SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

SwiftMask: A Real-Time Deep Learning Solution for Facemask-Wearing Detection

Afi Izzati Asran[1], Salmah Fattah[1,*], Farkhana Muchtar[2], Asni Tahir[3], Siti Hasnah Tanalol[3], Waleed Abdelrahman Yousif Mohammed[4]

[1] Faculty of Computing and Informatics, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia
[2] Fakulti Komputeran, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[3] Preparatory Centre for Science and Technology, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia
[4] Faculty of Computer Studies and Information Technology, Nile University, Khartoum, Sudan

ABSTRACT

Despite the availability of vaccines and the emergence of new variants, the COVID-19 pandemic persists. Relying solely on temperature checks and hand sanitizers falls short of preventing virus transmission in public spaces. The World Health Organization recommends mask usage to counter the problems in public areas. However, some individuals neglect to wear masks, heightening the risk of COVID-19 transmission. Manually monitoring mask adherence, particularly in public places, presents challenges. Thus, this study aims to employ three neural network models - MobileNetV2, VGG16, and ResNet50V2 - to detect mask usage in images and real-time videos. After analysing factors such as classification results, training plots, and detection on images, MobileNetV2 emerges as the optimal model for real-time video mask detection by utilizing the deep learning model. The study encompasses training and applying the trained model to real-time face mask detection. The primary objective is to evaluate models' real-time mask detection performance, fostering awareness and enforcing safety measures against virus spread. The automated approach (SwiftMask) remains pragmatic, unaffected by human emotions, safeguarding individuals, families, and communities.

Keywords:

COVID-19; Pandemic; Face mask; Deep learning; Real-time; MobileNetV2

1. Introduction

Following the successful development and distribution of COVID-19 vaccines, a substantial portion of the global population has received vaccination. Consequently, public venues such as shopping malls, parks, airports, and schools have reopened, attracting a significant influx of visitors. As [1] indicates, face-to-face interaction was prohibited due to the infectious nature of the virus. This unprecedented event, unparalleled in the past decade except during the SARS period, has created significant uncertainty and challenges especially in education.

* *Corresponding author.*
E-mail address: salmahf@ums.edu.my

https://doi.org/10.37934/araset.60.1.239252

Therefore, the World Health Organization (WHO) emphasizes preventive measures, including regular handwashing with soap or sanitizer, maintaining a safe distance of two meters from individuals with fever or cough, and the crucial act of wearing masks and monitoring body temperature as signs of illness.

The method of detecting face masks enables us to determine whether an individual is wearing a mask, as mentioned in [2]. The face mask detection technique facilitates individuals' identification of mask usage. By utilizing cameras and analytical tools, the face mask detection technology swiftly identifies whether an individual is wearing a mask, resembling the process of recognizing objects in a specific environment. Numerous systems for detecting objects have been introduced, with a notable application of deep learning methods, especially within medical fields. Concerning computer vision, a deep learning algorithm autonomously learns fundamental representations from an initial image. Then, it constructs further models based on these fundamentals, repeating the process for higher levels, as outlined in [3]. Deep learning algorithms excel at complex tasks such as recognizing speech and handwritten characters, effectively acquiring knowledge and categorizing data representations. This method extracts both basic and advanced features to ensure precise classification. Deep learning algorithms derive representations from raw images in computer vision, progressively refining these representations for improved accuracy. Recent successes in object recognition underline the potential of deep learning architectures, which can be harnessed for mask detection on human faces.

Consequently, the author proposes developing a high-performance face mask detection model applicable to images and videos. This study involves conducting experiments using convolution neural network (CNN) architectures like MobileNetV2, VGG16, and ResNet50V2. As stated in [4], utilizing the CNN model, the face mask recognition system employs a dataset comprising various facial images with and without masks. Comprehensive experiments and performance assessments are undertaken on the datasets to compare and choose the most precise model for real-time mask detection.

This study makes several significant contributions to the field of face mask detection. In addition to presenting real-time face mask detection results, we performed a t-test to compare MobileNetV2 with VGG16 and ResNet50V2, offering a statistical comparison of their performance. Furthermore, we evaluated the performance of these algorithms on static images, introducing a novel aspect that distinguishes our research. Our study bridges the gap between practical application and theoretical evaluation by combining real-time detection with detailed algorithmic analysis.

This paper is structured as follows: Section 2 explores previous research efforts by various authors, highlighting diverse methods and their findings. Section 3 details the study's methodology, while Section 4 presents an overview of the evaluation results and discussion. Finally, Section 5 concludes the paper.

2. Related Works

Despite the widespread vaccinations, there has been a notable increase in COVID-19 cases. Respiratory particles from someone close to someone infected can be transmitted to others in various ways, as referenced in [5]. As described by Kek *et al.,* [6], tiny pathogenic particles are responsible for airborne infections, which can remain suspended in the air and be transmitted over long distances, ultimately ending up inhaled by susceptible individuals. According to [7], researchers have conducted extensive experimental and computational studies to deepen our understanding of the virus's behaviour and mechanisms of action. The solution proposed by Winata *et al.,* [8] involves implementing child respirators equipped with bamboo-based activated carbon filters to address health protocol enforcement among children. In the past, doctors frequently utilized the nebulizer machine to treat respiratory pneumonia diseases. However, its usage is not recommended during COVID-19 due to the heightened risk of virus transmission, as in reference [9].

Numerous deep learning models and algorithms have been developed to recognize faces and extract facial features, even when parts are hidden, like teeth. Therefore, as a fundamental step for non-metric cameras, camera calibration plays a vital role in photogrammetry. It is essential to calibrate every instrument intended for observation to ensure the utmost accuracy of data, as recommended by Amran *et al.,* [10]. Deep Learning, mentioned in [11], is a subset of Machine Learning that uses algorithms inspired by the brain's workings. It relies on layers of neural networks to turn input into helpful output, as explained in [12]. Table 1 summarizes and compares various studies, showing what problems they tackle, methods, and data accuracy. Furthermore, Lad *et al.,* [13] compared three models, Sequential CNN, VGG16, and MobileNetV2, to see which is best for face mask recognition. They looked at factors like accuracy, training time, and training accuracy/loss plot to decide. MobileNetV2 performed the best, achieving the highest accuracy of 99.2%. The findings from [14] indicate that a sequential approach is optimal for a straightforward and linear architecture. Therefore, opting for the sequential API in advance is advisable, as it yields a model that balances accuracy and loss.

Furthermore, in [15], transfer learning methods were applied using VGG16 and MobileNetV2 for facial feature extraction and identity classification, resulting in an overall accuracy of 95.4%, with MobileNetV2 surpassing VGG16. Additionally, as per Table 1 [16], MobileNetV2, MTCNN (Multi-Task Cascaded Convolutional Neural Networks), and DNN (Deep Neural Network) models were utilized for face mask detection through transfer learning, achieving an impressive overall accuracy of 98.4%, with MobileNetV2 demonstrating superior performance.

In Table 1, most other researchers typically compared models like VGG16 with MobileNetV2 or combined CNN with VGG16 and MobileNetV2 for face mask detection. However, our project took a distinctive approach by introducing ResNet50V2 into the mix. ResNet50V2 has yet to be compared with popular models like MobileNetV2 and VGG16 in prior studies, making our approach unique. Moreover, in contrast to the work of Faisal *et al.,* [15], which employed transfer learning for facial feature extraction and classification, this study implemented a deep learning approach utilizing convolutional neural network (CNN) models—MobileNetV2, VGG16, and ResNet50V2—specifically for real-time face mask detection. The proposed solution represents a deviation from the approach in [15], as we prioritized algorithmic accuracy and conducted comprehensive testing of these models in real-time face mask detection scenarios.

Moreover, our study delves deeper in contrast to other researchers who concentrated exclusively on real-time face mask detection. Aside from delivering real-time face mask detection outcomes, we also furnish the training durations for each of the three algorithms. Furthermore, we performed a t-test to compare MobileNetV2 with VGG16 and ResNet50V2. Additionally, we assessed the efficacy of these algorithms on static images, representing a distinctive approach from other researchers.

Remarkably, the presented project achieved a superior accuracy of 99%, exceeding the results obtained by both Faisal *et al.,* [15] and Sowmya *et al.,* [16]. This study employs deep learning algorithms for face mask detection, utilizing TensorFlow/Keras and OpenCV to train CNN models. TensorFlow is a powerful software library that facilitates machine learning predictions and mathematical calculations using data flow graphs [17]. On the other hand, Keras simplifies the creation of high-level neural networks with Python, making it accessible for developers to experiment with various ideas [18]. OpenCV, as an open-source library, aids developers in tasks like image analysis and offers valuable tools and learning resources [19].

Our approach introduces four evaluation metrics, including accuracy, precision, recall, and f1-score, to assess the effectiveness of the CNN model. Furthermore, we appraise these models using static images and conduct a t-test before finalizing the model for real-time face mask detection. This methodology sets our research apart from the three existing systems, enabling us to identify the most effective model for the subsequent phase of real-time evaluation.

Table 1
Literature review of the existing studies

Source	Focused problem	Method	Accuracy	Result
Lad *et al.,* [13]	Real-time face mask detection	Sequential CNN, VGG-16 and MobileNetV2	99.2%	Accuracy, Training time and training accuracy or loss plot, real-time face mask detection
Faisal *et al.,* [15]	Face Mask Detection	VGG16 and MobileNetV2	95.4%	Accuracy with training accuracy or loss plot.
Sowmya *et al.,* [16]	Identify Recognition with Mask	MobileNetV2, MTCNN (Multi-Task Cascaded Convolutional Neural Networks), and DNN (Deep Neural Network	98.4%	Accuracy, training accuracy or loss plot and real-time face mask detection.
The proposed project	Real-time face mask detection	MobileNetV2, VGG16 and ResNet50V2	99%	Accuracy, Training time, training accuracy or loss plot, static images, t-test and real-time face mask detection.

3. Methodology

CNNs, abbreviated as Convolutional Neural Networks, are crucial in recognizing and categorizing images based on their acquired features. The construction of face mask detection systems encompasses two phases that integrate computer vision and deep learning techniques facilitated by Python, OpenCV, and TensorFlow/Keras, as depicted in Figure 1.

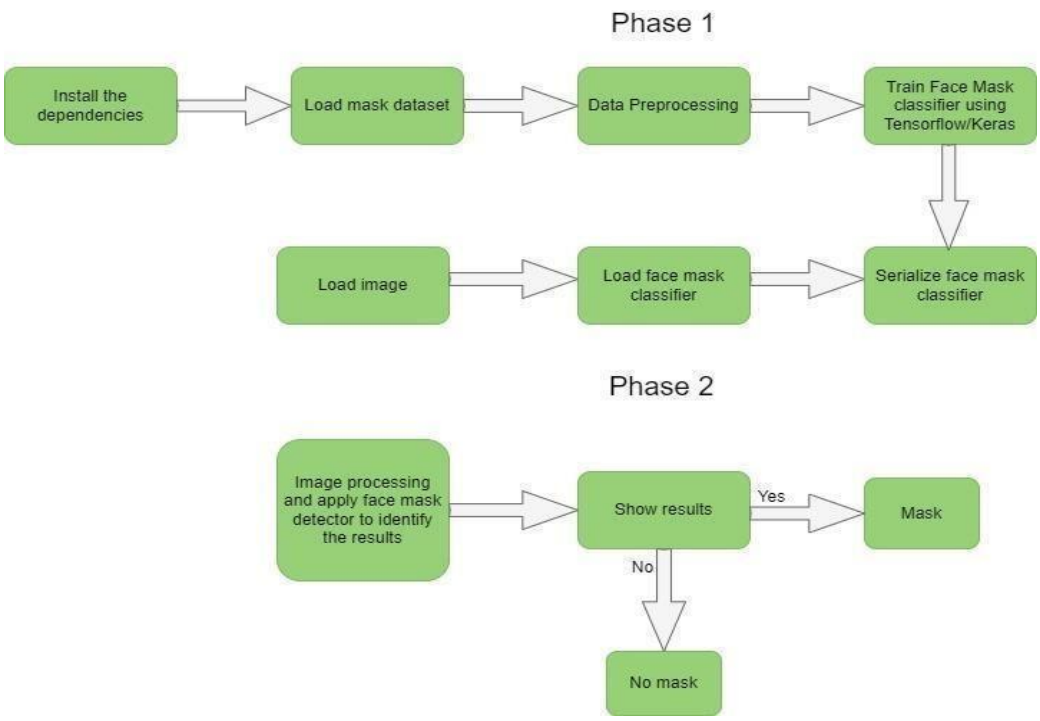

Fig. 1. Phases for building a face mask detection

3.1 Phase 1: Train Phase
3.1.1 Datasets

To collect data for face mask detection, we intend to utilize open sources in the proposed study. These sources may include freely available images and datasets accessible on GitHub, involving three distinct datasets. Selecting an appropriate and accurate dataset is paramount to ensure precise outputs and minimize potential errors, thereby preventing biased results. A more extensive collection of images contributes to enhanced accuracy. These open sources will encompass images depicting individuals both with and without masks. It is worth noting that the dataset used for image classification or training will remain consistent across all three models. Table 2 shows the dataset specifications of the proposed study.

Table 2
Dataset Specification

Source	Data with_mask	Data without_mask	Number of Data	File Type
Deb [29]	2,165	1,930	4,095	*jpg & *png

3.1.2 Data preprocessing

To build a highly accurate model, starting with a clean and error-free dataset is crucial. This study manually removes duplicates and mistakes to ensure the images are processed and in good condition. Dataset images can have problems like errors, copies, and noise, especially when collected from various open sources [20]. The model's input consists of these pre-processed images. The first step is ensuring all photos are the same size (224 x 224) because deep learning models can only work with arrays. This study used Python to convert the images into collections, as this step is essential. Next, the authors divided some of the dataset images into two groups: one for training and one for testing. Then, we applied various transformations like shifting, zooming, rotating, or changing the image's orientation. The process creates different styles of the same idea. The function generates these transformed images and tests them as a batch of data, effectively creating more images by altering some original image properties.

3.1.3 Image classification

Image classification is the final crucial stage in developing the face mask detection system. At this stage, researchers train deep learning models using labelled images or datasets to detect, recognize, and classify these images based on visual patterns. This study leveraged an open-source solution from TensorFlow and employed OpenCV and Python. The comprehensive approach encompasses the utilization of three CNN models.

i. <u>MobileNetV2:</u> In Experiment 1, the author used the Mobilenetv2 package from TensorFlow to prepare datasets containing images with and without masks for use with the Mobilenetv2 architecture. MobilenetV2, as described in [21], is a model already trained on a large dataset. The developers do not have to create and train a neural network from the ground up, saving them valuable development time. As indicated by Suresh *et al.,* [22], the proposed solution employs MobileNet as the foundational architecture and utilizes TensorFlow for model training, resulting in commendable performance in terms of accuracy, precision, recall, and F1-score. This approach demonstrates the effectiveness of leveraging pre-trained models for face mask detection.

ii. VGG16: In Experiment 2, the study utilized the VGG16 package from TensorFlow to preprocess two datasets containing images with and without masks for compatibility with the VGG16 architecture. The VGG16 model, introduced by Simonyan *et al.,* [23], operates as a convolutional neural network. The numerical value "16" within VGG16 denotes the total number of layers within the network, comprising thirteen convolutional layers, five Max Pooling layers, and three Dense layers, summing up to 21 layers in the architecture [24]. However, it's crucial to note that the network comprises only sixteen weight layers, simplifying the parameter learning process.

iii. ResNet50V2: In Experiment 3, the author employed the ResNet50V2 package from TensorFlow to preprocess datasets containing images with and without masks for compatibility with the ResNet_V2 architecture. The ResNet50 network represents a specialized neural network, distinguished by its greater depth than more common ResNet models. With a total of 48 layers, each layer contains various types of neurons, complemented by both MaxPool and Average Pool layers. Notably, the ResNet50V2 network surpasses the original ResNet50 and the newer ResNet101 networks' ability to recognize image patterns [25].

3.1.4 Apply the trained model face masks on static image

The three trained models evaluate the performance using static images, examining whether individuals in the images are wearing masks or not. Subsequently, the models will proceed to detect face masks in these images.

3.2 Phase 2: Apply the Trained Model on Real-Time Face Mask Detection

The selected model will be used with a webcam for real-time testing to detect whether individuals are wearing masks.

4. Results and Discussion

This section presents the results of evaluating three models with four assessment metrics: accuracy, precision, recall, and f1-score. The paper proceeds to discuss the performance of the three trained models in detecting face masks in static images. It then identifies the most effective real-time face mask detection model based on accuracy and performance in static images.

4.1 Classification Report for the Three Models Used

Figure 2, Figure 3 and Figure 4 shows the classification report for the MobileNetV2, VGG16, and ResNet50V2.

```
[INFO] evaluating network...
26/26 [==============================] - 16s 572ms/step
              precision    recall  f1-score   support

   with_mask       0.98      1.00      0.99       433
without_mask       0.99      0.98      0.99       386

    accuracy                           0.99       819
   macro avg       0.99      0.99      0.99       819
weighted avg       0.99      0.99      0.99       819
```

Fig. 2. Classification report for MobileNetV2

```
[INFO] evaluating network...
26/26 [==============================] - 148s 6s/step
              precision    recall  f1-score   support

   with_mask       0.92      0.94      0.93       433
without_mask       0.93      0.91      0.92       386

    accuracy                           0.92       819
   macro avg       0.92      0.92      0.92       819
weighted avg       0.92      0.92      0.92       819

[INFO] saving mask detector model...
```

Fig. 3. Classification report for VGG16

```
[INFO] evaluating network...
26/26 [==============================] - 50s 2s/step
              precision    recall  f1-score   support

   with_mask       0.99      1.00      0.99       433
without_mask       0.99      0.99      0.99       386

    accuracy                           0.99       819
   macro avg       0.99      0.99      0.99       819
weighted avg       0.99      0.99      0.99       819
```

Fig. 4. Classification report for ResNet50V2

The MobileNetV2 network, originating from research [26], employs a unique neural network architecture, introducing two key features: linear bottlenecks and shortcut links between these bottlenecks [27]. Regarding precision, MobileNetV2 achieves 98% for with_mask and 99% for without_mask (Table 3). The recall rates are 100% for with_mask and 98% for without_mask, while the f1-scores reach 99% for both with_mask and without_mask. The accuracy is 99%, with 99 correct predictions out of 100 examples.

On the other hand, the VGG16 method yields 92% precision for with_mask and 93% for without_mask. In recall, with_mask scores 94%, while without_mask reaches 91%. The f1-scores are 93% for with_mask and 92% for without_mask. The accuracy rate is 92%, indicating 92% correct predictions based on the classification report.

Lastly, ResNet50V2 demonstrates outstanding precision results of 99% for both with_mask and without_mask, accompanied by recall rates of 100% for with_mask and 99% for without_mask. The f1-scores for both with_mask and without_mask also reach 99%. Notably, the accuracy is 99%, mirroring MobileNetV2's performance with 99 correct predictions from 100 total examples.

Table 3
Summary of classification results of each model

Method	Accuracy	avgPrecision	avgRecall	avgF1-score
MobileNetV2	99%	98.5%	99%	99%
VGG16	92%	92.5%	92.5%	92.5%
ResNet50V2	99%	99%	99.5%	99%

Analysing the classification results, both MobileNetV2 and ResNet50V2 achieved an accuracy of 99%, while VGG16 scored 92%. According to Shafiq *et al.,* [28], the efficiency of ResNet50V2 exceeds that of VGG19, attributable to its 19.6 billion floating-point operations per second (FLOPS). However, based solely on accuracy results, which model to select for the next phase needs to be determined. The study must examine the training curve plot to make a decision. This analysis will help determine whether MobileNetV2 and ResNet50V2 exhibit an excellent fit plot, underfitting, or overfitting. This assessment is crucial in selecting the model for real-time video deployment.

4.2 Accuracy/Loss Plot for the Three Models Used

Figure 5, Figure 6 and Figure 7 depict the training and loss curves of the models, showing the loss on the training and validation datasets across training epochs. Figure 5 illustrates that over 20 epochs, MobileNetV2 achieved a training loss of 3.89% and an accuracy of 98.77%, while VGG16 attained a training loss of 5.30% and an accuracy of 98.18%. ResNet50V2 outperformed them with a training loss of 2.89% and an accuracy of 99.54%. For a comprehensive summary of training metrics, please refer to Table 4, which includes training loss, accuracy, val_loss, and val_acc for all three methods.

Fig. 5. Training Loss and Accuracy plot for MobileNetV2

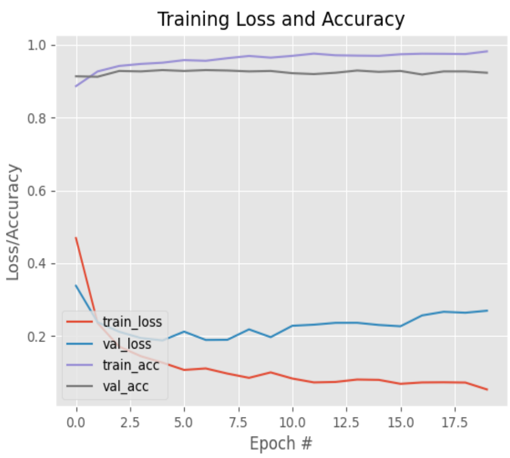

Fig. 6. Training Loss and Accuracy plot for VGG16

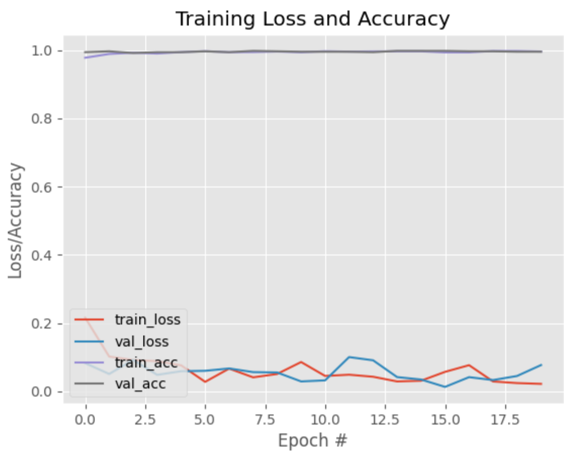

Fig. 7. Training Loss and Accuracy plot for ResNet50V2

Examining the plots for training loss and accuracy indicates that the plot for MobileNetV2 demonstrates an excellent fit. In Figure 5, the training and validation loss consistently decreases and stabilizes at a specific point, suggesting that the model is well-balanced and does not suffer from overfitting or underfitting.

Conversely, in the case of ResNet50V2, the plot exhibits signs of overfitting. Figure 7 demonstrates that the validation loss surpasses the training loss, implying that the model struggles to generalize with new data. Although it performs well with the training data, it falters when confronted with new data in the validation set. Consequently, this study selects MobileNetV2 as the preferred option for real-time video face mask detection due to its well-fitting loss and accuracy plot compared to the overfitting observed in the ResNet50V2 plot. This distinction renders it the more suitable choice for mask detection.

The training loss, indicated as train_loss in Table 4, depicts the error on the training dataset. On the other hand, the validation loss, or val_loss, represents the error observed when the validation dataset is processed through the trained network. As explained in [30], the train_acc, or training accuracy, reflects the model's accuracy on the examples it was initially built upon. In contrast, val_acc measures the quality of the model's predictions on new data.

Table 4
Training loss and accuracy results

Method	Train_loss	Train_acc	Val_loss	Val_acc
MobileNetV2	0.0389	0.9877	0.0385	0.9878
VGG16	0.0530	0.9818	0.2693	0.9231
ResNet50V2	0.0289	0.9954	0.0959	0.9902

Table 5 reveals that the training time for MobileNetV2 on the classification dataset is notably faster than VGG16 and ResNet50V2. MobileNetV2 achieved this task in just 28 minutes and 47 seconds, in stark contrast to VGG16's 4 hours and 38 seconds and ResNet50V2's 1 hour, 27 minutes, and 11 seconds. These results indicate that MobileNetV2 is the optimal choice, particularly for efficiently handling image classification tasks involving substantial datasets like the author's 4,095-image dataset. Consequently, MobileNetV2 emerges as the ideal algorithm for the subsequent phase, which involves real-time face mask detection.

Table 5
Training Time of Pre-trained Methods Used

Method	Training Time
MobileNetV2	00 hours:28 minutes:47 seconds
VGG16	04 hours:00 minutes:38 seconds
ResNet50V2	01 hours:27 minutes:11 seconds

4.3 Detection on Static Image for the Three Models

Considering the classification results, training loss/accuracy plots, training time, and static image detection, MobileNetV2 is the top choice for the next phase—real-time video face mask detection. Although ResNet50V2 shows promise with static images, its potential overfitting makes it less suitable for mask detection, whereas MobileNetV2 fits optimally. Notably, ResNet50V2 demonstrates the second-longest training time, whereas MobileNetV2 is the quickest.

4.4 Compute Differences using T-Test

MobileNetV2 outperforms VGG16 and ResNet50V2 in classifying face mask detection datasets in this project. Therefore, the study conducted a t-test to determine if a significant difference existed between MobileNetV2 and VGG16 and between MobileNetV2 and ResNet50V2.

The Table 6 shows the results of five accuracy tests from each of the three algorithms.

Table 6
Accuracy tests

Test	MobileNetV2	VGG16	ResNet50V2
1	98.77	98.18	99.54
2	98.71	97.93	99.41
3	98.92	97.87	99.41
4	98.83	97.84	99.57
5	98.67	98.15	99.60

MobileNetV2 achieved accuracy results of 98.77, 98.71, 98.92, 98.83, and 98.67 in five tests, while VGG16 scored 98.18, 97.93, 97.87, 97.84, and 98.15. ResNet50V2 consistently performed well with accuracy results of 99.54, 99.41, 99.41, 99.57, and 99.60 in the same tests. Table 7 presents the results of the t-test conducted between MobileNetV2 and VGG16, as well as between MobileNetV2 and ResNet50V2.

Table 7
T-test Results

Test	MobileNetV2	VGG16	ResNet50V2
t-statistic	9.3505	-12.1236	t-statistic
p-value	0.000013981	0.0000019827	p-value

The t-statistics for MobileNetV2 compared to VGG16 is 9.3505, while when compared to ResNet50V2, it is -12.1236, as tabulated in Table 7. The p-value for MobileNetV2 with VGG16 is 0.000013981, and with ResNet50V2, it is 0.0000019827. Therefore, since the p-value is < 0.05 in both cases, it indicates a significant difference between MobileNetV2 and VGG16 and ResNet50V2.

This study can confidently reject the null hypothesis, signifying a clear distinction between the two algorithms: MobileNetV2 compared to VGG16 and MobileNetV2 compared to ResNet50V2. MobileNetV2 is the optimal choice for the next phase, which involves applying real-time video face mask detection.

4.5 Real-Time Face Mask Wearing Detection

i. *Implementing the MobileNetV2 model in the OpenCV*
Input: Saved Model from MobileNetV2
Output: During the real-time webcam-based face mask detection process, when an individual wears a mask, a green square is displayed alongside the predicted accuracy percentage for mask-wearing. In contrast, if the person is not wearing a mask, a red square, and the expected accuracy percentage for not wearing a mask are shown.

ii. *Process Description*
Step 1: The system reads the video frame by frame and resize it as necessary for processing.

Step 2: It then invokes the preprocessing function.
Step 3: Consequently, based on the input data, the system makes predictions about whether the person is wearing a mask or not.
Step 4: Finally, it captures the results, including accuracy, using the webcam.

In Figure 8, real-time face mask detection correctly predicts individuals wearing masks with 100% accuracy. Conversely, in Figure 9, the face mask detector accurately identifies individuals as not wearing masks, achieving a 100% accuracy rate.

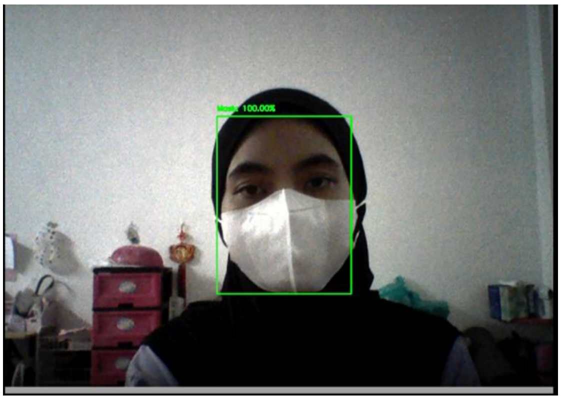

Fig. 8. Mask detected on one person

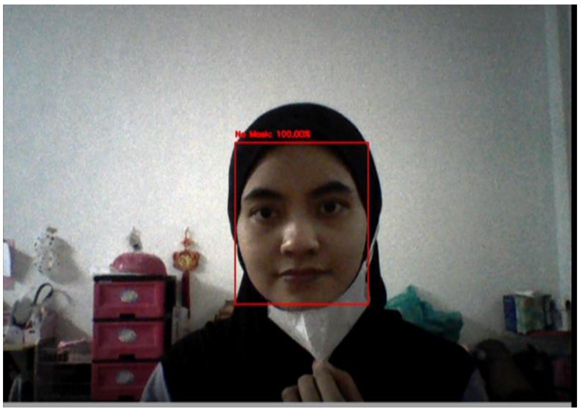

Fig. 9. No mask detected on one person

From Figure 10, real-time face mask detection accurately predicts both individuals wearing and without masks, achieving a 100% accuracy rate. In Figure 11, the face mask detector correctly identifies the two individuals wearing masks, with a 100% accuracy result.

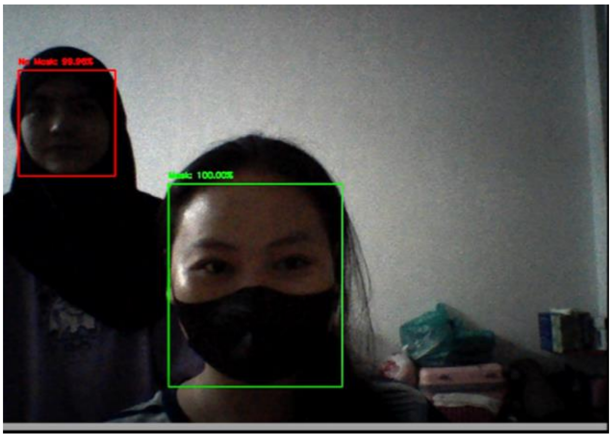

Fig. 10. Detection of two persons (with mask and no mask)

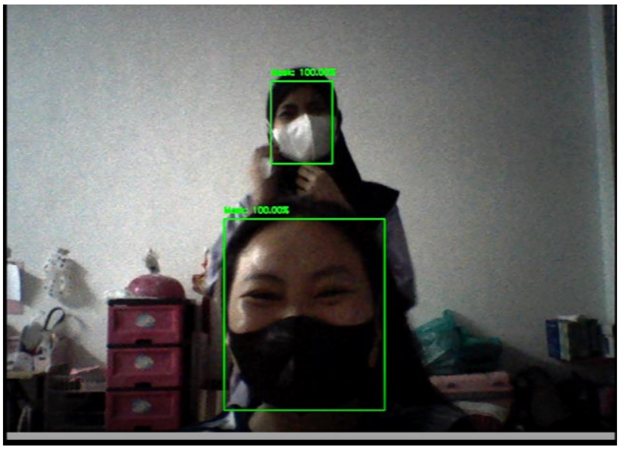

Fig. 11. Detection of two persons (both with masks)

5. Conclusions

In conclusion, this study introduced and evaluated using MobileNetV2, VGG16, and ResNet50V2 algorithms for static image and real-time face mask detection. The primary objective was to assess the effectiveness of deep learning in face mask detection. The study configured all three algorithms with identical architectures and hyperparameters. The study comprehensively covered dataset selection, data preprocessing, image classification, the utilization of pre-trained models, and the two-phase training process for the face mask detector. The evaluation relied on a classification report that included accuracy, precision, recall, and f1-score metrics. Among the algorithms, MobileNetV2 exhibited superior performance, leading to its recommendation for mask detection. The real-time mask detection results were promising, achieving 100% accuracy in correctly identifying individuals with and without masks. The approach, although informative, might only partially encompass the complexities and variations that can arise in diverse real-world environments and scenarios. As such, further research considering a broader range of environmental factors could enhance the applicability and robustness of the proposed model. For future research, diversifying the dataset to include various angles of people's faces will be crucial for improving the accuracy of face mask detection systems.

Acknowledgement
This research was not funded by any grant.

References

[1] Jie, C. Y., and N. Mat Ali. "COVID-19: What are the challenges of online learning? A literature review." *International Journal of Advanced Research in Future Ready Learning and Education* 23, no. 1 (2021): 23-29.
[2] Lauzon, Francis Quintal. "An introduction to deep learning." In *2012 11th international conference on information science, signal processing and their applications (ISSPA)*, pp. 1438-1439. IEEE, 2012. https://doi.org/10.1109/ISSPA.2012.6310529
[3] Loey, Mohamed, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and Nour Eldeen M. Khalifa. "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic." *Measurement* 167 (2021): 108288. https://doi.org/10.1016/j.measurement.2020.108288
[4] Kaur, Gagandeep, Ritesh Sinha, Puneet Kumar Tiwari, Srijan Kumar Yadav, Prabhash Pandey, Rohit Raj, Anshu Vashisth, and Manik Rakhra. "Face mask recognition system using CNN model." *Neuroscience Informatics* 2, no. 3 (2022): 100035. https://doi.org/10.1016/j.neuri.2021.100035
[5] World Health Organization. "Coronavirus disease (COVID-19) outbreak: rights, roles and responsibilities of health workers, including key considerations for occupational safety and health: interim guidance, 19 March 2020." In *Coronavirus disease (COVID-19) outbreak: rights, roles and responsibilities of health workers, including key considerations for occupational safety and health: interim guidance, 19 March 2020*. 2020.
[6] Kek, Hong Yee, Huiyi Tan, Desmond Daniel Chin Vui Sheng, Yi Lee, Nur Dayana Ismail, Muhd Suhaimi Deris, Haslinda Mohamed Kamar, and Keng Yinn Wong. "A CFD assessment on ventilation strategies in mitigating healthcare-associated infection in single patient ward." *Progress in Energy and Environment* (2023): 35-45. https://doi.org/10.37934/progee.24.1.3545
[7] Samah, Farhana Abu, Siti Zalita Talib, Nur Ainun Mokhtar, and Nurulbahiyah Ahmad Khairudin. "Molecular Docking Studies Of Potential Drugs For Covid19 Targetting On The Coronavirus Hemagglutinin Esterase." *Journal of Research in Nanoscience and Nanotechnology* 4, no. 1 (2021): 13-18.
[8] Winata, I. Made Putra Arya, Putu Emilia Dewi, Putu Brahmanda Sudarsana, and Made Sucipta. "Air-Flow Simulation in Child Respirator for Covid-19 Personal Protection Equipment Using Bamboo-Based Activated Carbon Filter." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 91, no. 1 (2022): 83-91. https://doi.org/10.37934/arfmts.91.1.8391
[9] Abd Rahman, Muhammad Faqhrurrazi, Nor Zelawati Asmuin, Ishkrizat Taib, Juwaidi Nazar, Azizan Ismail, and Riyadhthusollehan Khairulfuaad. "Investigate flow characteristics of metered-dose inhaler (MDI) disposable inhaler spacer (AeroCup) for COVID-19 patient by using computational fluid dynamic (CFD)." *CFD Letters* 12, no. 12 (2020): 63-74. https://doi.org/10.37934/cfdl.12.12.6374
[10] Amran, Muhammad Syahiran, and Khairul Nizam Tahar. "Assessment on different digital camera calibration software for photogrammetric applications." *Journal of Advanced Research in Applied Mechanics* 30, no. 1 (2017).
[11] Asaad, Renas Rajab, and Revink Masoud Abdulhakim. "The Concept of Data Mining and Knowledge Extraction Techniques." *Qubahan Academic Journal* 1, no. 2 (2021): 17-20. https://doi.org/10.48161/qaj.v1n2a43
[12] Almufti, Saman M., Ridwan B. Marqas, Zakiya A. Nayef, and Tamara S. Mohamed. "Real time face-mask detection with arduino to prevent covid-19 Spreading." *Qubahan Academic Journal* 1, no. 2 (2021): 39-46. https://doi.org/10.48161/qaj.v1n2a47
[13] Lad, Akash Manish, Anubhav Mishra, and Arul Rajagopalan. "Comparative analysis of convolutional neural network architectures for real time COVID-19 facial mask detection." In *Journal of Physics: Conference Series*, vol. 1969, no. 1, p. 012037. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1969/1/012037

[14] Hisham, Sri Dewi, Shaza Eva Mohamad, Mohd Ibrahim Shapiai, Koji Iwamoto, Aimi Alina Hussin, Norhayati Abdullah, and Fazrena Nadia Md Akhir. "Comparison of Conventional CNN Sequential API and Functional API for Microalgae Identification." *Journal of Advanced Research in Micro and Nano Engineering* 17, no. 1 (2024): 96-104. https://doi.org/10.37934/armne.17.1.96104

[15] Adhinata, Faisal Dharma, Nia Annisa Ferani Tanjung, Widi Widayat, Gracia Rizka Pasfica, and Fadlan Raka Satura. "Comparative study of VGG16 and MobileNetv2 for masked face recognition." *J. Ilm. Tek. Elektro Komput. dan Inform* 7, no. 2 (2021): 230-237. https://doi.org/10.26555/jiteki.v7i2.20758

[16] K. N, Sowmya, Rekha P. M, Trishala Kumari, and Baru Debtera. "Multistage Framework for Automatic Face Mask Detection Using Deep Learning." *Computational Intelligence and Neuroscience* 2022, no. 1 (2022): 1500047. https://doi.org/10.1155/2022/1500047

[17] Pang, Bo, Erik Nijkamp, and Ying Nian Wu. "Deep learning with tensorflow: A review." *Journal of Educational and Behavioral Statistics* 45, no. 2 (2020): 227-248. https://doi.org/10.3102/1076998619872761

[18] Arnold, Taylor B. "kerasR: R interface to the keras deep learning library." *J. Open Source Softw.* 2, no. 14 (2017): 296. https://doi.org/10.21105/joss.00296

[19] Christa, G. Harriat, J. Jesica, K. Anisha, and K. Martin Sagayam. "CNN-based mask detection system using openCV and MobileNetV2." In *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, pp. 115-119. IEEE, 2021.

[20] Petrescu, Relly Victoria, Raffaella Aversa, Antonio Apicella, and Florian Ion Petrescu. "Green Energy to Protecting the Environment (Energia Verde Para Proteger O Meio Ambiente)." *GEINTEC Journal* 7, no. 1 (2017). https://doi.org/10.7198/S2237-072220170001011

[21] Zhou, Yangling, Pejman Ghassemi, Michelle Chen, David McBride, Jon P. Casamento, T. Joshua Pfefer, and Quanzeng Wang. "Clinical evaluation of fever-screening thermography: impact of consensus guidelines and facial measurement location." *Journal of Biomedical optics* 25, no. 9 (2020): 097002. https://doi.org/10.21203/rs.3.rs-28139/v1

[22] Suresh, K., M. B. Palangappa, and S. Bhuvan. "Face mask detection by using optimistic convolutional neural network." In *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, pp. 1084-1089. IEEE, 2021. https://doi.org/10.1109/ICICT50816.2021.9358653

[23] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[24] Gopalakrishnan, Kasthurirangan, Siddhartha K. Khaitan, Alok Choudhary, and Ankit Agrawal. "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection." *Construction and building materials* 157 (2017): 322-330. https://doi.org/10.1016/j.conbuildmat.2017.09.110

[25] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016. https://doi.org/10.1109/CVPR.2016.90

[26] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510-4520. 2018. https://doi.org/10.1109/CVPR.2018.00474

[27] Yu, Donghang, Qing Xu, Haitao Guo, Chuan Zhao, Yuzhun Lin, and Daoji Li. "An efficient and lightweight convolutional neural network for remote sensing image scene classification." *Sensors* 20, no. 7 (2020): 1999. https://doi.org/10.3390/s20071999

[28] Shafiq, Muhammad, and Zhaoquan Gu. "Deep residual learning for image recognition: A survey." *Applied Sciences* 12, no. 18 (2022): 8972. https://doi.org/10.3390/app12188972

[29] Deb, C. "Face-Mask-Detection: Face Mask Detection system based on computer vision and deep learning using OpenCV and Tensorflow." https://github.com/chandrikadeb7/Face-Mask-Detection

[30] Pedro Alves. "Train accuracy vs test accuracy vs confusion matrix". https://datascience.stackexchange.com/questions/28426/train-accuracy-vs-test-accuracy-vs-confusionmatrix#:~:text=Accuracy%3A%20The%20amount%20of%20correct,examples%20it%20hasn't%20seen