ICIC Express Letters
Part B: Applications
Volume **14**, Number **11**, November **2023**

pp. 1125–1132

AUTOMATIC CLASSIFICATION OF FACE MASK WEARING CONDITIONS USING PRE-TRAINED CNN WITH FINE-TUNING

Albertus Joko Santoso[1,*] and Raymond Erz Saragih[2]

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

[2]Department of Informatics
Universitas Universal
Kompleks Maha Vihara Duta Maitreya, Bukit Beruntung, Sei Panas, Batam 29456, Indonesia
raymondes@uvers.ac.id

Received February 2023; accepted May 2023

Abstract. *People worldwide must follow the rules in public to reduce the transmission of COVID-19, such as wearing a mask. Problems arise when an individual or group violates the restrictions since they may transfer or get infected with COVID-19. Authorities may manually monitor public locations, but vast areas require additional staff. Therefore, this work proposed an automatic classification of face mask-wearing conditions using pre-trained Convolutional Neural Network (CNN) models. The automatic classification of face mask-wearing conditions uses deep learning-based face detection and several pre-trained CNNs to classify the people with, without, or incorrectly using their masks. Fine-tuning was used to train the pre-trained models. This work uses image sharpening to make the model correctly classify the face mask-wearing condition. The highest accuracy is 98.708% by EfficientNetB2 and EfficientNetB3.*
Keywords: CNN, Deep learning, Face mask, Fine-tuning

1. **Introduction.** The Coronavirus disease, called COVID-19, has profoundly influenced people's daily lives worldwide [1]. The pandemic has caused impacts on various fields, such as education [2,3], tourism [4-6], business [7,8], and health [9]. Several months after the initial COVID-19 outbreak in Wuhan, China, the virus rapidly spread to other countries and forced the government worldwide to take preventive action to reduce the transmission of the virus [9-11].

As the world struggled with the Alpha variant of SARS-CoV-2, new variants emerged (such as Beta and Gamma). The Delta version was first discovered in India, with a greater transmission rate [12,13] and was classified as a Variant of Concern (VOC) [14]. Nations worldwide have mandated specific guidelines for people outside in public to stop the spread of COVID-19. Wearing a medical or cloth mask, keeping a safe physical distance, regularly washing their hands, using hand sanitizer, avoiding crowds, and canceling public events are all required [15].

The two primary preventive measures one can do in public are wearing a mask and preserving physical distance. However, problems arise when an individual or group violates the restrictions since they may transfer or get infected with COVID-19. In public locations, the responsible authorities must take precautionary steps to admonish people who violate such rules. Specific public spaces have large areas, which makes manual monitoring difficult and necessitates the presence of personnel to monitor the public areas.

DOI: 10.24507/icicelb.14.11.1125

Over the years, researchers have continued to develop deep learning methods for computer vision, and deep learning model capabilities can outperform previous advanced machine learning methods [16]. One deep learning technique that falls into the supervised category is the Convolutional Neural Network (CNN). CNN has a wide range of applications, such as computer vision, natural language processing, speech recognition, face recognition [17,18], facial expression recognition [19], speech emotion recognition [20], agriculture [21,22], and forestry [23]. Training CNN from scratch requires a significant amount of data [24]. A common approach to addressing the problem is to use transfer learning [17,25,26]. In the case of image recognition, numerous CNN models were trained on large-scale datasets, such as ImageNet, and it is possible to use these pre-trained models without training them from scratch [17]. As a result, the transfer learning method, more specifically, fine-tuning, is utilized in this research.

Monitoring the public usage of a mask might be automated using CNN since it is feasible to train the model to distinguish between individuals who wear and those who do not. This work uses CNN to develop a model capable of distinguishing between people wearing and those not wearing masks. Another issue has arisen due to some people failing to wear their masks in public places [24] properly; hence, the proposed approach would also categorize those who do not wear them properly. Several works proposed automatic face mask detection. [29] used three pre-trained deep-learning models, MobileNet, GoogleNet, and ResNet50, to create automatic face mask detection. The proposed method used a global pooling block. The researchers evaluated the models using two public datasets and could achieve 99% and 100% on each dataset. An Internet of Things (IoT) system for rapid screening and face mask detection was proposed in [30]. The proposed method used pre-trained VGG16, ResNet50, Inception V3, MobileNetV2, and a CNN architecture and trained through fine-tuning. The highest accuracy achieved was using VGG16, with 99.81%. In [31], the authors used ResNet50 and trained through transfer learning for mask detection, comparing AlexNet and MobileNet. The work also utilized OpenCV to detect a face in the training dataset. The ResNet50 achieved a high accuracy of 98.2%. The proposed method from previous works used several models such as GoogleNet, ResNet50, MobileNet, and AlexNet. However, research on the performance of newer models is required.

This research presented an automatic classification of face mask-wearing conditions, which are with, without, or incorrectly wearing the mask. The novelty offered in this work is utilizing newer CNN models, EfficientNet (B0, B1, B2, B3, and B4) [27] and MobileNetV3 Small [28], and trained through fine-tuning using the combination of several face mask datasets as well as comparing the result of fine-tuning the last or last two blocks, which previous works did not perform. This research offered image sharpening for images with numerous distant faces. Finally, this work presents the performance of each model on images with several faces and a single face facing different directions. The application of image sharpening enables the model to distinguish face mask-wearing conditions correctly. The experimental results show that fine-tuning the last two blocks of each model can improve the accuracy and the utilization of image sharpening enables the proposed model to classify each face in the image correctly.

The following is the organization of this paper. The proposed method is addressed in Section 2. Section 3 presents the results and discussion of the experiments. Lastly, Section 4 offers the conclusion and the possibilities for future research.

2. Method.

This study offers an automatic classification of face mask-wearing conditions that employ deep learning-based face detection and pre-trained convolutional neural networks. The automatic classification of face mask-wearing conditions requires steps shown in Figure 1.

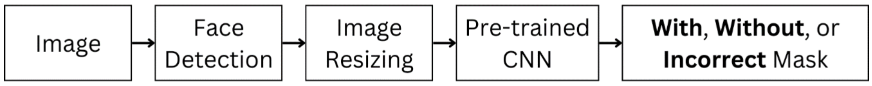


Figure 1. Proposed model of automatic classification of face mask-wearing condition

Face detection determines the face's location in the image. The face detection utilizes the CenterFace, deep learning-based face detection with a faster detection speed than MTCNN [32]. Further, each cropped face is resized to the required input size of each model. Image sharpening is done to each detected face in the image since the faces might be blurry. The sharpening utilizes the function provided in Tensorflow Addons. The classification uses several EfficientNet [27] versions and MobileNetV3 Small [28], classifying each detected face into three categories: with, without, or incorrectly wearing the mask.

This work trains six pre-trained EfficientNet (B0-B4) and MobileNetV3 Small. The models are relatively small compared to previous models, such as VGG16, ResNet50, and Inception V3. The fine-tuning process consists of two-stage [33]. The first stage is warm-up training, during which all model layers are frozen except for newly added layers. The second stage involves unfreezing the last layers and training the unfrozen layers and the newly added layers with a lower learning rate. The added layers are the average global pooling layer and softmax layer. The softmax layer is modified based on the number of classes. The initial and fine-tuning stages use 25 epochs to train the models. Training the models in both stages uses the Adam optimizer and unfreezing the layers in this work based on each model's last block [25]. Standard approaches to reducing overfitting include adding a dropout layer and image augmentation [25].

The datasets used in this work are the masked Face-Net [34], the properly-wearing-masked dataset [35], the face mask detection dataset [36], and the masked dataset with gan generated human face [37]. This work combines the images of the four chosen datasets and uses 48,000 images of three classes, with, without, and incorrectly worn masks. The training data uses 80% of the dataset and 20% for validation.

3. Results and Discussion.

This section presents the results of fine-tuning the six models. Table 1 presents the results of each model's accuracy evaluation on the training and validation datasets.

Table 1. Results of face mask-wearing classification

Models	Accuracy results			
	Before fine-tuning		After fine-tuning	
	Training (%)	Validation (%)	Training (%)	Validation (%)
EfficientNetB0	88.034	93.781	**99.036**	**98.510**
EfficientNetB1	87.784	93.615	98.339	98.438
EfficientNetB2	88.422	93.354	98.328	98.344
EfficientNetB3	88.464	92.823	98.555	98.438
EfficientNetB4	88.820	93.719	98.490	98.500
MobileNetV3 Small	88.167	93.781	98.193	97.823

Based on Table 1, EfficientNetB0 achieves the highest accuracy of 99.036% on the training set and 98.510% on the validation set, whereas the MobileNetV3 Small achieves the lowest. EfficientNetB4 achieves a slight difference from EfficientNetB0. Before fine-tuning, the accuracy of the training set ranges from 87%-88%. After fine-tuning the last block of each model, the accuracy rises from 98%-99%, a 9.6%-11% improvement. The accuracy in validation data increases after fine-tuning as well, which is roughly up to more than 4%-5%, with EfficientNetB3 achieving the highest increase of 5.615%.

Based on [25], it is feasible to fine-tune several last blocks of the pre-trained model. Therefore, to compare the results between fine-tuning only the last block with several last blocks requires further experiments. Table 2 shows the results of fine-tuning the last two blocks of each model.

Table 2. Results of fine-tuning the last two blocks of each model

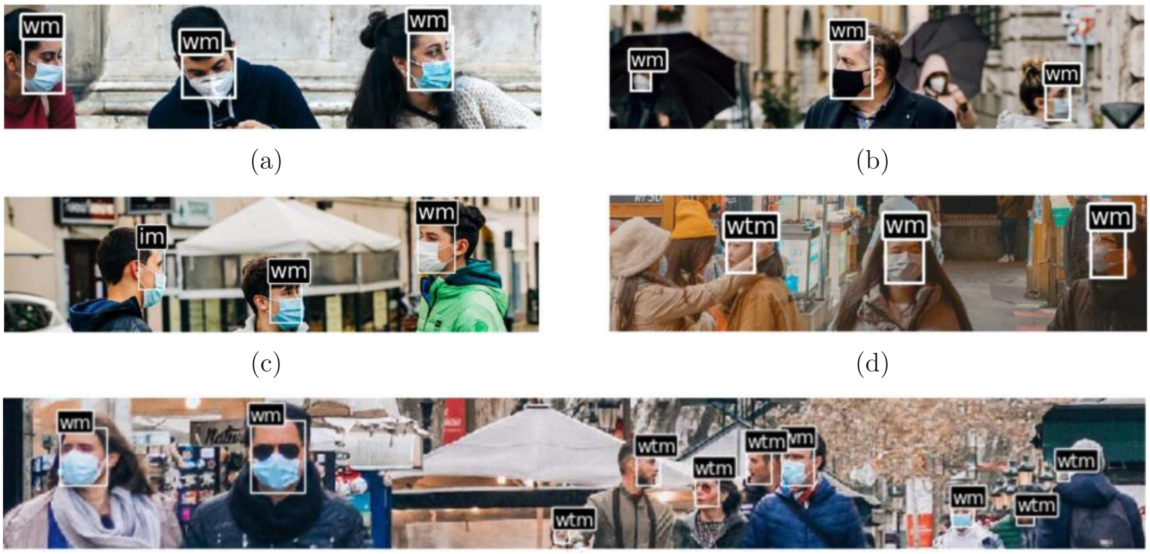
Models	Accuracy results			
	Before fine-tuning		After fine-tuning	
	Training (%)	Validation (%)	Training (%)	Validation (%)
EfficientNetB0	87.948	93.698	99.161	98.552
EfficientNetB1	87.794	93.323	99.177	98.531
EfficientNetB2	88.201	93.052	99.219	**98.708**
EfficientNetB3	88.180	92.646	99.352	**98.708**
EfficientNetB4	88.841	93.333	99.409	98.646
MobileNetV3 Small	88.055	93.458	99.292	98.333

Fine-tuning the last two blocks of each model improves the accuracy of training and validation datasets. Based on Table 2, EfficientNetB2 and EfficientNetB3 achieve the highest accuracy on the validation set, although the accuracy achieved differs from each other, which is in the range of 98%. The difference between before and after fine-tuning the model is more significant than the previous result in Table 1, specifically for the MobileNetV3 Small, which achieves 0.833% higher. EfficientNetB3 achieved the highest increase of 6.062%. In general, the accuracy of the training dataset improves by 10.5%-11%. In Table 1, fine-tuning the last block of EfficientNetB4 increased the training dataset by 9.67%, while fine-tuning the last two blocks increased the accuracy of the training dataset by 10.568%, which is a significant improvement. This work shows that further fine-tuning several model blocks can improve the model's accuracy in classifying the face mask-wearing conditions.

The evaluations of each model use images containing many faces and a single face. Figure 2 shows the results of detecting multiple faces and classifying the face mask-wearing condition of each face. The model classifies whether the face is with a mask (wm), without a mask (wtm), or an incorrectly worn mask (im). The images provided were the results of using EfficientNetB2, although using EfficientNetB3 achieves similar results.

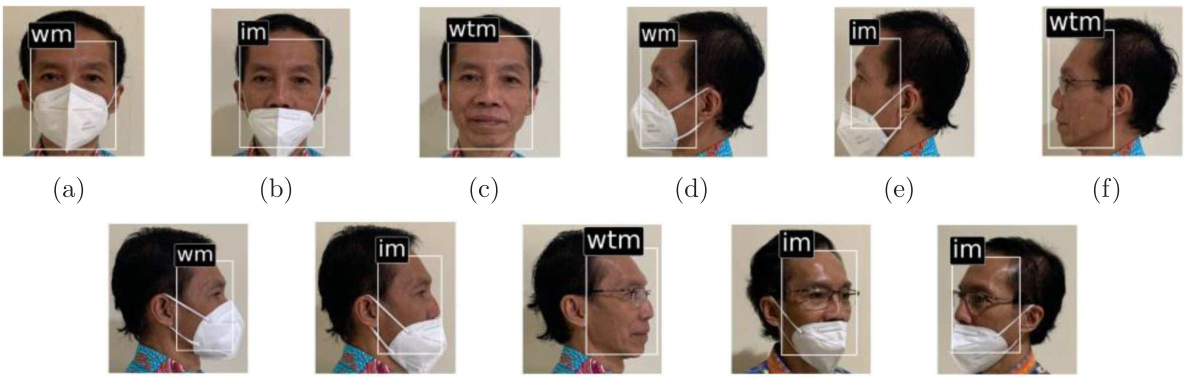
Figure 2 shows that the CenterFace face detection was unable to detect several faces as they are entirely obscured by the hood of a jacket, such as in Figure 2(b), or a hat and hair, such as in Figure 2(d). Overall, EfficientNetB2 can correctly predict the face mask-wearing condition of the people in the image. However, the result in Figure 2(c) shows an incorrect prediction. The predicted face mask-wearing condition of the person on the left is an incorrectly worn mask, while the truth is with a mask. Furthermore, the EfficientNetB2 was tested using images with a single face. Each image's single face faces forward, fully to the left and right, and around 30 to 45 degrees to the left or right, with different face mask-wearing conditions. Figures 3(a)-3(k) show the results of each test image.

As seen in Figure 3, EfficientNetB2 can correctly classify each image, although the single face faces fully to the right or left and around 30 to 45 degrees to the left or right. Although EfficientNetB3 achieves similar accuracy to EfficientNetB2, the model incorrectly classifies one facing forward faces, as shown in Figure 4. The EfficientNetB3 classifies incorrectly worn mask as with mask. In contrast, the EfficientNetB2 could classify it correctly, as shown in Figure 3(b).



(a) (b) (c) (d) (e)

Figure 2. Results of classifying face mask-wearing conditions in (a) [36], (b) [37], (c) [38], (d) [39], (e) [40]



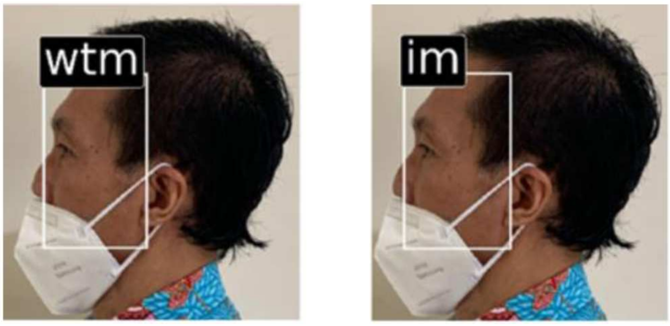
(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Figure 3. Results of classifying face mask-wearing conditions on a single face

(a) (b) (c)

Figure 4. Results of classifying face mask-wearing conditions using EfficientNetB3

Even though the image contains a single face, image sharpening can deliver a different result. Figure 5 shows the difference before and after applying image sharpening.



(a) (b)

FIGURE 5. The difference before (a) and after (b) applying image sharpening

The difference is noticeable. Before applying the image sharpening, the model incorrectly classifies an incorrectly worn mask as without a mask. After using the image sharpening, the result is accurate. Therefore, it is essential to apply image sharpening, regardless of whether there are multiple faces at varying distances or a single face close to the camera. Image sharpening can help the proposed model correctly classify the wearing condition of a face mask in an image.

4. **Conclusions.** This study proposes an automatic classification of face mask-wearing conditions to monitor the use of face masks in public. This study classified the face mask-wearing conditions using six pre-trained models: EfficientNet (B0, B1, B2, B3, and B4) and the MobileNetV3 Small. Each model was trained to classify faces with masks, incorrectly worn masks, and without masks. Fine-tuning the last block of each model could achieve a great result of 97.823%-98.510%. However, fine-tuning each model's last two blocks can improve accuracy. EfficientNetB2 and EfficientNetB3 achieved the highest accuracy of 98.708%. The utilization of image sharpening enables the proposed model to classify each face in the image correctly; therefore, image sharpening is essential. Although both EfficientNetB2 and EfficientNetB3 achieve a similar result, EfficientNetB3 has drawbacks, such as misclassification of the face facing forward.

In the future, the plan is to create a more robust model that could correctly classify the condition of face mask-wearing on a person by adding more training data. Additionally, it is possible to use other image enhancement techniques, such as deep learning-based image enhancement. It is also important to experiment on various newer models, particularly lightweight ones.

Acknowledgment. The authors gratefully acknowledge that Universitas Atma Jaya Yogyakarta supports this work.

REFERENCES

[1] K. Apiratwarakul et al., Emergency medical services amid new wave of coronavirus disease 2019 outbreak in Khon Kaen, Thailand, *Open Access Macedonian Journal of Medical Sciences*, vol.9, no.E, pp.492-495, DOI: 10.3889/oamjms.2021.6243, 2021.

[2] S. Pokhrel and R. Chhetri, A literature review on impact of COVID-19 pandemic on teaching and learning, *Higher Education for the Future*, vol.8, no.1, pp.133-141, DOI: 10.1177/2347631120983481, 2021.

[3] F. M. Reimers, Learning from a pandemic. The impact of COVID-19 on education around the world, in *Primary and Secondary Education during COVID-19*, Springer, Cham, Switzerland, pp.1-37, DOI: 10.1007/978-3-030-81500-4_1, 2022.

[4] L. P. Foo, M. Y. Chin, K. L. Tan and K. T. Phuah, The impact of COVID-19 on tourism industry in Malaysia, *Current Issues in Tourism*, vol.24, no.19, pp.2735-2739, DOI: 10.1080/13683500.2020.1777951, 2021.
[5] S. S. Yeh, Tourism recovery strategy against COVID-19 pandemic, *Tourism Recreation Research*, vol.46, no.2, pp.188-194, DOI: 10.1080/02508281.2020.1805933, 2021.
[6] N. A. A. Rahman, S. A. Rahim, M. F. Ahmad and B. A. M. Hafizuddin-Syah, Exploring COVID-19 pandemic: Its impact to global aviation industry and the key strategy, *International Journal of Advanced Science and Technology*, vol.29, no.6 (Special Issue), pp.1829-1836, 2020.
[7] B. H. Meyer, B. Prescott and X. S. Sheng, The impact of the COVID-19 pandemic on business expectations, *International Journal of Forecasting*, vol.38, no.2, pp.529-544, DOI: 10.1016/j.ijforecast.2021.02.009, 2022.
[8] D. Teresiene et al., The impact of the COVID-19 pandemic on consumer and business confidence indicators, *Journal of Risk and Financial Management*, vol.14, no.4, 159, DOI: 10.3390/jrfm14040159, 2021.
[9] J. Torales, M. O'Higgins, J. M. Castaldelli-Maia and A. Ventriglio, The outbreak of COVID-19 coronavirus and its impact on global mental health, *International Journal of Social Psychiatry*, vol.66, no.4, pp.317-320, DOI: 10.1177/0020764020915212, 2020.
[10] C.-W. Hsieh, M. Wang, N. W. M. Wong and L. K. Ho, A whole-of-nation approach to COVID-19: Taiwan's National Epidemic Prevention Team, *International Political Science Review*, vol.42, no.3, pp.300-315, DOI: 10.1177/01925121211012291, 2021.
[11] B. Rayamajhee et al., How well the government of Nepal is responding to COVID-19? An experience from a resource-limited country to confront unprecedented pandemic, *Front Public Health*, vol.9, pp.1-12, DOI: 10.3389/fpubh.2021.597808, 2021.
[12] L. Bian et al., Impact of the Delta variant on vaccine efficacy and response strategies, *Expert Review of Vaccines*, vol.20, no.10, pp.1201-1209, DOI: 10.1080/14760584.2021.1976153, 2021.
[13] E. Shim, Projecting the impact of SARS-CoV-2 variants and the vaccination program on the fourth wave of the COVID-19 pandemic in South Korea, *International Journal of Environmental Research and Public Health*, vol.18, no.14, 7578, DOI: 10.3390/ijerph18147578, 2021.
[14] A. M. Tareq, T. B. Emran, K. Dhama, M. Dhawan and T. E. Tallei, Impact of SARS-CoV-2 delta variant (B.1.617.2) in surging second wave of COVID-19 and efficacy of vaccines in tackling the ongoing pandemic, *Human Vaccines and Immunotherapeutics*, vol.17, no.11, pp.4126-4127, DOI: 10.1080/21645515.2021.1963601, 2021.
[15] O. Dadras et al., Effects of COVID-19 prevention procedures on other common infections: A systematic review, *European Journal of Medical Research*, vol.26, no.1, pp.1-13, DOI: 10.1186/s40001-021-00539-1, 2021.
[16] A. Voulodimos, N. Doulamis, A. Doulamis and E. Protopapadakis, Deep learning for computer vision: A brief review, *Computational Intelligence and Neuroscience*, vol.2018, pp.1-13, DOI: 10.1155/2018/7068349, 2018.
[17] L. Alzubaidi et al., Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions, *Journal of Big Data*, vol.8, no.1, p.53, DOI: 10.1186/s40537-021-00444-8, 2021.
[18] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, A survey of convolutional neural networks: Analysis, applications, and prospects, *IEEE Transactions on Neural Networks and Learning Systems*, pp.1-21, DOI: 10.1109/tnnls.2021.3084827, 2021.
[19] Q. T. Ngo and S. Yoon, Facial expression recognition based on weighted-cluster loss and deep transfer learning using a highly imbalanced dataset, *Sensors*, vol.20, no.9, 2639, DOI: 10.3390/s20092639, 2020.
[20] H. Cheng and X. Tang, Speech emotion recognition based on interactive convolutional neural network, *2020 3rd IEEE International Conference on Information Communication and Signal Processing (ICICSP 2020)*, vol.28, no.2, pp.163-167, DOI: 10.1109/ICICSP50920.2020.9232071, 2020.
[21] M. Momeny, A. Jahanbakhshi, K. Jafarnezhad and Y. D. Zhang, Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach, *Postharvest Biology and Technology*, vol.166, 111204, DOI: 10.1016/j.postharvbio.2020.111204, 2020.
[22] I. R. Paranavithana and V. R. Kalansuriya, Deep convolutional neural network model for tea bud(s) classification, *IAENG International Journal of Computer Science*, vol.48, no.3, pp.1-6, 2021.
[23] T. T. Yang, S. Y. Zhou and A. J. Xu, Rapid image detection of tree trunks using a convolutional neural network and transfer learning, *IAENG International Journal of Computer Science*, vol.48, no.2, pp.1-8, 2021.
[24] R. Patel and A. Chaware, Transfer learning with fine-tuned MobileNetV2 for diabetic retinopathy, *2020 International Conference for Emerging Technology (INCET)*, pp.1-4, DOI: 10.1109/INCET49848.2020.9154014, 2020.

[25] I. Kandel and M. Castelli, How deeply to fine-tune a convolutional neural network: A case study using a histopathology dataset, *Applied Sciences*, vol.10, no.10, 3359, DOI: 10.3390/app10103359, 2020.
[26] H. Noprisson, E. Ermatita, A. Abdiansah, V. Ayumi, M. Purba and H. Setiawan, Fine-tuning transfer learning model in woven fabric pattern classification, *International Journal of Innovative Computing, Information and Control*, vol.18, no.6, pp.1885-1894, DOI: 10.24507/ijicic.18.06.1885, 2022.
[27] M. Tan and Q. V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, *The 36th International Conference on Machine Learning (ICML 2019)*, vol.2019, pp.10691-10700, http://arxiv.org/abs/1905.11946, 2019.
[28] A. Howard et al., Searching for mobileNetV3, *Proc. of the IEEE International Conference on Computer Vision*, vol.2019, pp.1314-1324, DOI: 10.1109/ICCV.2019.00140, 2019.
[29] I. B. Venkateswarlu, J. Kakarla and S. Prakash, Face mask detection using MobileNet and global pooling block, *2020 IEEE 4th Conference on Information and Communication Technology (CICT)*, pp.1-5, DOI: 10.1109/CICT51604.2020.9312083, 2020.
[30] S. Hussain et al., IoT and deep learning based approach for rapid screening and face mask detection for infection spread control of COVID-19, *Applied Sciences (Switzerland)*, vol.11, no.8, DOI: 10.3390/app11083495, 2021.
[31] S. Sethi, M. Kathuria and T. Kaushik, A real-time integrated face mask detector to curtail spread of Coronavirus, *Computer Modeling in Engineering and Sciences*, vol.127, no.2, pp.389-409, DOI: 10.32604/cmes.2021.014478, 2021.
[32] Y. Xu, W. Yan, G. Yang, J. Luo, T. Li and J. He, CenterFace: Joint face detection and alignment using face as point, *Scientific Programming Towards a Smart World 2020*, vol.2020, pp.1-8, DOI: 10.1155/2020/7845384, 2020.
[33] D. O. Melinte and L. Vladareanu, Facial expressions recognition for human-robot interaction using deep convolutional neural networks with rectified adam optimizer, *Sensors*, vol.20, no.8, 2393, DOI: 10.3390/s20082393, 2020.
[34] A. Cabani, K. Hammoudi, H. Benhabiles and M. Melkemi, MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19, *Smart Health*, vol.19, 100144, DOI: 10.1016/j.smhl.2020.100144, 2021.
[35] *Properly-Wearing-Masked Detect Dataset*, https://github.com/ethancvaa/Properly-Wearing-Masked-Detect-Dataset, Accessed on August 03, 2021.
[36] V. Kumar, *Face Mask Detection*, https://www.kaggle.com/datasets/vijaykumar1799/face-mask-detection, 2021.
[37] Y. Goel, *Masked Dataset with Gan Generated Human Face*, https://www.kaggle.com/datasets/goelyash/masked-dataset-with-gan-generated-human-face, 2021.
[38] G. C. Marino, *Unsplash*, https://unsplash.com/photos/4AF5GiN8Qzs, Accessed on August 18, 2021.
[39] G. C. Marino, *Unsplash*, https://unsplash.com/photos/PSXgPRSSUCQ, Accessed on August 18, 2021.
[40] G. C. Marino, *Unsplash*, https://unsplash.com/photos/5rEP3kApQIs, Accessed on September 24, 2022.
[41] Y S, *Unsplash*, https://unsplash.com/photos/Wudw2aTYTig, Accessed on August 18, 2021.
[42] Y. Aziz, *Unsplash*, https://unsplash.com/photos/X1k2iK-BGT0, Accessed on September 24, 2022.