ICIC Express Letters
Volume **16**, Number **4**, April **2022**

 ISSN 1881-803X
pp. **433–440**

AUTOMATIC FACE MASK DETECTION BASED ON MOBILENET V2 AND DENSENET 121 MODELS

Albertus Joko Santoso* and Raymond Erz Saragih

Department of Informatics
Universitas Atma Jaya Yogyakarta
Jl. Babarsari No. 43, Janti, Caturtunggal, Depok, Sleman
Daerah Istimewa Yogyakarta 55281, Indonesia
raymondes@uvers.ac.id
*Corresponding author: albjoko@staff.uajy.ac.id

Received May 2021; accepted August 2021

Abstract. *The COVID-19 pandemic has brought significant impacts to the world. In Indonesia, public places such as malls, restaurants, shops, private and government offices, and public areas obliged visitors to wear masks. Unfortunately, there are times when visitors do not obey the rules by not wearing a mask; therefore, surveillance must be conducted. However, manual surveillance to check if a person wearing a mask can be a tedious task. This research aims to propose an automatic face mask detection that can detect if a person is using a mask or not. The proposed method combines face detection and classification using deep learning. The face detection is done using USM sharpening, CenterFace, and two pre-trained models, the MobileNet V2 and DenseNet 121 are used to classify if a person wears a face mask or not. The pre-trained models were fine-tuned using two datasets. Google Colab and libraries such as Tensorflow, Keras, and Scikit-learn were utilized. The research results show that the MobileNet V2 achieves higher performance and has a faster execution time.*
Keywords: COVID-19, Face mask detection, Deep learning, USM sharpening, CenterFace, MobileNet V2, DenseNet 121

1. **Introduction.** COVID-19 has brought significant impacts to the world [1]. COVID-19 has pushed governments around the world to create preventive actions to suppress the spread of COVID-19. Countries necessitate people to maintain personal hygiene through handwashing, and social distancing to prevent further transmission [2]. Another precautionary step is wearing a face mask to protect the wearer and others in spreading the COVID-19 [3].

The COVID-19 has a significant and pervasive impact on Indonesia. COVID-19 spread across all 34 provinces in Indonesia by April 2020 [4]. Batam, as one of the cities in Riau Island Province, Indonesia, is affected by COVID-19. The regional government has issued regulations to prevent the spread of COVID-19 in Batam, and one of them is wearing a face mask [5]. In Batam, public places, such as malls, restaurants, offices, and areas with many visitors, have complied with the regulations and obliged visitors to wear masks. Unfortunately, there are times when visitors do not wear masks as they have started to become careless. In such cases, there is a need to conduct surveillance in public areas to make sure that people use their masks. However, this can be a tedious task if the surveillance is done manually.

Machine learning has shown a significant impact on automating tasks, such as classification [6] and face recognition [7]. Specifically, deep learning has been a resounding success across numerous application areas, such as computer vision and image processing [8]. One of the commonly used approaches is the Convolutional Neural Network (CNN)

DOI: 10.24507/icicel.16.04.433

[9]. Deep learning can be applied to automating the surveillance of wearing face masks [10]. Therefore, this study aims to create an automatic face mask detection. The proposed method combines face detection and CNN. The face detection is done using CenterFace [6] and the pre-trained MobileNet V2 [11] and DenseNet 121 [12] are used to classify if a person wears a face mask or not.

This paper is organized as follows. Section 2 addresses several works related to face mask detection. The materials and proposed method are specified in Section 3. The results and comparison with existing studies are presented in Section 4. Lastly, the conclusion of the proposed method and future works are presented in Section 5.

2. **Related Works.** The emergence of the COVID-19 pandemic has caused people to use a mask as one of the prevention steps. Loey et al. proposed face mask detection using ResNet-50 through transfer learning for feature extraction with Support Vector Machine (SVM), Decision Trees, and ensemble algorithm for classification [13]. Three datasets were used to evaluate the proposed method, and the highest accuracy was achieved by using ResNet-50 with SVM, which is 99.64%, 99.49%, and 100%.

Oumina et al. proposed face mask detection using pre-trained VGG-19, Xception, and MobileNet V2 [14]. The pre-trained models were used as a feature extractor with SVM and K-Nearest Neighbor (KNN) as classifiers. The highest accuracy achieved is 97.11% by using MobileNet V2 and SVM. Venkateswarlu et al. proposed face mask detection using pre-trained MobileNet, ResNet50, and GoogleNet [15]. A global pooling layer was added to the MobileNet and could achieve an accuracy of 99.56% on the first dataset and 100% on the second dataset. The MobileNet has the smallest number of parameters and fastest training time. Sanjaya and Rakhmawan proposed face mask detection using the pre-trained MobileNet V2 [16]. Face detection was utilized and the MobileNet V2 was used as the predictor. The proposed model achieved an accuracy of 96.85%. Chowdary et al. [17] proposed a face mask detection method using a pre-trained Inception V3. The model was trained on Simulated Masked Face Dataset (SMFD) and achieved 99.9% on training and 100% on testing. Mercaldo and Santone proposed a real-time face mask detection using MobileNetV2 [18]. The MobileNetV2 was trained through transfer learning and achieved an accuracy of 98%. Jiang et al. proposed a face mask detection based on the YOLO v3, which is the SE-YOLOv3 [19]. The proposed method was trained using a dataset of three classes, with a mask, without a mask, and incorrect mask. The proposed method has a very good performance in mask detection and could be deployed in a real-time situation.

In this paper, the aim is to create an automatic face mask detection. The face detection used is CenterFace [20], which is a fast face detector, even using the CPU. Pre-trained MobileNet V2 [11] and the DenseNet [12], specifically the DenseNet 121, were utilized and were trained on the face mask datasets. Both models are small, compared to other models in Keras library such as VGG-16 that has the size of 528 MB, and Inception V3 with the size of 92 MB, while MobileNet V2 and DenseNet 121 have the size of 14 MB and 33 MB, respectively. The performance of both models will be compared. To train both models, transfer learning with fine-tuning is used. Previous works did not utilize a sharpening method; therefore, this work utilized one of the sharpening methods, which is the Unsharp Masking (USM) sharpening. The USM sharpening is utilized to enhance the input image, therefore reducing blurry faces in an image.

3. **Materials and Methods.** Two public datasets are used in this study. The first dataset is a face mask detection dataset from [21], which consists of 4,180 images of a face with mask and 1,569 images of a face with no mask. The second dataset was created by [22]. The dataset consists of 11,792 images of a face with a mask and without a mask. The dataset is divided into three sets: the training set, the validation set, and the testing set.

This research aims to classify if a person is wearing a mask or not. The proposed method is shown in Figure 1.

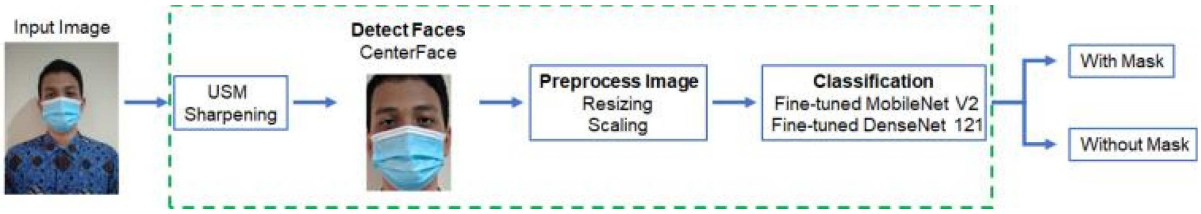


FIGURE 1. Proposed method

The system will receive an input image and the following is applying Unsharp Masking (USM) sharpening. Unsharp Masking (USM) is a generally used sharpening method that could increase the acuity of an image [23,24]. The USM works by blurring an image input followed by inverting and scaling down to produce a mask that will be combined with the original input image [23]. The USM was previously used to enhance images taken from a satellite and sharpen the edges [24]. In this work, the USM sharpening was utilized to create an enhanced image, thus reducing blurry faces with a mask. The face detector, which is CenterFace [20], will detect faces in the image. CenterFace is a fast deep learning-based face detector and a facial landmark locator. CenterFace can be run using a CPU with fast processing time and accuracy. Each detected face will be cropped for further processing. In the preprocessing stage, the cropped faces in the image will be resized into the size of 224×224 pixels and scaled following the required MobileNet V2 and DenseNet 121 input. The trained MobileNet V2 and DenseNet 121 will then classify the face and return the prediction result.

The DenseNet was proposed by [12] to create a new CNN architecture with few parameters and less computation. The MobileNet V2 is one of the known CNN models that was created by [11]. The MobileNet V2 was created to enable the model to run on mobile devices; therefore, it was created to work efficiently. The DenseNet and MobileNet V2 used in this work were trained on the ImageNet database, which consists of millions of images with a variety of classes [25]. Both pre-trained models are provided in the Keras library. The DenseNet in this work utilizes the DenseNet 121. The transfer learning method was used in this work. Transfer learning is an approach in which a pre-trained CNN is used for extracting features or can be retrained through fine-tuning [26]. As a feature extractor, the last fully connected layer of the pre-trained CNN is removed and train a classifier using the features extracted by the model. The fine-tuning approach is replacing and further retraining a classifier on top of the pre-trained CNN model, as well as fine-tuning the weights of several layers [27]. Training a CNN from scratch requires a significant number of images and computational resources. The transfer learning approach, however, becomes a solution when the dataset is relatively small and lacks computational resources. Therefore, in this work, the transfer learning with fine-tuning strategy is used on the MobileNet V2 and DenseNet 121, because the available dataset is relatively small.

4. **Results and Discussion.** In this work, the experiments were conducted using Google Colab and utilize several libraries such as Tensorflow, Keras, and Scikit-learn to create and evaluate the model. The first dataset was split into 80% training set and 20% testing set. The second dataset follows the division of the dataset, which are the training set, validation set, and testing set. Data augmentation was used and applied on the training sets to generate variations of the data as well as to reduce overfitting. The experiments evaluate first by freezing all the layers in the MobileNet V2 and DenseNet 121, and second

by unfreezing several top layers of each model. Each model is compiled using the Adam optimizer. The evaluation was done using accuracy, precision, recall, and F1-score.

The evaluation of each model on the first dataset is shown in Table 1 and the second dataset in Table 2. The precision, recall, and F1-score presented are the macro average score.

TABLE 1. Evaluation results on the first dataset

Model	Accuracy	Precision	Recall	F1-score
MobileNet V2 (Untuned)	94.696%	93.586%	92.971%	93.272%
MobileNet V2 (Fine-tuned)	98.435%	97.938%	98.128%	98.033%
DenseNet 121 (Untuned)	96.609%	96.931%	94.486%	95.617%
DenseNet 121 (Fine-tuned)	97.652%	97.869%	96.198%	96.991%

TABLE 2. Evaluation results on the second dataset

Model	Accuracy	Precision	Recall	F1-score
MobileNet V2 (Untuned)	99.294%	99.322%	99.275%	99.294%
MobileNet V2 (Fine-tuned)	99.798%	99.794%	99.804%	99.798%
DenseNet 121 (Untuned)	96.774%	96.827%	96.740%	96.769%
DenseNet 121 (Fine-tuned)	99.698%	99.700%	99.695%	99.697%

The result shows that before fine-tuning the MobileNet V2, the model achieves an accuracy of 94.696%. However, after fine-tuning, the model achieves an increase of 3.739% on the accuracy, which is 98.435%. Similarly, before fine-tuning the DenseNet 121, the model achieves an accuracy of 96.609%, and after fine-tuning, the accuracy increases to 97.652%, although, the accuracy increase on the DenseNet 121 is lower than the MobileNet V2, which is 1.043%. By fine-tuning, there is an increase in precision, recall, and F1-score as well. Comparing the overall performance results after fine-tuning, the MobileNet V2 achieves slightly higher than DenseNet 121, with an accuracy difference of 0.783%.

The results of both models on the second dataset are high. Before fine-tuning, the accuracy of MobileNet V2 is 99.294%, and after fine-tuning the MobileNet V2, the accuracy slightly increases to 99.798%. The fine-tuned MobileNet V2 as expected achieves slightly higher precision, recall, and F1-score as well, which are 99.794%, 99.804%, and 99.798%, respectively. Similarly, after fine-tuning, the performance of DenseNet 121 increases, with accuracy, precision, recall, and F1-score of 99.698%, 99.700%, 99.695%, and 99.697%, respectively. Comparing both models, the MobileNet V2 overall has higher performance than the DenseNet 121, although, both models could achieve relatively high performance:

The models are then tested to detect and classify faces with a mask or without a mask. Figure 2 shows the result in which the model tried to predict a single face in an image:

(a) (b) (c) [28] (d) (e)

FIGURE 2. Result of detecting mask on an image with a single face

Figures 2(a), 2(b), and 2(c) [28] with a face wearing a mask, while Figures 2(d) and 2(e) with a face without wearing a mask.

The result shows that the trained model could correctly predict a face with a mask and without a mask, even though the face is near the camera or quite far from the camera. However, in a real-world case, the taken image consists of multiple people and with quite distant from the camera. Therefore, several tests were conducted on an image with multiple distant faces. Figures 3(a) [29], 3(b) [30], 3(c) [31], 3(d) [32], and 3(e) [33] show the result of the prediction.

(a) [29] (b) [30]

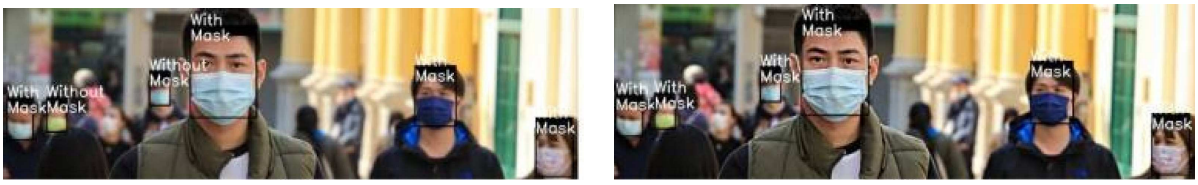
(c) [31] (d) [32]

(e) [33]

FIGURE 3. Prediction results of an image with multiple distant faces

Based on the result shown in Figure 3, overall, the face detection and the face mask classifier can correctly predict faces with a mask and without a mask. However, there are several shortcomings, such as faces that could not be detected as well as false predictions on several faces. The causes are that several faces in the image are blurred, especially when the face is too far from the camera and partially visible. As stated above, in this work the USM sharpening was utilized to the input image. The difference between not applying and applying USM sharpening to an image is presented in Figure 4.

As shown in Figure 4(a), by not applying USM sharpening, there are two wrong predictions. The model predicted each person not wearing a mask, although they are using a mask. The reason is that the faces are blurry. However, when applying USM sharpening as shown in Figure 4(b), the model could correctly predict the same faces using a mask. However, the usage of USM sharpening results in a longer execution time as additional processing is required. The execution time of each model on the test images is presented in Table 3 for image with a single face and Table 4 for image with multiple distant faces. The execution time is compared between using CPU and GPU. The multiple distant faces images have the size of 1920 pixels $\times$ 1280 pixels and 1920 pixels $\times$ 2560 pixels.



(a) (b)

FIGURE 4. Results of (a) not applying USM sharpening and (b) applying USM sharpening

TABLE 3. Execution time on image with single face

Model	CPU (second)					GPU (second)				
	1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
MobileNet V2	0.65	0.55	0.56	0.65	0.52	0.55	0.38	0.47	0.55	0.40
DenseNet 121	0.77	0.76	0.73	0.78	0.77	0.60	0.40	0.54	0.59	0.41

TABLE 4. Execution time on image with multiple distant faces

Model	CPU (second)					GPU (second)				
	1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
MobileNet V2	4.9	7.14	4.3	4.5	10	3.97	5.94	3.64	3.79	6.95
DenseNet 121	6.56	7.98	5.2	5.62	17.98	4.18	6.2	3.79	3.92	7.67

The result from Table 3 and Table 4 shows that overall, the MobileNet V2 has a faster execution time than the DenseNet 121. The utilization of GPU however brings significant reduction. The image size, on the other hand, is vital to the execution time. The bigger the image results in the longer the execution time. Table 5 shows the comparison between the proposed method with several previous works.

TABLE 5. Comparison with previous works

Author	Method	Best result
[13]	ResNet-50 with SVM, Decision Trees, and Ensemble	ResNet-50 with SVM – 99.64%, 99.49%, & 100%
[14]	VGG-19, Xception, MobileNet V2 with SVM and KNN	MobileNet V2 with SVM – 97.11%
[15]	MobileNet, ResNet50, and GoogleNet	MobileNet – 99.56% & 100%
[16]	MobileNet V2	96.85%
[17]	Inception V3	100%
[18]	MobileNet V2	98%
Proposed method	MobileNet V2, DenseNet 121	MobileNet V2 – 98.435% & 99.798%

Based on the best results of several previous works, the achieved best accuracy is between 96.85% to 100%. Our proposed method achieved the best accuracy of 98.435% and 99.798% by using MobileNet V2 and is considered satisfactory, compared to previous works. However, the work of [15] and [17], achieved the highest result, which is 100%, by using MobileNet and Inception V3. The work of [14], [16], and [18] used MobileNet V2 as well and achieved a considerably high accuracy result.

5. **Conclusions.** In this work, a face mask detection system is proposed. Face mask detection consists of face detection and classification by the CNN model. Face detection utilizes the CenterFace. The classification was done by fine-tuning pre-trained MobileNet V2 and the DenseNet 121, using two mask datasets. The USM sharpening was utilized to enhance the image and has shown improvement to the results. The difference in performances between the untuned and fine-tuned model was presented. The results show that by fine-tuning the model, higher accuracy can be achieved. The MobileNet V2 achieves an accuracy of 98.435% on the first dataset and 99.798% on the second dataset, while the DenseNet 121 achieves an accuracy of 97.652% on the first dataset and 99.698% on the second dataset. The model is then tested using images with a single face and images with multiple distant faces. The results show that the model could correctly predict all the images with a single face and could overall correctly predict images with multiple distant faces. However, there are several shortcomings when predicting multiple distant faces, such as blurred faces and the face is too far from the camera, as well as partially covered faces. Comparing the execution time, the MobileNet V2 has a faster execution time than the DenseNet 121 and using GPU could bring a significant reduction to the execution time. In the future, the plans are to create a more robust face mask detection system using different models, as well as adding masked face recognition and apply the face mask detection in the entrance of public areas, such as the mall, government office, in which the visitors are required to use a mask.

Acknowledgment. This research was funded by Universitas Atma Jaya Yogyakarta, Indonesia (No.: 024/In/R).

REFERENCES

[1] A. Dutta and H. W. Fischer, The local governance of COVID-19: Disease prevention and social security in rural India, *World Development*, vol.138, DOI: 10.1016/j.worlddev.2020.105234, 2021.

[2] C. J. Worby and H. H. Chang, Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic, *Nat. Commun.*, vol.11, no.1, pp.1-9, 2020.

[3] X. Liu and S. Zhang, COVID-19: Face masks and human-to-human transmission, *Influenza Other Respi. Viruses*, vol.14, no.4, pp.472-473, 2020.

[4] R. E. Caraka et al., Impact of COVID-19 large scale restriction on environment and economy in Indonesia, *Glob. J. Environ. Sci. Manag.*, vol.6, Special Issue, pp.65-84, 2020.

[5] *Information Center Regarding COVID-19 in Batam City*, https://lawancorona.batam.go.id/, Accessed on 03-Feb-2021.

[6] P. C. Sen, M. Hajra and M. Ghosh, Supervised classification algorithms in machine learning: A survey and review, in *Advances in Intelligent Systems and Computing*, J. K. Mandal and D. Bhattacharya (eds.), Singapore, Springer Singapore, 2020.

[7] H. Yang, W. Gan, F. Chen and X. Li, Face recognition using shearlets edges fusion, *International Journal of Innovative Computing, Information and Control*, vol.15, no.4, pp.1309-1322, 2019.

[8] M. Z. Alom et al., A state-of-the-art survey on deep learning theory and architectures, *Electronics*, vol.8, no.3, DOI: 10.3390/electronics8030292, 2019.

[9] H. Jo, D. Kim, K.-W. Pak and M. Kim, Road damage detection over road scanner images using deep convolutional neural network, *ICIC Express Letters*, vol.14, no.10, pp.1001-1008, 2020.

[10] M. Loey, G. Manogaran, M. H. N. Taha and N. E. M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection, *Sustain. Cities Soc.*, vol.65, DOI: 10.1016/j.scs.2020.102600, 2021.

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, MobileNetV2: Inverted residuals and linear bottlenecks, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.4510-4520, 2018.

[12] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, Densely connected convolutional networks, *Proc. of the 30th IEEE Conf. Comput. Vis. Pattern Recognition (CVPR2017)*, pp.2261-2269, 2017.

[13] M. Loey, G. Manogaran, M. H. N. Taha and N. E. M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, *Meas. J. Int. Meas. Confed.*, vol.167, DOI: 10.1016/j.measurement.2020.108288, 2021.

[14] A. Oumina, N. El Makhfi and M. Hamdi, Control the COVID-19 pandemic: Face mask detection using transfer learning, *2020 IEEE the 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, pp.1-5, 2020.
[15] I. B. Venkateswarlu, J. Kakarla and S. Prakash, Face mask detection using MobileNet and global pooling block, *2020 IEEE the 4th Conference on Information & Communication Technology (CICT)*, pp.1-5, 2020.
[16] S. A. Sanjaya and S. A. Rakhmawan, Face mask detection using MobileNetV2 in the era of COVID-19 pandemic, *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, pp.6-10, DOI: 10.1109/ICDABI51230.2020.9325631, 2020.
[17] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra and S. Agarwal, Face mask detection using transfer learning of InceptionV3, *Big Data Analytics*, pp.81-90, 2020.
[18] F. Mercaldo and A. Santone, Transfer learning for mobile real-time face mask detection and localization, *J. Am. Med. Informatics Assoc.*, 2021.
[19] X. Jiang, T. Gao, Z. Zhu and Y. Zhao, Real-time face mask detection method based on YOLOv3, *Electronics*, vol.10, no.7, DOI: 10.3390/electronics10070837, 2021.
[20] Y. Xu, W. Yan, G. Yang, J. Luo, T. Li and J. He, CenterFace: Joint face detection and alignment using face as point, *Sci. Program.*, pp.1-8, 2020.
[21] Face Mask Detection Dataset, *Kaggle*, https://www.kaggle.com/wobotintelligence/face-mask-detection-dataset, Accessed on 12-Feb-2021.
[22] A. Jangra, Face Mask ∼12K Images Dataset, *Kaggle*, https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset, Accessed on 12-Feb-2021.
[23] J. L. Clark, C. P. Wadhwani, K. Abramovitch, D. D. Rice and M. T. Kattadiyil, Effect of image sharpening on radiographic image quality, *J. Prosthet. Dent.*, vol.120, no.6, pp.927-933, 2018.
[24] X. Liu, J. Tao, X. Yu, J. Cheng and L. Guo, The rapid method for road extraction from high-resolution satellite images based on USM algorithm, *2012 International Conference on Image Analysis and Signal Processing*, pp.1-6, 2012.
[25] C. C. Aggarwal, *Neural Networks and Deep Learning*, Springer International Publishing, Cham, 2018.
[26] N. K. Manaswi, *Deep Learning with Applications Using Python*, Apress, Karnataka, 2018.
[27] R. Patel and A. Chaware, Transfer learning with fine-tuned MobileNetV2 for diabetic retinopathy, *2020 International Conference for Emerging Technology (INCET)*, pp.1-4, 2020.
[28] C. Feng, In China's epidemic, there is a staff member on each street to check the temperature registration, *Unsplash*, https://unsplash.com/photos/54DqfFCVoFQ, Accessed on 25-Feb-2021.
[29] M. P. Agency, People in Macau's Senado square during Chinese New Year season, Macau, China, *Unsplash*, https://unsplash.com/photos/62yAr99fX28, Accessed on 25-Feb-2021.
[30] G. C. Marino, People wearing face masks stroll along a street in Sulmona, Abruzzo, Italy, *Unsplash*, https://unsplash.com/photos/ms6tf_QVeSQ, Accessed on 26-Feb-2021.
[31] M. P. Agency, People in Macau's central district wearing face masks due to COVID-19 pandemic, Macau, China, *Unsplash*, https://unsplash.com/photos/CADsAEnFMjg, Accessed on 25-Feb-2021.
[32] M. P. Agency, Elderlies wearing face masks at a bus stop in Macau, China in the surroundings of the public hospital (RW), *Unsplash*, https://unsplash.com/photos/UC4fPkva91o, Accessed on 25-Feb-2021.
[33] M. P. Agency, Pandemic crowds, Macau, China, *Unsplash*, https://unsplash.com/photos/4yXV0JIKyo, Accessed on 25-Feb-2021.