OPEN ACCESS

International Journal of Advances in Artificial Intelligence and Machine Learning Vol 1 No 1 (2024) https://doi.org/10.58723/ijaaiml.v1i1.286

Check for updates

Research Article

AI-Powered Face Mask Detection Utilizing MobileNetV2 for Health Monitoring

Received: Oct 04, 2024	Revised: Oct 21, 2024	Accepted: Nov 05, 2024	Publish: Nov 06, 2024

Misinem*, Eka Puji Agustini, Maria Ulfa

Abstract:
The COVID-19 pandemic has highlighted the critical need for face masks to prevent virus transmission. Ensuring consistent mask usage in crowded public spaces remains a challenge, especially with manual monitoring methods that are inefficient and prone to error. To address this, this research introduces a real-time face mask detection system leveraging MobileNet-V2, a lightweight and efficient deep learning model known for its high performance in image classification tasks. The system utilizes a dataset from Kaggle comprising 11,792 images, divided into training (10,000), validation (800), and testing (992) sets. MobileNet-V2 was fine-tuned for this task, using its inverted residual layers to extract features and enhance performance effectively. Data augmentation techniques were applied to improve the model's ability to generalize across diverse scenarios. The MobileNet-V2 model achieved an impressive 98.69% accuracy on the testing dataset, demonstrating exceptional reliability in identifying individuals wearing masks versus those without. Standard evaluation metrics, including precision, recall, and a confusion matrix, confirmed its robustness. This system's ability to operate in real-time makes it ideal for public health surveillance in environments such as airports, shopping malls, and public transport. The proposed face mask detection system is both accurate and scalable, offering an efficient solution for enforcing mask-wearing protocols in public spaces. The system's integration of advanced deep learning techniques ensures its reliability in real-time monitoring, contributing to better public health management. Future work will focus on further optimizing the model and expanding its application to other health-related monitoring tasks, enhancing its value for public health surveillance.

Keywords: Classification, Convolutional Neural Network, Kaggle dataset, MobileNetV2

1. INTRODUCTION

The COVID-19 pandemic, declared in March 2020, has necessitated the use of face masks to prevent virus transmission. Despite the easing of mask mandates, the resurgence of infectious diseases underscores the need for automated face mask detection systems. This study aims to develop a real-time face mask detection system using deep learning techniques. Face masks are an essential tool for reducing the chance of COVID-19 transmission. Even though face masks are no longer mandatory as the situation and vaccination rates rise, it is crucial to be aware of the future pandemic and resurgence of infectious diseases (Tabish, 2020). Therefore, an automated system for face mask detection is essential for future use.

Monitoring compliance with mask-wearing guidelines in public spaces is challenging and resource-intensive (Tabish, 2020). An automated system can alleviate the burden on human resources and ensure consistent enforcement of mask-wearing protocols (Curtis et al., 2022). Machine learning or deep learning techniques can help develop automated face mask detection systems (Kumar et al., 2022). These systems use cameras and image processing algorithms to identify the appearance of the face mask. Wearing a face mask indoors would no longer be required, but it would still be needed for all healthcare facilities, on public transportation, and for people who are positive for COVID-19.

There are still people who are infected with COVID-19, and there are also new COVID-19 variants that are rapidly spreading in China, such as BA.5.2 and BF.7, an Omicron sub-variant that had infected many people, by the 31st of December 2022 in Indonesia and other countries (Tallei et al., 2023). Respiratory infections are not new; apart from COVID-19, other respiratory diseases require the use of a face mask, such as Influenza H5N1 and H1N1. The third most frequent cause of mortality is lower respiratory tract infections, which are also the most prevalent infectious cause of death. Therefore, we must recognize the prediction of additional pandemics in the future and develop a system that can effectively identify face masks for the foreseeable future to play a critical role in respiratory infection control, especially in public places where social distancing is not always possible (Jewell et al., 2020).

Numerous studies have explored the application of deep learning techniques, particularly Convolutional Neural Networks (CNNs), to address the need for automated face mask detection. These methods have proven highly effective in image classification tasks, making them ideal for identifying the presence of face masks in images or video streams. The literature provides a range of approaches that focus on optimizing accuracy, computational efficiency, and real-time processing.

A significant contribution to this field is the use of MobileNet-V2, a lightweight CNN architecture designed for mobile and embedded vision applications. Its inverted residual layers and linear bottleneck enhance its performance while maintaining low computational costs, making it suitable for real-time face mask detection on resource-constrained devices. A study by Loey et al. (2021) demonstrated the effectiveness of MobileNet-V2 in a face mask detection system, achieving high accuracy with minimal computational overhead and making it deployable in real-world environments such as public spaces and healthcare facilities.

Other deep learning models, such as VGG16, ResNet50, and InceptionV3, have also been used in face mask detection, but these models typically require more processing power and are less suited to real-time applications (Nyarko et al., 2022). MobileNet-V2's architecture, specifically optimized for speed and efficiency, outperforms these models when considering the balance between accuracy and real-time performance.

Furthermore, data augmentation techniques are commonly used in deep learning models to enhance robustness. Methods such as image rotation, flipping, and zooming help improve the model's ability to generalize across diverse environments, ensuring high detection accuracy even in varied lighting or occlusion conditions (Kumar et al., 2020).

These studies highlight the growing importance of deep learning in developing automated face mask detection systems, particularly in the context of future pandemics and respiratory infection control. By leveraging state-of-the-art models like MobileNet-V2, these systems can ensure efficient, accurate, and real-time monitoring of mask compliance in public spaces.

MobileNetV2

This research presents a deep-learning approach using a Convolutional Neural Network (CNN) and the pre-trained MobileNet-V2 model. The CNN model is trained on a dataset consisting of images of individuals with and without masks.

MobileNetV2 Due to its lightweight architecture, MobileNet-V2 is a pre-trained deep-learning model designed for mobile and embedded vision applications, which was proposed by Sandler et al. (2018). The primary modifications in the MobileNet-V2 architecture involved the adoption of inverted bottleneck blocks and the incorporation of residual connections. These changes improved the model’s efficiency and performance for mobile and embedded device applications (Dong et al., 2020).

The MobileNet-V2 model consists of 53 layers. The architecture starts with a standard 3x3 convolutional layer having 32 channels, followed by 17 bottleneck blocks, each employing the specified channel expansion rate, number of output channels, and stride (Anditto et al., 2022). The network concludes with a regular 1x1 convolutional layer. A global average pooling layer is applied for classification, and then predictions are made using the classification layer.

2. MATERIAL AND METHOD

The research framework involves preliminary studies, literature review, data acquisition, model development, prototype design, and testing. The dataset was acquired from Kaggle by Ashish (2020). It consists of images categorized into two classes: “withMask” and “withoutMask” on the link https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset/data.

Data Acquisition

The dataset comprises a total of 11,792 images divided into three directories: training, validation, and testing. The training directory contains 10,000 images, with 5,000 images of people wearing masks and 5,000 images of people not wearing masks. The validation directory includes 800 images, with 400 images of people wearing masks and 400 images of people not wearing masks. The testing directory contains 992 images, with 509 images of people wearing masks and 483 images of people not wearing masks.

Data Pre-processing and Data Augmentation

The raw images were accessed from the dataset and stored in local folders. All images were standardized to a resolution of 256x256 pixels to ensure consistency and compatibility with the deep learning models. Next, data augmentation in Python code is shown in Figure 1.

```
# Image size
IMG_SIZE = (256, 256)

# data argumentation
train_datagen = ImageDataGenerator(rescale=1/255.0,
                                   rotation_range=45,
                                   shear_range=0.2,
                                   zoom_range=0.2,
                                   horizontal_flip=True,
                                   vertical_flip=True
                                   )

# data argumentation
test_datagen = ImageDataGenerator(rescale= 1 / 255.0)

train_dataset = train_datagen.flow_from_directory(train_dir, target_size=(IMG_SIZE),
                                          color_mode="rgb",
                                          batch_size=200,
                                          shuffle=True,
                                          class_mode="categorical")

test_dataset = test_datagen.flow_from_directory(test_dir, target_size=(IMG_SIZE),
                                          color_mode="rgb",
                                          batch_size=64,
                                          shuffle=True,
                                          class_mode="categorical")

validation_dataset = train_datagen.flow_from_directory(val_dir, target_size=(IMG_SIZE),
                                          color_mode="rgb",
                                          batch_size=64,
                                          shuffle=True,
                                          class_mode="categorical")
```

```
Found 10000 images belonging to 2 classes.
Found 992 images belonging to 2 classes.
Found 800 images belonging to 2 classes.
```

Figure 1. Data Augmentation

Based on Figure 1, Data Augmentation plays a critical role in enhancing the performance of deep learning models, especially in tasks like face mask detection, where the variability in real-world data is high. Data augmentation involves applying various transformations to the original dataset to increase its size and diversity artificially. This makes the model more robust and better able to generalize to new, unseen data. In the context of face mask detection, data augmentation can help the model learn to recognize masks in different conditions, such as variations in lighting, angles, face orientations, and occlusions.

The sample image was displayed in Figure 2, WithMask and WithoutMask.

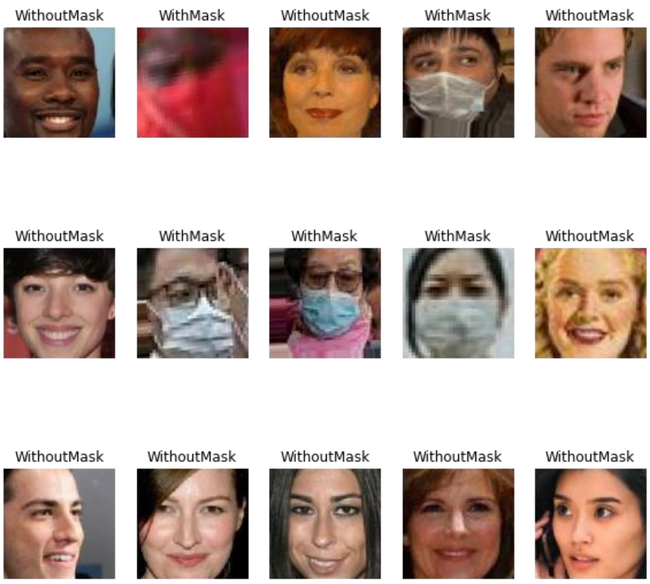

Figure 2. The sample image was WithMask and WithoutMask

The categorical terms “with mask” and “without mask” were transformed into numerical formats for model training. “with mask” was labeled as 0, and “without mask” was labeled as 1.

Various transformations were applied to the images to create additional variations and improve model generalization. These transformations included rescaling, rotation, horizontal and vertical flipping, and random rotation.

Model Development

The MobileNetV2 model is a lightweight architecture well-suited for mobile and embedded vision tasks, offering fast and efficient computation without sacrificing too much accuracy. By loading the model with pre-trained weights from ImageNet, the code leverages the model's knowledge, which has already been trained on a vast and diverse dataset. This allows for faster convergence during training and better initial performance, mainly when the available dataset for the task is limited.

Additionally, the model is initialized with the parameter include_top=False, which removes the fully connected classification layers at the top of the pre-trained model. This is a common practice when fine-tuning a pre-trained model for a specific task that requires a different set of output classes than the original ImageNet task, which contains 1,000 classes. In this setup, the user would likely add their task-specific classification layers, such as binary or multiclass classification heads, depending on the problem at hand, as shown in Figure 3.

```
# using mobilenetv2 model
mobilenet = MobileNetV2(weights='imagenet',include_top=False,input_s
hape=(256,256,3))
# make pre trained model into non trainable bcoz its takes much time
for layer in mobilenet.layers:
    layer.trainable = False
```

```
Downloading data from https://storage.googleapis.com/tensorflow/k
eras-applications/mobilenet_v2/mobilenet_v2_weights_tf_dim_orderi
ng_tf_kernels_1.0_224_no_top.h5
9412608/9406464 [==============================] - 0s 0us/step
9420800/9406464 [==============================] - 0s 0us/step
```

Figure 3. The MobileNetV2 model

Based on Figure 3, to further optimize the training process, the pre-trained layers in MobileNetV2 are set to be non-trainable by freezing them with the command layer.trainable = False. This ensures that the pre-learned weights from the ImageNet dataset are not updated during training, allowing the model to retain the valuable features learned from that dataset, such as edge detection, patterns, and textures. Freezing the layers helps reduce the computational requirements and training time, as only the new custom layers added on top will be updated during the learning process. This technique is particularly effective when the dataset is smaller or when time and computational resources are limited, helping to avoid overfitting.

The model downloads the pre-trained weights from TensorFlow's storage. This is a standard step when using pre-trained models, ensuring that the necessary weights are available for transfer learning. Overall, the setup optimizes fast, efficient training by leveraging transfer learning, making it highly effective for tasks such as face mask detection, wherereal-time performance and accuracy are critical. By using MobileNetV2, the model achieves a balance between computational efficiency and firm performance, particularly in scenarios where high-speed inference is essential.

3. RESULT AND DISCUSSION

The experimental findings from the MobileNet-V2 model were evaluated using the same training and validation sets. The test set was used to assess the model’s performance. The experimental results play a critical role in guiding the decision on the most suitable model for face mask detection.

MobileNetV2 Results

The training log presented shows the performance of a deep learning model across five epochs, providing valuable insights into its learning behavior. The model demonstrates steady improvement in both training and validation accuracy throughout the epochs. In the first epoch, the training accuracy starts at 64.14%, which progressively increases to 98.65% by the fifth epoch. Similarly, the validation accuracy improves from 94.38% in the first epoch to 98.62%

in the fifth epoch. These results suggest that the model is learning effectively and is generalizing well to the validation dataset.

In terms of loss, the training loss begins at 3.3110 in the first epoch and decreases to 0.0392 by the fifth epoch. This steady decline in training loss indicates that the model is successfully minimizing its errors during training. The validation loss follows a similar pattern, starting at 0.4011 and reducing to 0.0356 by the end of the fifth epoch. The combination of decreasing loss and increasing accuracy shows that the model is effectively learning the underlying patterns in the data, as shown in Figure 4 and Figure 5.

```
history = model.fit(train_dataset,
                    validation_data=validation_dataset,
                    epochs=5,
                    verbose=1)
```

```
Epoch 1/5
50/50 [==============================] - 346s 7s/step - loss: 3.3110 - accuracy: 0.6414 - val_
loss: 0.4011 - val_accuracy: 0.9438
Epoch 2/5
50/50 [==============================] - 315s 6s/step - loss: 0.2170 - accuracy: 0.9488 - val_
loss: 0.0921 - val_accuracy: 0.9675
Epoch 3/5
50/50 [==============================] - 316s 6s/step - loss: 0.0758 - accuracy: 0.9742 - val_
loss: 0.0508 - val_accuracy: 0.9837
Epoch 4/5
50/50 [==============================] - 321s 6s/step - loss: 0.0466 - accuracy: 0.9856 - val_
loss: 0.0404 - val_accuracy: 0.9837
Epoch 5/5
50/50 [==============================] - 316s 6s/step - loss: 0.0392 - accuracy: 0.9865 - val_
loss: 0.0356 - val_accuracy: 0.9862
```

Figure 4. The training MobileNetV2 model

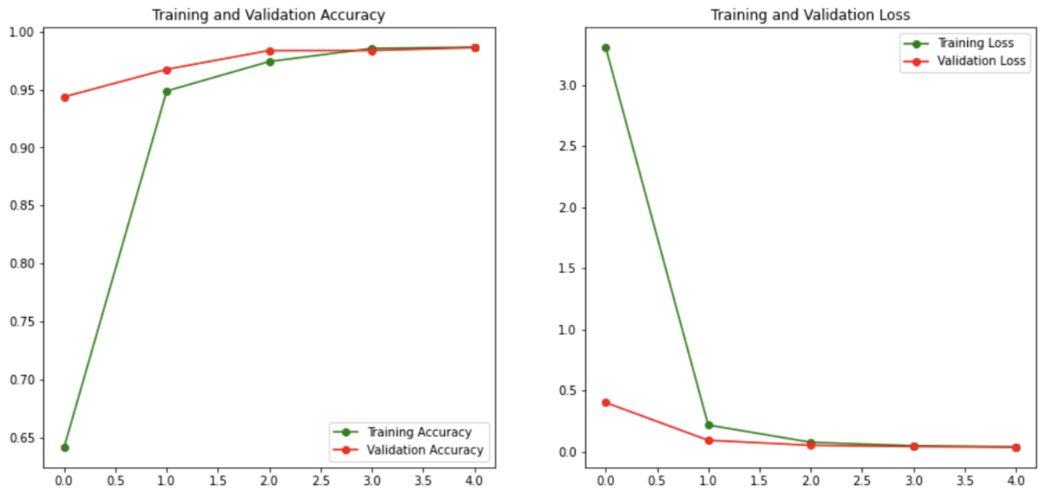

Figure 5. The training and validation accuracy and loss

The high validation accuracy, which is very close to the training accuracy, suggests that the model is not overfitting despite the rapid learning progress. This is a positive sign, as it indicates that the model is not just memorizing the training data but is able to generalize well to unseen validation data. However, the consistently high performance in such a short number of epochs might raise some concerns about potential overfitting, mainly if the dataset size is small.

To ensure the model's robustness, a few precautions could be considered. Implementing early stopping would allow the training process to halt once the validation performance stops improving, preventing unnecessary training and reducing the risk of overfitting. Additionally, the use of regularization techniques such as dropout could further mitigate the risk of overfitting. Another potential improvement would be adjusting the learning rate. Given the rapid convergence observed, lowering the learning rate and training for more epochs might help the model learn more effectively from the data and refine its feature extraction process.

The model exhibits strong performance, with both training and validation accuracy exceeding 98% by the fifth epoch. While the results are promising, further fine-tuning and monitoring for overfitting would help ensure that the model remains robust and performs well in real-world applications.

4. CONCLUSION

The research successfully developed a high-accuracy face mask detection system using the MobileNet-V2 model. The model achieved impressive results, showing that the model is learning effectively and minimizing errors, making it the most suitable choice for real-time face mask detection applications. The dataset used for training and testing was well-structured and comprehensive, providing a balanced representation of individuals with and without masks. The pre-processing steps ensured that the data was in a suitable format for training the deep learning models, contributing to the high accuracy achieved by the MobileNet-V2 model.

This consistent improvement in accuracy and the low final loss values highlight that the MobileNetV2 model is well-suited for the task of real-time face mask detection. Its ability to achieve such high accuracy in a short training period suggests it can perform well in real-world environments. The model's balanced performance across training and validation datasets indicates that it is not overfitting, making it a reliable tool for ensuring mask compliance in public spaces.

The MobileNetV2 model proves to be a robust and scalable solution for face mask detection, providing high accuracy and efficiency and the potential for real-time deployment in health monitoring systems.

AUTHOR INFORMATION

Corresponding Authors

Misinem, Universitas Bina Darma, Palembang, Indonesia.

Email: misinem@binadarma.ac.id

Authors

Eka Puji Agustini, Universitas Bina Darma, Palembang, Indonesia.

Maria Ulfa, Universitas Bina Darma, Palembang, Indonesia.

REFFERENCE

Anditto, R., & Roestam, R. (2022). SECURITY MONITORING USING IMPROVED MOBILENET V2 WITH FINE-TUNING TO PREVENT THEFT IN RESIDENTIAL AREAS DURING THE COVID-19 PANDEMIC. *SINTECH (Science and Information Technology) Journal*, *5*, 87–94. https://doi.org/10.31598/sintechjournal.v5i1.1023

Ashish Jangra. (2000). *Face Mask Detection ~12K Images Dataset*. https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset/data

Dong, K., Zhou, C., Ruan, Y., & Li, Y. (2020). MobileNetV2 Model for Image Classification. *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, 476–480. https://doi.org/10.1109/ITCA52113.2020.00106

Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. *JAMA*, *323*(19), 1893–1894. https://doi.org/10.1001/jama.2020.6585

Kumar, T. A., Rajmohan, R., Pavithra, M., Ajagbe, S. A., Hodhod, R., & Gaber, T. (2022). Automatic Face Mask Detection System in Public Transportation in Smart Cities Using IoT and Deep Learning. *Electronics*, *11*(6). https://doi.org/10.3390/electronics11060904

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable Cities and Society*, *65*, 102600. https://doi.org/https://doi.org/10.1016/j.scs.2020.102600

Nyarko, B. N. E., Bin, W., Zhou, J., Agordzo, G. K., Odoom, J., & Koukoyi, E. (2022). Comparative Analysis of AlexNet, Resnet-50, and Inception-V3 Models on Masked Face Recognition. *2022 IEEE World AI IoT Congress (AIIoT)*, 337–343. https://doi.org/10.1109/AIIoT54504.2022.9817327

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Tabish, S. A. (2020). Covid-19 Pandemic: Emerging Perspectives and Future Trends. *Journal of Public Health Research*, *9*(1), jphr.2020.1786. https://doi.org/10.4081/jphr.2020.1786

Tallei, T. E., Alhumaid, S., AlMusa, Z., Fatimawali, Kusumawaty, D., Alynbiawi, A., Alshukairi, A. N., & Rabaan, A. A. (2023). Update on the omicron sub-variants BA.4 and BA.5. *Reviews in medical virology*, *33*(1), e2391. https://doi.org/10.1002/rmv.2391