated a prediction accuracy of 98%, the same as the rate for InceptionV3. Therefore, the study looked into the precision and recall scores of both models since the datasets were unbalanced. It also checked the performance gaps of each class, which implied that ResNet 50 outperformed InceptionV3. There was significantly enough correlation in each class to allow a modest biased prediction in the actual testing.

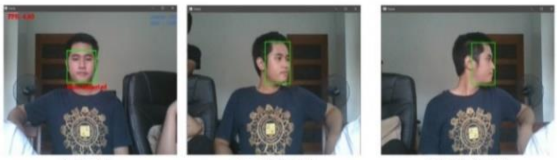
3.1. Face detector analysis

Table 2 shows the result of the model detection in terms of distance, angle occlusions, and multiple face detection. The lowest accuracy in the 1-meter category is HAAR, while the lowest in 2 meters is DLIB. MTCNN outperformed all the tested methods, both in the 1 meter and 2 meter categories. In terms of angle, HAAR and MTCNN face detectors recorded a significant difference of 15% at a 90° angle. FaceNet was close to MTCNN in attaining the highest performance, with a 1% difference in partial and half face occlusions. MTCNN delivered the highest performance in multiple detections. Overall, an investigation of the capability of face detectors was needed by applying algorithms to effectively detect face mask-wearing conditions.

Table 2. Face detectors experimental analysis results.

Analysis		**Haar**		**DLIB**		**FaceNet**		**MTCNN**	
		Pre	Acc	Pre	Acc	Pre	Acc	Pre	Acc
Distance	**1m**	53%	56%	56%	68%	64%	68%	76%	78%
	2m	32%	47%	27%	30%	50%	50%	67%	67%
Angle	**0°**	46%	60%	54%	58%	79%	81%	86%	88%
	45°	37%	44%	11%	25%	55%	61%	78%	86%
	90°	0%	26%	0%	.07%	12%	17%	39%	41%
Partial Occlusion		25%	25%	40%	40%	63%	63%	64%	64%
Half Face Occlusion		20%	20%	29%	29%	50%	50%	50%	52%
Multiple Face Detection		67%		25%		33%		78%	

Figure 4 shows the sample face detector algorithm could easily detect a person's face from different angles with 0°, 45°, and 90° frontal sides. HAAR and DLIB gave lower accuracy results at a specific angle, as shown in Table 2.



0° frontal 45° side frontal 90° side frontal

Fig. 4. Face detection experimental analysis in 0°, 45°, and 90° angles.

3.2. Face mask classification analysis using MTCNN

Figure 5 shows the detected image bounding in the box, labeled to determine the three categories and the five landmarks.

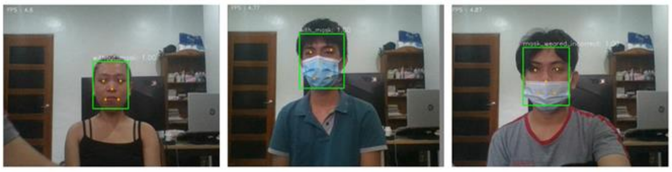


Fig. 5. Face mask classification (a) no face mask, (b) correct wearing of face mask, and (c) incorrect wearing of face mask.

Table 3 shows the overall performance classifications in terms of distance, angle, and occlusion. The performances recorded an average score of 98% in the 1-meter category and 80% in the 2-meter category. The NFM recorded 100% in both precision and accuracy. In angles, 0° recorded the highest classification accuracy of 100% in IWFM and NFM classifications. The model also predicted partial occlusion with the highest score of 99%, compared to 82% for half-face occlusion. The result was acceptable since the face detector performed well when there were no obstructions on the face. Lastly, multiple detections recorded a high result in IWFM and NFM while the CWFM recorded only 70%. Therefore, there’s a need to enhance the datasets’ representation, more samples to improve accuracy, and utilize a high-end camera.

Table 3. ResNet-50 with MTCNN experimental analysis results.

Analysis		Correct Wearing		Incorrect Wearing		No Face Mask	
		Pre	Acc	Pre	Acc	Pre	Acc
Distance	**1m**	100%	96%	95%	98%	100%	100%
	2m	95%	83%	92%	80%	92%	78%
Angle	**0°**	100%	96%	100%	100%	100%	100%
	45°	98%	79%	98%	95%	100%	99%
	90°	94%	73%	96%	84%	96%	97%
Partial Occlusion		98%	89%	97%	90%	97%	99%
Half Face Occlusion		96%	82%	97%	82%	97%	88%
Multiple Face Prediction		91%	70%	98%	99%	97%	97%

Figure 6 shows the model detected facial landmarks in partial face occlusion and generated labels of NFM and IWFM.

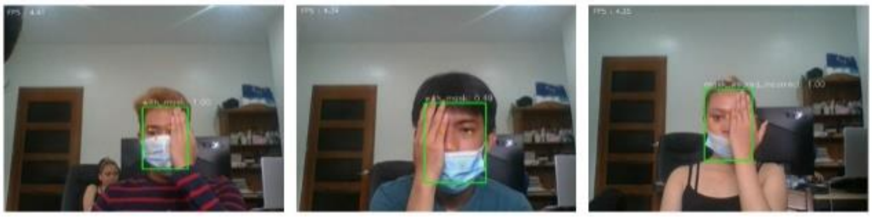


Fig. 6. Occlusion with classification (a) partial occlusion with face mask (b) half face occlusion with incorrect wearing (c) half face occlusion with incorrect wearing.

4. Application

Figure 7 shows the face mask detector prototype application system that allows the user to monitor daily activities in various indoor and outdoor environments. Even under low ambient light and partial occlusion, it can classify face mask-wearing conditions. The system can detect many people at the same time and count those who are not following the proper safety protocols for wearing face masks so that the person in charge knows where and when strict safety precautions are required in that specific region. To use the model, you need to import the model and label files, save the changes, and the system will monitor them in real time.

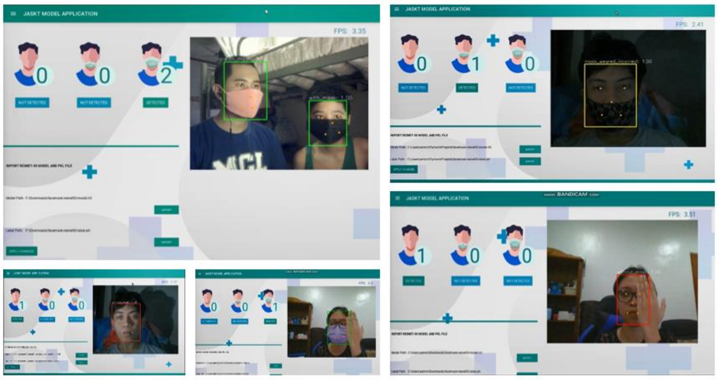


Fig. 7. Face mask detector prototype application system.

5. Conclusion

Based Based on the results, the study concluded that the best-performing face detector is the MTCNN. Among the object classifiers, the ResNet-50 model detector recorded an average classification score of 72.5% for distances of 1 meter and 2 meters, 70% for angles of 0°, 45°, and 90°, 78% for multiple detections, and 58% for partial and half-face occlusions, which means that the available face detectors have a limit in performing detection based on the given conditions. Therefore, mathematical analysis and algorithms should still be employed.

Furthermore, the ResNet-50 model marked a significant classification performance of 98%, which can be good for detecting the identified categories. As a result, the model recorded an average classification performance on the three conditions: a) CWFM scored 88% in distance, 83% in angle, 70% in multiple detections, and 86% in occlusion; b) IWFM scored 89% in distance, 93% in angle, 99% in multiple detections, and 91% in occlusion; and c) NFM scored 89% in distance, 99% in angles, 97% in multiple detections, and 93% in occlusion.

Notably, all these categories show the same pattern; when the condition increases, the performance decreases. Additionally, both the IWFM and NFM conditions recorded higher scores than the CWFM condition even though the latter gained the highest number of datasets with a total dataset of 8,155. And, to further enhance the performance, the study recommends the following:

- Use a high-end camera to improve the performance of the 2-meter distance.

- Employ algorithms or additional mathematical functions to optimize the detection of face occlusion, angles, and increase multiple face detection.
- Consider different landmark possibilities of the face to determine whether a person is correctly or incorrectly wearing a face mask.
- Increase the batch and epoch sizes to improve the model performance.
- Improve the number of datasets.

Abbreviations

CNN	Convulutional Neural Networks
CWFM	Correct Wearing of Fae Mask
FP	False Positive
FN	False Negative
IWFM	Incorrect Wearing of Face Mask
MAFA	Masked Face
MTCNN	Multi-Task Cascaded Convolutional Neural Networks
NFM	No Face Mask
R-CNN	Region Based Convolutional Neural Networks
RESNET-50	Residual Neural Network 50
SSDMNV2	Single Shot Multibox Detector and MobileNetV2
TP	True Positive
TN	True Negative
SVM	Support Vector Machine
WHO	World Health Organization

References

1. Altmann, D.; Douek, D.; and Boyton, R. (2020). What policymakers need to know about COVID-19 protective immunity. *The Lancet*, 395(10236), 1527-1529.
2. World Health Organization (2020). Mask use in the context of COVID-19. Retrieved October 10, 2021, from https://www.who.int.
3. Ringer, J. (2020). Which type of face mask is most effective against COVID-19? *Loma Linda University Health*.Retrieved September 25, 2021, from https://news.illu.edu.
4. Wang, J.; Pan, L.; Tang, S.; Ji, J.; and Shi, X. (2020). Mask use during COVID-19:A risk-adjusted strategy. *Environmental Pollution*, 266(1), 115099.
5. Hassaballah, M.; Bekhet, S.; Rashed, A.; and Zhang, G. (2018). Facial Features Detection and Localization. *Studies in Computational Intelligence*, 804, 33-59.
6. Vikram, K.; and Padmavathi, S. (2017). Facial parts detection using Viola-Jones algorithm. *Proceedings of the 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 1-4.
7. Cuimei, L.; Zhiliang, Q.; Nan, J.; and Jianhua, W. (2017). Human face detection algorithm via Haar cascade classifier combined with three additional classifiers. *Proceedings of the 13th IEEE International Conference on Electronic Measurement & Instruments* (ICEMI), Yangzhou, China, 483-487.
8. Padilla, R.; Costa Filho, C.F.F.; and Costa, M.G.F. (2012). Evaluation of haar cascade classifiers designed for face detection. *World Academy of Science,*

Engineering and Technology International Journal of Computer and Information Engineering, 6(4), 466-469.

9. Joshi, A.S.; Joshi, S.S.; Kanahasabai, G.; Kapil, R.; and Gupta, S. (2020) Deep learning framework to detect face masks from video footage. *Proceedings of the 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, Bhimtal, India, 435-440.
10. Zhang, K.; Zhang, Z.; and Li, Z. (2016). Joint face detection and alignment using multi-task cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499-1503.
11. Johnston, B.; and Chazal, P. (2018). A review of image-based automatic facial landmark identification techniques. *EURASIP Journal on Image and Video Processing*, 2018, Article number: 86.
12. Ge, S.; Li, J.; Ye, Q.; and Luo, Z. (2017) Detecting masked faces in the wild with LLE-CNNs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2682-2690.
13. Loey, M.; Manogaran, G.; Taha, M.H.N.; and Khalifa, N.E.M. (2020). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, *Measuremen*t, 167 (2021), 108288.
14. Qin, B.; and Li, D. (2020). Identifying face mask-wearing condition using image super-resolution with classification network to prevent COVID-19, *Sensors*, 20(18), 5236.
15. Ejaz, M.S.; Islam, M.R. ; Sifatullah, M.; and Sarker, A. (2019) Implementation of principal component analysis on masked and non-masked face recognition. *1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT),* Dhaka, Bangladesh, 1-5.
16. Nieto-Rodríguez, A.; Mucientes, M.; and Brea, V.M. (2015). Mask and maskless face classification system to detect breach protocols in the operating room. *Proceedings of the 9th International Conference on Distributed Smart Cameras*, New York, United States, 207-208.
17. Yadav, S. (2020). Deep learning-based safe social distancing and face mask detection in public areas for COVID-19 safety guidelines adherence. *International Journal for Research in Applied Science and Engineering Technology*, 8 (7), 1368-1375.
18. Serign, M.B.; Fang, M. (2020). An improved face recognition algorithm and its application in attendance management system. *Array*, 5, 100014.
19. Yuan, Z. (2020). Face detection and recognition based on visual attention mechanism guidance model in unrestricted posture. *Scientific Programming Toward a Smart World*, 2020, 8861987.
20. Talahua, J.; Buele, J.; Calvopina, P.; and Aldas, J.V. (2021). Facial recognition system for people with and without face mask in times of the COVID-19 pandemic. *Sustainability*, 13(12), 6900.