ICIC Express Letters
Part B: Applications
Volume **14**, Number **4**, April **2023**

ICIC International ©2023 ISSN 2185-2766
pp. **415–422**

FACE MASK RECOGNITION SYSTEM ON RESTAURANT WITH SELF-PAYMENT SYSTEM USING CONVOLUTIONAL NEURAL NETWORK (CNN)

Ryan Wibawa, Kevin Aprilla Tarto, Jovan Gannon Sukitata
Kristien Margi Suryaningrum and Rezki Yunanda

Computer Science Department
School of Computer Science
Bina Nusantara University
JL. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ ryan.wibawa; kevin.tarto; jovan.sukitata }@binus.ac.id; { kristien.s; rezki.yunanda }@binus.edu

Received July 2022; accepted October 2022

Abstract. *COVID-19 is a disease that affects many aspects of life transmitted by verbal interaction. Nowadays the rapid growth of COVID-19 has become an international issue due to violation of the face mask rules. This research will provide a comparison of the deep learning class, Convolutional Neural Network (CNN) which is used as the basis of the face mask recognition system and to adapt it into a payment verification system. This research will use MobileNetv2 and YOLO-v4 with its pretrained model using group of images composed of person using face mask and person not using face mask. Each model successfully performs the detection task. The result shows that MobileNetv2 has achieved a better overall percentage compared to the YOLO-v4 algorithm. Hence, MobileNetv2 has been chosen as the algorithm used for the payment verification system.*
Keywords: COVID-19, Rapid growth, Deep learning, Convolutional Neural Network (CNN), Face mask recognition, Payment verification system, MobileNetv2, YOLO-v4

1. **Introduction.** In the last months of 2019, COVID-19 was discovered in Wuhan, China, and alerted the world to the risk of the upcoming pandemic [1,2]. Uncontrolled COVID-19 has stirred a worldwide problem with the risk of death spreading globally around the world to 213 countries and territories. The lack of active health professionals and the inability to resist COVID-19 has created vulnerability for humanity. World Health Organization (WHO) has pronounced COVID-19 as a pandemic [3]. COVID-19 causes various symptoms ranging from complications, and respiratory illness, to severe illnesses such as pneumonia, organ failure, and finally, death due to the absence of highly effective drugs, vaccines, and medical resources [4,5]. WHO has been prescribing many precautions to wear face masks, maintain social distancing, and the appliance of disinfectants and Personal Protective Equipment (PPE) to prevent the spread process of COVID-19 [6]. The spread process of COVID-19 has caused many impacts on different aspects of our life, starting from the economic, educational, to health aspects. For this reason, precaution measures that we often call the new normal must be fulfilled to minimalize the spread process of COVID-19.

While a new normal has been applied, some people are not wearing face masks in public places due to carelessness [7]. This situation cannot be solved by human resources alone, due to the insufficient number of official medical and police personnel. Thus, a new solution can be found in a more modern approach, which is a face mask recognition system, a system that is capable of recognizing the usage of the face mask [8]. This system is based on computer vision that can detect images, videos, and even real-time inspection

DOI: 10.24507/icicelb.14.04.415

[9]. By using face mask recognition, it can ease the public authorities' duties to confirm people wear face masks. However, to do so, computer vision always requires the same perception that is common between humans and machines. It works very well to make a machine have a human-like perceiving ability [10]. The current study has shown that convolutional neural network can be used to accomplish tasks of object detection. Object detection is a very handy system since it can detect and label various objects.

In the past few years, many object detection architectures were built and improved, such as MobileNetv2 and YOLO-v4. MobileNetv2 uses a small amount of computational power so it takes less power and is primarily used in embedded systems which can fit into many devices. While YOLO-v4 uses more computational power, the architecture specializes in object detection and implements other tools to further enhance the results of object detection. Through this research, we will create a face mask recognition for payment verification system and compare the two object detection architectures, namely YOLO-v4 and MobileNetv2, to assess which architecture is suitable for the face mask recognition for payment verification system.

2. **Literature Review.** Few pieces of research have shown the result of various deep learning frameworks in executing a face mask recognition task. Research in MobileNetv2 gives results with great accuracy [9,11]. In addition, research on YOLO-v4 shows a high percentage in all precision, recall, and F1-support [12]. Few kinds of research have shown the comparison between the performance of various object detection algorithms in accomplishing face mask recognition tasks. Research shows a comparison between MobileNetv2 with SVM or KNN and VGG19-KNN and Xception-SVM. MobileNetv2-SVM achieves the best accuracy among the rest [13]. There is also existing research showing the real-time implementation of face mask detection with detection algorithms at Politeknik Negeri Batam [14]. Another research also shows a modified YOLO-v4 was implemented in several schools in China [15].

3. **Proposed Method.** Figure 1 shows the flowchart of the method. First, dataset will be input for the training in corresponding algorithm, MobileNetv2 and YOLO-v4. Then, MobileNetv2 and YOLO-v4 will be trained using its constraint and dataset and the model of the two algorithms will be saved. The two models then will be tested in real time using webcam. If the models scan the person is not using a mask, the models will show a red box in the person's face, a note for no mask, and the percentage of probability. If the models scan the person is using a mask, the models will show a green box in the person's face, a note for mask, and the percentage of probability.

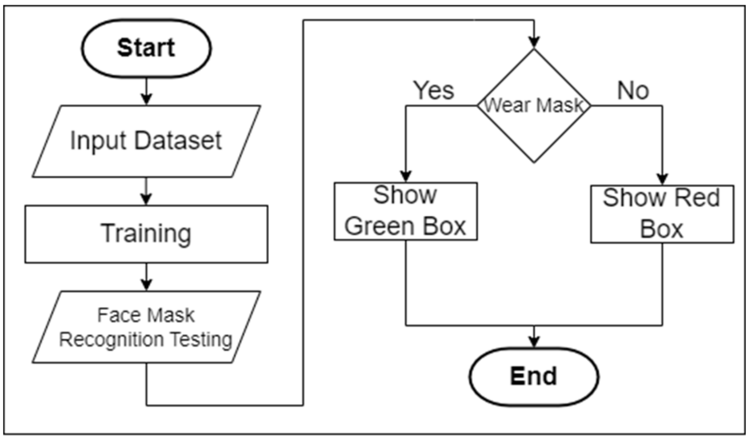


Figure 1. Flowchart diagram

3.1. **MobileNetv2.** MobileNet is an extremely efficient algorithm framework designed by Google for embedded systems using the Convolutional Neural Network (CNN) as a base. MobileNet is a reasonably accurate algorithm, but it uses fewer resources than other types of algorithms [11,16]. MobileNet is designed specifically for mobile applications requiring high speed. MobileNet evaluates the trade between accuracy with computations and between accuracy with the model size [17]. MobileNet only takes a low computation power which will be much better than any other algorithm, but with compromising significant accuracy of the result [18]. The first version (MobileNetv1) had a low model size and decent complexity cost due to the depth-wise separable convolution, so it is commonly used for low processing applications. It also allows collaboration with other classifiers to make it more accurate [1,13]. In one research, MobileNetv2, an improved model of MobileNetv1 can achieve a very high percentage in detecting face masks by combining with other models [7]. MobileNetv2 used the architecture of convolution building of blocks with a bottleneck depth that has 1 block stride with 3 levels of layers, which on the first layer has 1×1 convolutions with "ReLU6", then the second layer has convolution depth-wise, and the last layer contains a 1×1 linear convolution [19]. Also, MobileNetv2 can store the model with the TensorFlow lite format which is usually used in mobile applications that identifies a real-time object as an object detector for identifying some masked regions that are needed [20].

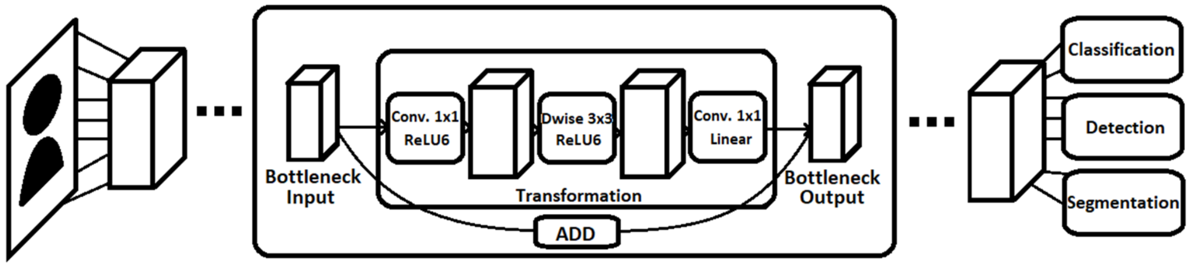


FIGURE 2. MobileNetv2 architecture

3.2. **YOLO.** YOLO is a type of algorithm specializing in the detection of an object [14,21]. YOLO accomplishes the detection of an object within one step by assuming the object detection as a regression problem, which allows YOLO to accomplish the detection speed of 45 fps on the Titan X GPU [22]. The original model of YOLO operates by dividing the image into a 19×19 grid of cells, each of which is responsible for executing the prediction of 2 bounding boxes by providing a level of confidence level if the existence of an object could be confirmed in both bounding boxes. The predictions made by the bounding box are not maximally suppressed. And the prediction having the highest value of Intersection over Union (IoU) scores will be assigned as the prediction result [12,23]. Several studies on face mask recognition managed to build a face mask recognition system using the improved algorithm of YOLO.

The research by AlexeyAB has improved YOLO into YOLO-v4 by providing SPP block, modifying SOTA method, proposing new data enhancement methods Mosaic and Self-Adversarial Training (SAT), and using PANet instead of FPN. The accuracy of YOLO-v4 has integrated the characteristics of YOLO-v1, YOLO-v2, and YOLO-v3 and has proven to be better than YOLO-v3. To utilize the superior learning ability of Cross-Stage Partial Network (CSPNet), CSPDarkNet53 has been constructed as the backbone for extracting features from images provided. For the neck, YOLO-v4 uses Spatial Pyramid Pooling (SPP) which consists of blocks that are used to increase the receptive field and combine parameters from various levels in the backbone [12,15,24].

Apart from the excellence that it has, the original YOLO model had several limitations. One of the difficulties is to detect objects that appear in groups of objects that have unusual aspect ratios or configurations, this will happen because each grid cell is responsible for predicting only 2 bounding boxes of fixed aspect ratios. In addition, localization inaccuracies could happen because the low and high-dimension localization errors will be treated equally by the loss function [23]. As an example, YOLO-v4 still has its shortcomings, such as the extraction feature of the multi-scale objects that is still insufficient, long training time, high calculation cost, and the high number of redundant information that is caused by too many gray bars in the model [24]. In one of the studies, in the training, insignificant channels will be cut off, so the model will be more compact and later be fine-tuned to get higher accuracy [15].

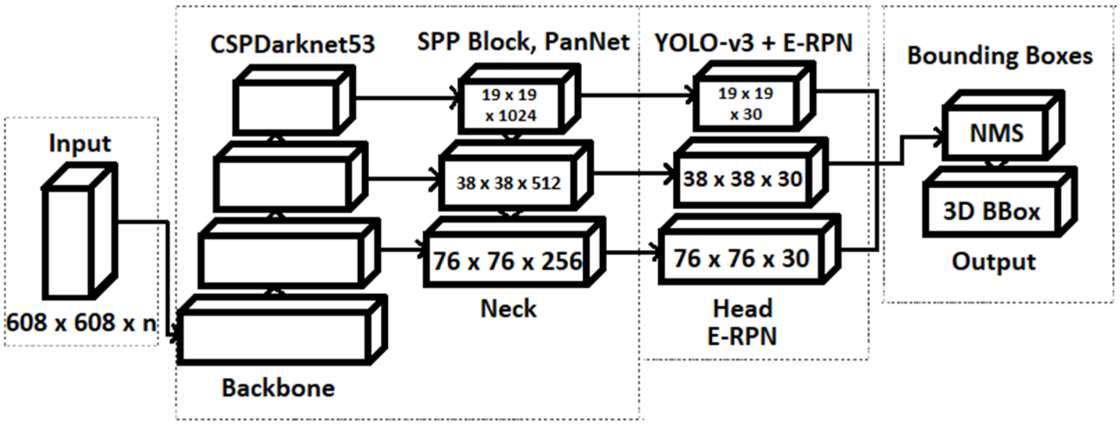


FIGURE 3. YOLO-v4 architecture

3.3. **UML diagram.** Figure 4 shows a use case diagram for the implementation of face mask recognition in the payment system. For the payment system, the first step is that the customer will order the food at the cashier counter. Then, the system will have a camera in it and scan the customer's face. After the system has scanned the customer's face and the customer wears a mask, then the customer can pay for the meals. If the customer does not wear a face mask, the customer cannot pay for the meals until the customer uses the face mask and is verified by the system.

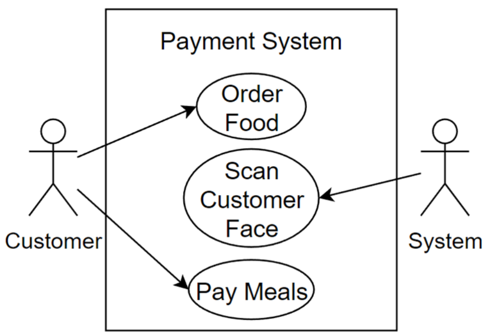


FIGURE 4. Use case diagram

Figure 5 shows an activity diagram for the use case "Scan Customer Face". The diagram shows how the system scans the customer whether the customer is wearing the face mask or not. It starts with the starting node, the black circle. First, the system will detect customers' faces in the live video. After it detects the customer's face, it will start scanning the customer's face. In the scanning, if the customer does not wear a mask, the system

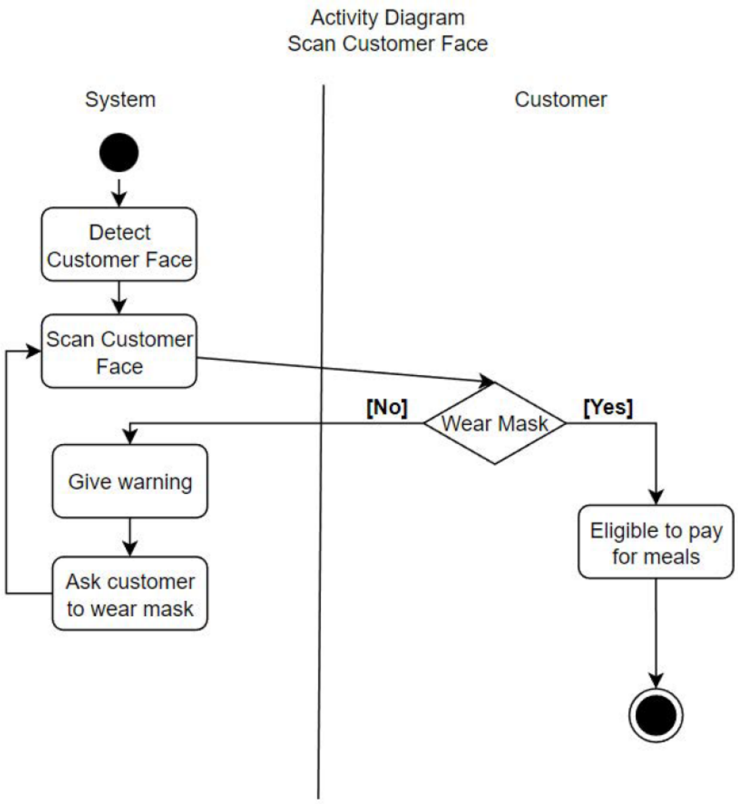


Figure 5. Activity diagram

will give a warning and ask the customer to wear a face mask. It will be repeated until the customer has to wear a face mask. If the customer does wear a face mask, the customer will be eligible to pay for the meals.

4. **Experiment and Results.** The comparison between the algorithms will be shown by various variables. The variables that will be used to compare the algorithms are precision, recall, and F1-Score. The formulae for the variables are shown below.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

where TP = True Positive; FP = False Positive

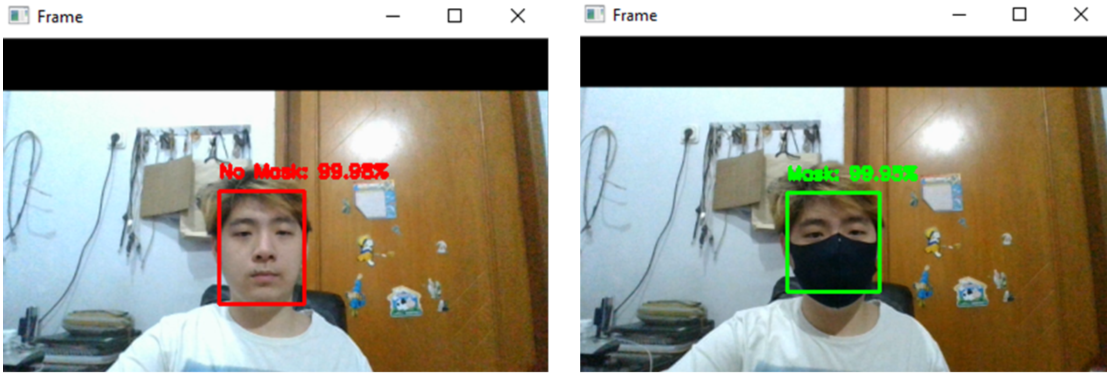
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

where TP = True Positive; FN = False Negative

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

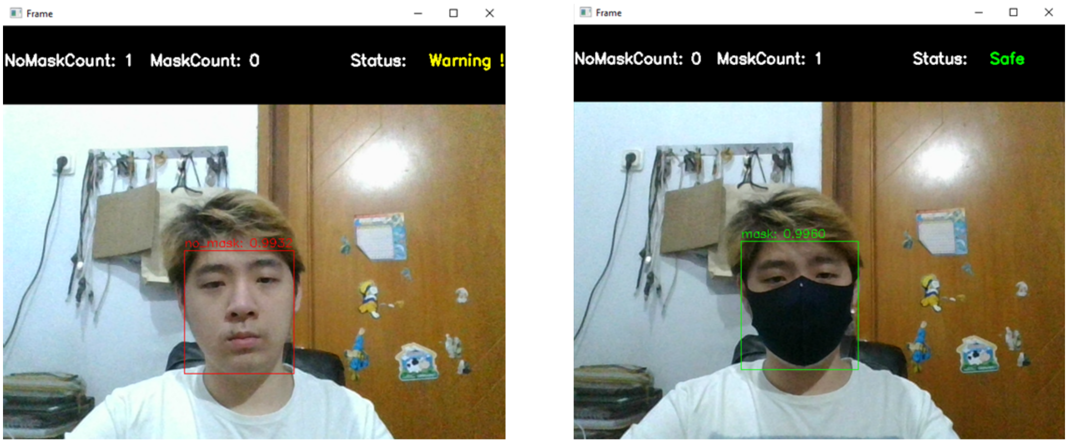
4.1. **MobileNetv2 face mask recognition.** Figure 6 shows the detection result from the MobileNetv2 algorithm. If the user uses the mask, it will show a green box with the percentage of the probability. Meanwhile, if the user does not use the mask, it will show a red box and show the percentage of the probability. MobileNetv2 also successfully ran with a good FPS score (tested with AMD Ryzen 7 4800H CPU and NVIDIA Geforce RTX 2060 GPU).

4.2. **YOLO-v4 face mask recognition.** Figure 7 shows the detection result from the YOLO-v4 algorithm. If the user uses the mask, it will show a green box with the percentage of the probability. Meanwhile, if the user does not use the mask, it will show a red box and show the percentage of the probability. YOLO-v4 face mask recognition ran



No Mask Mask

FIGURE 6. MobileNetv2 testing



No Mask Mask

FIGURE 7. YOLO-v4 testing

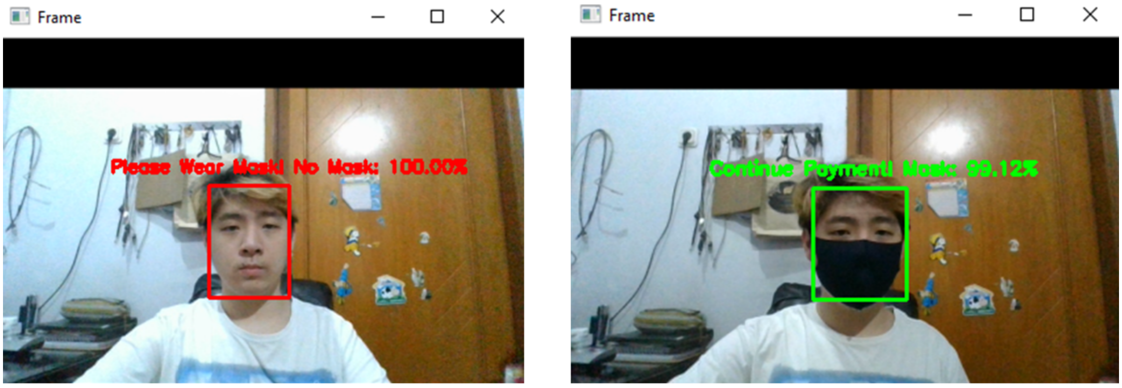
TABLE 1. Calculation based on the variables in each algorithm

Algorithm	Precision	Recall	F1-Score
MobileNetv2	0.98	0.98	0.98
YOLO-v4	0.92	0.95	0.93

with a low FPS score (tested with AMD Ryzen 7 4800H CPU and NVIDIA Geforce RTX 2060 GPU).

4.3. **Self-payment system implementation.** Based on the calculation from Table 1, MobileNetv2 has a higher percentage in the testing phase with 0.98 precision, 0.98 recall, and 0.98 F1-Score. And based on the executed testing phase, it is proven that YOLO-v4 will not be able to function smoothly in a certain environment due to its higher computation requirement compared to MobileNetv2. Hence, MobileNetv2 will be used as the implementation for an additional step in a self-payment system with some changes added in the text provided for mask and no mask detected.

Figure 8 shows the detection result from the MobileNetv2 algorithm that has been modified. If the user does not wear the face mask, the system will show a red box with the warning, "Please Wear Mask!" and the writing "No Mask:" and the percentage of the probability. Meanwhile, if the user wears the face mask, the system will show a green box on the customer's face with the writing "Continue Payment! Mask:" and the percentage of the probability (tested with AMD Ryzen 7 4800H CPU and NVIDIA Geforce RTX 2060 GPU).



No Mask Mask

Figure 8. MobileNetv2 implementation

5. **Conclusion.** The compared algorithms, MobileNetv2 and YOLO-v4, have a great performance in detecting whether someone is wearing a mask or not. In the testing phase, results showed that MobileNetv2 has a larger percentage in all the variables compared. Thus, MobileNetv2 is chosen as the final algorithm to be implemented as a part of the restaurant self-payment system. The outcome indicates that a self-payment system utilizing the MobileNetv2 algorithm is capable of efficiently detecting whether or not a consumer is wearing a face mask. This object detection approach can be used in additional self-paying direct transactional elements for future research. With any luck, this will aid in lowering the COVID-19 spread rate in a self-pay restaurant.

REFERENCES

[1] R. Iyer, P. S. Ringe, R. V. Iyer and K. P. Bhensdadiya, Comparison of YOLOv3, YOLOv5s and MobileNet-SSD V2 for real-time mask detection, *International Journal of Research in Engineering and Technology*, https://www.researchgate.net/publication/353211011, 2021.

[2] G. Yang et al., Face mask recognition system with YOLOV5 based on image recognition, *2020 IEEE 6th International Conference on Computer and Communications (ICCC2020)*, pp.1398-1404, doi: 10.1109/ICCC51575.2020.9345042, 2020.

[3] S. Singh, U. Ahuja, M. Kumar, K. Kumar and M. Sachdeva, Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment, *Multimedia Tools and Applications*, vol.80, no.13, pp.19753-19768, doi: 10.1007/s11042-021-10711-8, 2021.

[4] M. Guillermo, A. R. A. Pascua, R. K. Billones, E. Sybingco, A. Fillone and E. Dadios, COVID-19 risk assessment through multiple face mask detection using MobileNetV2 DNN, *The 9th International Symposium on Computational Intelligence and Industrial Applications (ISCIIA2020)*, 2020.

[5] D. U. Mata, Face mask detection using convolutional neural network, *J. Nat. Remedies*, vol.21, no.12, pp.14-19, 2021.

[6] S. Abbasi, H. Abdi and A. Ahmadi, A face-mask detection approach based on YOLO applied for a new collected dataset, *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, Tehran, Iran, doi: 10.1109/CSICC52343.2021.9420599, 2021.

[7] A. J. Santoso and R. E. Saragih, Automatic face mask detection based on MobileNet V2 and DenseNet 121 models, *ICIC Express Letters*, vol.16, no.4, pp.433-440, doi: 10.24507/icicel.16.04.433, 2022.

[8] S. Manjula and V. K. R, Real world face mask detection using MobileNetV2 and Raspberry Pi, *International Journal of Engineering Research and Applications*, vol.11, pp.26-32, doi: 10.9790/9622-1110012632, 2021.

[9] 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, *Sustainable Cities and Society*, vol.66, doi: 10.1016/j.scs.2020.102692, 2021.

[10] C. Gupta, Coronamask: A face mask detector for real-time data, *International Journal of Advanced Trends in Computer Science and Engineering*, vol.9, no.4, pp.5624-5630, doi: 10.30534/ijatcse/2020/212942020, 2020.

[11] S. K. Dey, A. Howlader and C. Deb, MobileNet Mask: A multi-phase face mask detection model to prevent person-to-person transmission of SARS-CoV-2, *Advances in Intelligent Systems and Computing*, vol.1309, pp.603-613, doi: 10.1007/978-981-33-4673-4_49, 2021.

[12] K. Bhambani, T. Jain and K. A. Sultanpure, Real-time face mask and social distancing violation detection system using YOLO, doi: 10.1109/B-HTC50970.2020.9297902, 2020.

[13] A. Oumina, N. el Makhfi and M. Hamdi, Control the COVID-19 pandemic: Face mask detection using transfer learning, doi: 10.1109/ICECOCS50124.2020.9314511, 2020.

[14] S. Susanto, F. A. Putra, R. Analia and I. K. L. N. Suciningtyas, The face mask detection for preventing the spread of COVID-19 at Politeknik Negeri Batam, *2020 3rd International Conference on Applied Engineering (ICAE)*, doi: 10.1109/ICAE50557.2020.9350556, 2020.

[15] X. Wang, Z. Chen, B. Wei and M. Ling, Application of pruning Yolo-V4 with center loss in mask wearing recognition for gymnasiums and sports grounds of colleges and universities, *2020 IEEE 6th International Conference on Computer and Communications (ICCC2020)*, pp.1373-1377, doi: 10.1109/ICCC51575.2020.9345257, 2020.

[16] W. Vijitkunsawat and P. Chantngarm, Study of the performance of machine learning algorithms for face mask detection, *InCIT 2020 – The 5th International Conference on Information Technology*, pp.39-43, doi: 10.1109/InCIT50588.2020.9310963, 2020.

[17] Y. Zhou, The efficient implementation of face mask detection using MobileNet, *Journal of Physics: Conference Series*, vol.2181, no.1, doi: 10.1088/1742-6596/2181/1/012022, 2022.

[18] P. Garg, S. Jadhav, U. N. Pandey and P. T. Suradkar, Face mask detection using Mobile Net V2, *International Journal of Advanced Research in Science, Communication and Technology*, pp.143-146, doi: 10.48175/ijarsct-1188, 2021.

[19] M. Almghraby and A. O. Elnady, Face mask detection in real-time using MobileNetv2, *International Journal of Engineering and Advanced Technology*, vol.10, no.6, pp.104-108, doi: 10.35940/ijeat.F3050.0810621, 2021.

[20] M. Mohan, M. Sojasomanan and M. Kaleeswari, Automatic face mask detection using Python, *Journal of Science Technology and Research*, vol.2, no.1, 2021.

[21] M. R. Bhuiyan, S. A. Khushbu and M. S. Islam, A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3, *The 11th International Conference on Computing, Communication and Networking Technologies*, 2020.

[22] S. Yadav, 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*, vol.8, no.7, pp.1368-1375, doi: 10.22214/ijraset.2020.30560, 2020.

[23] R. Liu and Z. Ren, Application of Yolo on mask detection task, *2021 IEEE 13th International Conference on Computer Research and Development (ICCRD2021)*, pp.130-136, doi: 10.1109/ICCRD51685.2021.9386366, 2021.

[24] J. Yu and W. Zhang, Face mask wearing detection algorithm based on improved YOLO-v4, *Sensors*, vol.21, no.9, doi: 10.3390/s21093263, 2021.