Real-Time Face Mask Classification with Convolutional Neural Network for Proper and Improper Face Mask Wearing

FATIN AMANINA AZIS, HAZWANI SUHAIMI, PG EMEROYLARIFFION ABAS

Faculty of Integrated Technologies, Universiti Brunei Darussalam, Gadong, Brunei Darussalam, BE1410, Brunei Darussalam

Corresponding author: Fatin Amanina Azis (e-mail: 20h8455@ubd.edu.bn).

⋮ **ABSTRACT** Since the discovery of COVID-19, the wearing of a face mask has been recognized as an effective means of curbing the spread of most infectious respiratory diseases. A face mask must completely enclose the lips and nose properly for effective prevention of the disease. Some people still refuse to wear the mask, either out of annoyance or difficulty, or they are just wearing it incorrectly, which diminishes the mask's effectiveness and renders it worthless. The deep learning models described in this research provide a mechanism for assessing whether a face mask is being worn correctly or incorrectly using images. For both training and testing, the suggested method makes use of MaskedFace-Net dataset that contains annotated photos of an individual's face with proper and improper masks. Threshold optimizations are applied to produce significant results of prediction when comparing ResNet50, MobileNetV2 and DenseNet121 models. It is observed that better performance can be achieved with having accuracy as the target evaluation metric and reaching accuracy levels of 97.6%, 99.0%, and 99.8% for ResNet50, DenseNet121, and MobileNetV2, respectively after threshold optimization. As an outcome, DenseNet121 outperformed the other evaluated models when accuracy, recall, and precision metrics were used to assess the testing set. The face mask categorization can be used to automatically monitor face masks in real-time in public locations like hospitals, airports, shopping complexes and congested spaces to verify compliance with the published guidelines by the higher authorities in a country, making the results valuable for future use.

⋮ **KEYWORDS** Face mask; CNN; Real-Time; Infectious Respiratory Diseases; MaskedFace-Net; Image Processing.

I. INTRODUCTION

COMMUNICABLE diseases and their prevention are a highly researched area in the medical field and will continue so, as many people still continuously contract different infectious diseases throughout the world. Different viral infections were discovered a long time ago and some even exist until now, including tuberculosis, swine flu and the newly emerging infection such as monkey pox. Thus, it is of vital importance that science keeps up to date with these different diseases. This is especially true for COVID-19, with its different variations, high transmission rate and severity [1]. COVID-19 is caused by virus, a type of microorganism that has the ability to reproduce very fast which makes it deadly contagious, since it can be easily transmitted from person-to-person through respiratory droplets carrying the virus [2]–[5]. If no preventive action is taken, the disease can spread widely very quickly. Learning from the recent global outbreak of COVID-19, it is especially important to mitigate and to formulate a solution to curb the spread of the disease. Naturally,

solution used to mitigate COVID-19 may also be used to prevent future outbreaks of other similar airborne viruses, as well as can be used as basis for future improvements and preparation for handling similar situations better.

One of the actions taken by health officials to reduce transmissions of airborne diseases, particularly, COVID-19, between individuals is through the wearing of face masks. There are different types of face mask available, with different capabilities in blocking particles that can possibly carry the virus. These include surgical and medical face masks, with different level of protections and comforts, prices, and some which are disposable one-time use masks. At the start of the COVID-19 pandemic, N95 face masks were recommended only for front-liners due to supply shortages; with KN95 and KF95 masks, having almost similar efficiencies to the N95 mask, filling the gaps, especially for public uses [6], [7]. Both surgical and respirator masks can reach filtration efficiency of 95% in blocking aerosol particles, however, surgical masks are generally cheaper and hence, more preferred for everyday use

[8]. Although the wearing of face mask has been proven to be somehow effective in controlling transmission, and its usage has even been made mandatory in high-risk areas by health officials, the effectiveness of face masks is also dependent on how it is worn. Generally, the face mask needs to entirely cover the mouth, nose and chin for it to be effective [9], with the mask providing a tight seal without any gaps. Improper wearing of a face mask may reduce the level of protection that it provides and may even deem the face mask worthless. As such, both the choice of face mask and its proper wearing are of utmost importance to ensure maximum protection can be derived.

To ensure that the public complies with the requirements for wearing face masks, and more importantly, for wearing them properly, security officers and officials perform manual monitoring at the entrance of premises and public places [1], particularly, at times when transmission rates are particularly high. Naturally, this manual monitoring is resource intensive, requiring large manpower, and may be prone to human error. An alternative, which has been put forward by scientists to address these issues, is by automating the process via the introduction of an automatic detection system using computer vision techniques. Numerous studies on the classification of face masks have recently been presented, however the research problem is still less explored. The use of smaller datasets and the rarity of direct model comparisons are a few probable explanations for the limitations in the research.

Different CNN models were studied as classifiers for the detection of improper wearing of face masks. Whilst different CNN models had been used for other applications [10]–[13], they were not sufficiently explored for the identification of improper mask wearing [14]–[17]. Residual Network (ResNet) [18] was used for image recognition task, with residual blocks used in deep residual networks to boost model precision. The core concept behind ResNet is the use of skip connections, which is present in the residual blocks [19]. It operates in two ways. First, skip connections resolve the vanishing gradient problem by creating a different path for the gradient to use [20]. Secondly, it can also learn an identity function of the model, and thereby, ensure that the model's higher levels do not function any worse than its bottom layer. Essentially, the skip connections combine the results of earlier layers with the results of stacked layers, enabling the training of far deeper networks feasible. However, due to the long time needed to train the layers, ResNet50's building block model was redesigned into a bottleneck design; whereby instead of two layers in original ResNet34, a stack of three layers bottleneck blocks was utilised [21]. ResNet models are suitable for image classifications, with some researchers using ResNet50 for face mask recognition to achieve accuracies of 89.5% [22] and 95% [23].

Another CNN models which have been used for face mask recognition are MobileNets. MobileNets are generally compact, low-power models with low-latency, and may be modified to accommodate different use cases' resource limitations. MobileNetV2 enhances the existing MobileNet models and it is a highly effective feature extractor for object detection and segmentation tasks [24]. By employing depth-wise separable convolutions as effective building blocks, MobileNetV2 expands on the concepts of MobileNetV1. The inner part of the MobileNetV2 model contains the model's capacity to change from pixels to image categories, while the bottlenecks of the network encode the intermediary inputs and outputs. MobileNetV2 processed faster than the previous MobileNets [13], with MobileNetV2 having been demonstrated for face mask detection task to give 89% [25] and 98% accuracies [26].

DenseNet121 has four average pooling layers and 120 convolutional layers. Each dense block contains a different number of layers with two convolutions apiece, a bottleneck layer with a kernel size of 1x1 and a convolution layer with a kernel size of 3x3 [27]. By applying the composite function operation, an output from the first layer serves as input to the second layer. The convolution layer, pooling layer, batch normalization, and non-linear activation layers make up this composite operation, which solves the vanishing gradient problem [28]. Several previous works also utilised DenseNet121 for face mask recognition, some of them reported a high accuracy of 98.5% [29]. Among the three classification models, it has been found that DenseNet121 reports a higher accuracy for face mask classification on large datasets, but it cannot be directly compared to the other models as they are using different datasets. Therefore, since there are no direct comparisons for the three classification models using identical large datasets which include images for improper mask wearing detections such as MaskedFace-Net dataset [30], there is a need for direct comparison between the models. Hence, in this paper, popular CNN models: ResNet50, MobileNetV2 and DenseNet121 are explored to detect proper and improper face mask wearing using MaskedFace-Net dataset.

II. METHODOLOGY

Figure 1 depicts the process adopted in this paper, which is divided into two main phases: training and testing phases. The face mask classification task is based on binary classification to identify proper and improper wearing of face mask during the outbreak of infectious respiratory diseases using facial images of people. Three different CNN models: ResNet50, MobileNetV2 and DenseNet121, have been selected for the classification task. These CNN models are trained using images from the training dataset during training phases, with the performance of the trained models compared with each other during the testing phase. At the testing phase the most effective CNN model is compared and selected for detecting face mask wearing conditions of a single person. Furthermore, the best trained model is selected, and used in real-time demonstrations using real-time images from a simple image capture setup, to demonstrate the effectiveness and simplicity of the face mask detection method.
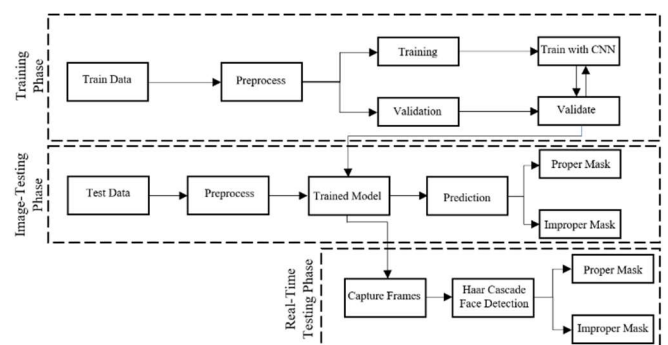

Figure 1. The block diagram of training and testing phases.

A. DATASET AND PRE-PROCESSING

A large open-source dataset, MaskedFace-Net [30], has been

used in this paper. The dataset contains clear facial images of a single person with a medical face mask categorised into properly worn (Correctly Masked Face Dataset) and improperly worn (Incorrectly Masked Face Dataset) face masks. It is the largest dataset for the face mask classification task, specifically curated during the first wave of COVID-19, containing 133,783 images. Before feeding into the models, input images need to be pre-processed to make them more suitable as input to the CNN models. The images are first scaled to a chosen pixel size, which can significantly crop the original images. The images are then, augmented by rotating, flipping, sharpening, blurring, and transforming colors of the images to gray scales resulting in different boundaries for edge detection.

For real-time testing, real-time images were taken from a web camera, which capture images at a rate of 30 frames per second. OpenCV and Tensorflow Keras Python modules were used for pre-processing of these images. Every frame was fed into a pre-trained Haar Cascade face detection, which uses a light object detection algorithm to identify human faces in a real-time video irrespective of their scales and locations. The Haar Cascade face detection can detect multiple faces present in a single frame hierarchically and time-efficiently [31]. Face regions were searched within a cluster of pixels, and for every detected face, the face region was cropped. A scanning operation was then performed in a cascading manner to check the eye regions on the detected face region, to obtain reference coordinates of the eyes, which were then used to draw a bounding box around the face. These bounded face regions were then passed to the pre-trained model for prediction. It is noted that the Haar Cascade classifier is not required as a pre-processing task on the MaskedFace-Net dataset, as the dataset already provides cropped face regions.

B. CLASSIFICATION METHODS

Following pre-processing, the augmented images were then passed onto the selected classification models to allow the models to learn the parameters and characteristics of the augmented data for the classification task. The pre-processed MaskedFace-Net dataset were split into training and testing datasets, with the augmented training dataset split further into training and validation sets. Validation is required to ensure proper training at every epoch as well as to reduce overfitting issues. The training phase outputted a trained CNN model, which were then used during the testing phase, both in the image-testing phase and real-time testing phase. Identical data splits were used to train and test all three CNN models, to ensure fair comparison between the models.

Three CNN models are used in this work: ResNet50 [32], MobileNetV2 [14] and DenseNet121 [33]. All three models are used for face mask classification tasks; however, no direct comparison has been previously performed on the three popular CNN models directly. Additionally, comparatively small datasets were used in previous research works. ResNet50 is a 50-layer network that stacks residual blocks on top of another to back-propagate the gradient during training. Among the three models, ResNet50 has the largest number of parameters, with over 25.6 million trainable parameters with a training size of 98MB. This is followed by DenseNet121, a highly dense 121-layer network with 8.1 million trainable parameters and a training size of 33MB. MobileNetV2, which is a 53-layer network composed of fully convolutional layer and residual bottleneck layers, is the lightest among the models. It has about 3.5 million trainable parameters with about 14MB training size.

Table 1 shows the summary of the model layers, parameters, and sizes.

Table 1. Summary of model layers, parameters, and sizes

Model	No. of layers	No. of parameters	Model size
ResNet50	50	25.6 million	98MB
DenseNet121	121	8.1 million	33MB
MobileNetV2	53	3.5 million	14MB

Threshold optimization, by finding the optimum probability thresholds that give the highest prediction accuracy, had also been performed. For binary classification, probability threshold $p_{Th} = 0.5$ represents the default probability threshold, with the model classifying the input data as positive or negative depending on whether they are above or below the threshold. These are then used to assign predicted class label to the input data. Threshold optimization seeks to find an alternative threshold for the face mask classification such that a given performance measure $P_{measure}$ is optimised, during the training phase. This performance measure can be accuracy, recall, precision, Area Under Curve-Receiver Operating Characteristics (AUC-ROC) and F1-score.

Mathematically, the optimisation problem can be expressed as

$$\max_{p_{Th}} P_{measure} \qquad (1)$$

Subject to:

$$0 \leq p_{Th} \leq 1$$

C. PERFORMANCE METRICS

To evaluate the performance of the model for the dataset, different metrics have been used to select the best performing model. The performance metrics which have been applied in this work include accuracy, precision, and recall. As the dataset is a balanced dataset, accuracy is suitable to be used as one of the evaluation metrics. Accuracy quantifies the ratio of the number of correct predictions to the total number of predictions. On the other hand, precision measures how exact the model is in identifying a positive class correctly, whilst recall evaluates how exact the model is in predicting all the positive class in the dataset. To calculate the metrics, confusion matrix is used to easily identify four elements that can be used to calculate the evaluation metrics encompassing true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). True positive is known as the number of positive outcomes correctly predicted as positive by the model. False positive is the number of negative outcomes incorrectly predicted as negative by the model and true negative is the number of negative outcomes the model incorrectly predicts as negative. Based on the four elements, the formulae are shown below to calculate respective performance metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

$$Precision = \frac{TP}{TP + FP} \qquad (4)$$

Machine learning algorithms commonly predict the likelihood or probability of the input data belonging to the different class label, before predicting the precise class label based on the probabilities. Given probabilities of the input data

belonging to the different class labels, a probability threshold controls the choice of whether the input data is classified as positive or negative. For normal projected probabilities or scores ranging from 0 to 1, default probability threshold is 0.5 [34]. In a binary classification problem such as the face mask classification with the class labels 0 and 1 and a default probability threshold of 0.5, values below the threshold are assigned to class 0, and values above or equal to the threshold are assigned to class 1. This threshold can be altered to optimize the prediction of classes, which may give different predictions for the same probabilities. As a result, the performance metrics, including accuracy, precision, recall, Area Under Curve-Receiver Operating Characteristics (AUC-ROC) and F1-score may also vary. Threshold optimizations were integrated in different studies [34]–[37] to improve the prediction performance by changing the threshold iteratively to find the optimal threshold which gives the highest performance measures.

III. RESULTS AND DISCUSSION

The three CNN classification models have been assessed on the large open-source MaskedFace-Net [30] dataset, containing facial images of persons wearing medical masks correctly and incorrectly. The dataset consists of two sub-datasets: 1) the Incorrectly Masked Face Dataset (IMFD) dataset, containing images of incorrectly worn face mask, and 2) the Correctly Masked Face Dataset (CMFD), containing images of correctly worn face mask. There are 133,783 images in the collection, which are then reduced to 119,400 images to evenly distribute between the CMFD and IMFD datasets. The images were first pre-processed, which involved resizing the images to 150x150 pixels in RGB color space, image alterations, and data augmentation techniques such as rescaling, shearing, zooming, and flipping. These were carried out to maintain consistency, enhance quality, and increase the dataset's size by giving the models a wide variety of images for the training. Table 2 summarizes the experiment's specifications and parameters.

The MaskedFace-Net dataset was randomly split 90%:10%% for training and testing, respectively; to give 109,380 random training and 11,940 random testing datasets. The training dataset was further randomly split into 85,580 and 23,800 images for training and validation during the training process, respectively. Performances of the classification models were then measured using the testing dataset. Figure 2 depicts the dataset allocation of training, validation, and testing along with illustrative images from each class.
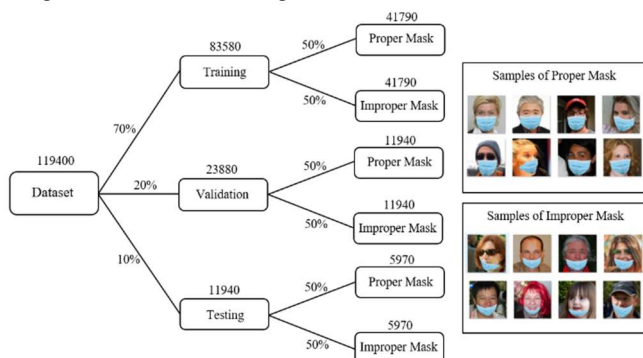

Figure 2. Dataset allocation

Table 2. Training machine details and training parameters' values

Parameter	Detail/ Value
Processor	AMD Ryzan Threadripper 3960X 24-Core Processor, 3793MHz
RAM	64 GB
System Type	x64-based pc
Total Training Data	83,580
Total Validation Data	23,880
Input Shape	150 x 150
Optimizer	Adam
Learning Rate	0.001
Dropout	0.4
Momentum	0.9
Epoch	100
Batch Size	32

Figure 3 gives the confusion matrices of the trained ResNet50, MobileNetV2, and DenseNet121 CNN models using the default probability threshold of 0.5 i.e. $p_{Th} = 0.5$. The trained ResNet50, MobileNetV2 and DenseNet121 were able to accurately predict 5,904 images, 5,929 images, and 5,954 images of the properly worn face mask images, respectively. These are 98.9%, 99.3%, and 99.7% of the total 5,970 properly worn face mask images, respectively. On the other hand, the trained ResNet50, DenseNet121, and MobileNetV2 were able to accurately predict 5,714 images (95.7%), 5,887 images (98.6%) and 5,955 images (99.7%) from the total 5,970 improperly worn face mask images, respectively.
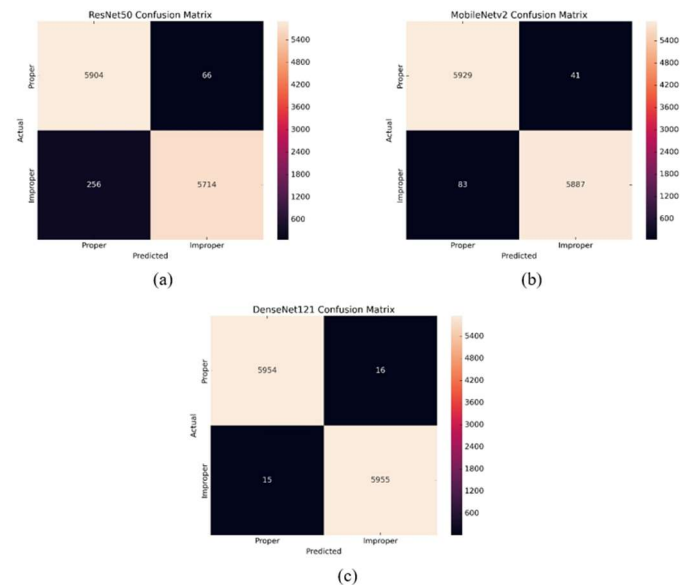

Figure 3. Confusion matrices for default threshold of (a) ResNet50; (b) MobileNetv2; and (c) DenseNet121 CNN models.

The thresholds are then changed iteratively using an increment of 0.01 to find the optimum prediction of results for both proper and improper wearing of face masks, based on the optimisation in equation (1). Accuracy, precision, recall, AUC-ROC and F1-score are chosen as the performance measures $P_{measure}$ which have to be optimised. Table 3, 4 and 5 show the performance evaluation metrics for the three models after the optimisation.

Table 3. Threshold optimizations for ResNet50

ResNet50						
Target Metric	p_{Th}	Resulting Accuracy	Resulting Precision	Resulting Recall	Resulting AUC-ROC	Resulting F1-Score
Accuracy	0.40	0.976	0.980	0.972	0.976	0.976
Precision	0.75	0.960	0. 996	0.923	0.960	0.958
Recall	0.10	0.959	0.928	0.994	0.959	0.960
AUC-ROC	0.40	0.976	0.980	0.972	0.976	0.976
F1-Score	0.35	0.976	0.976	0.976	0.976	0.976

Table 4. Threshold optimizations for MobileNetV2

MobileNetV2						
Target Metric	p_{Th}	Resulting Accuracy	Resulting Precision	Resulting Recall	Resulting AUC-ROC	Resulting F1-Score
Accuracy	0.35	0.990	0.991	0.989	0.990	0.990
Precision	0.75	0.988	0.996	0.979	0.988	0.988
Recall	0.10	0.986	0.978	0.994	0.986	0.986
AUC-ROC	0.35	0.990	0.991	0.989	0.990	0.990
F1-Score	0.35	0.990	0.991	0.989	0.990	0.990

Table 5. Threshold optimizations for DenseNet121

DenseNet121						
Target Metric	p_{Th}	Resulting Accuracy	Resulting Precision	Resulting Recall	Resulting AUC-ROC	Resulting F1-Score
Accuracy	0.40	0.998	0.997	0.999	0.998	0.998
Precision	0.75	0.997	0.999	0.996	0.997	0.997
Recall	0.10	0.994	0.990	0.999	0.994	0.994
AUC-ROC	0.40	0.998	0.997	0.998	0.998	0.998
F1-Score	0.40	0.998	0.997	0.999	0.998	0.998

It can be observed from the results that choosing accuracy, AUC-ROC and F1-score as target performance measures $P_{measure}$ gives identical and best-performing outcomes. The optimum probability thresholds, p_{Th} for ResNet50 are 0.40, 0.35, 0.75 and 0.1 with accuracy and AUC-ROC; F1-score, precision and recall as target performance measures, respectively. These are similar for DenseNet121, with the exception of F1-score as target performance measure. On the other hand, MobileNetV2 utilises a slightly lower probability threshold of $p_{Th} = 0.35$ with accuracy, AUC-ROC and F1-score as target performance measures, with probability threshold p_{Th} of 0.75 and 0.1 for precision and recall as target performance measures, respectively. Comparing between the 3 models, it can be summarized that DenseNet121 outperforms both ResNet50 and MobileNetV2 with optimum accuracy of 99.8% for DenseNet121 followed by MobileNetV2 with 99.0% and ResNet50 with 97.6%. $p_{Th} = 0.4$ in DenseNet121 also results in a balanced and better performance, with other performance metrics giving above 97% performance, including precision with 99.7%, recall with 99.9%, AUC-ROC with 99.8% and F1-score with 99.8%.

Figure 4 shows the confusion matrices for the best performing ResNet50, MobileNetV2 and DenseNet121 after threshold optimizations with accuracy set as the target criteria. Probability thresholds are 0.4, and 0.35 and 0.4 for ResNet50, MobileNetV2 and DenseNet121, respectively. In Figure 4, it is shown that ResNet50, MobileNetV2, and DenseNet121 are able to predict images of proper and improper mask wearing more effectively than using the default threshold value in Figure 3. ResNet50 mistakenly predicted many images of

proper and improper mask wearing. Further research into the inaccurately predicted images reveals that images of people without covering the chin are more likely to be misinterpreted as proper mask wearing. In the future, more images of persons with exposed chins may be used to train the classifiers, to improve the model's ability to recognize these kinds of images.
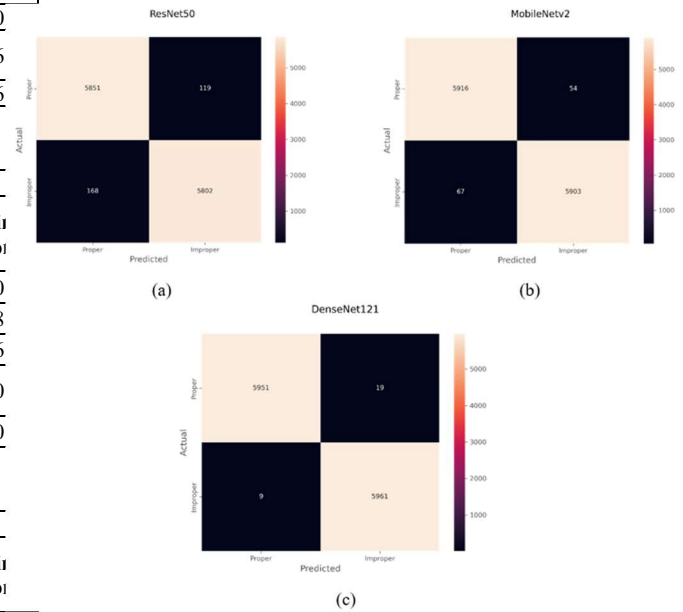

Figure 4. Confusion matrices after threshold optimizations of (a) ResNet50; (b) MobileNetv2; and (c) DenseNet121 CNN models.

The face mask classification may be used by enforcement agencies to ensure compliance to the face mask requirement, especially in an enclosed space. Emphasis may also be laid more on precision and recall values. Using a model with a high recall value is crucial during the apex of an outbreak of infectious respiratory disease, because wearing a face mask is required to stop the spread of the disease. This is especially true if it is the only path to monitor whether people are wearing face masks. Low recall value can cause a sense of insecurity. It appears everyone is correctly adhering to the face mask guidelines, although some individuals may be doing so purposefully or unintentionally improperly. This sense of insecurity from incorrect predictions might hinder the slowing down of disease transmission. On the other side, a low recall value demands extra authority confirmation. For example, the 9 inaccurately predicted images of people who were, in fact, appropriately wearing their face mask would require additional, superfluous validations using the DenseNet121 model. As a result, a high recall model is typically employed when the cost of additional validation is significant. The DenseNet121 model gives precision of 99.7%, with 5,961 out of 5,970 images for improper face mask wearing correctly classified, which is higher than ResNet50 and MobileNetV2. It also gives a high recall of 99.9%. These suggest that DenseNet121 is less likely to anticipate an improper mask wearing image incorrectly than a proper mask wearing image.

The best performing model, DenseNet121 is then employed for real-time demonstration. Figure 5 and Figure 6 show samples of some of the real-time predictions of proper and improper mask wearing, respectively. The results were captured via a live web camera. For every face detected by the

Haar Cascade function, a red bounding box was drawn to scale with the face size in the image. The classification class result is shown in green for properly worn face mask, which covers the mouth, nose and chin. On the other hand, the classification class result is shown in red for improperly worn face mask which exposed any of the mouth, nose or chin. It has been shown that the model can successfully detect facial features of a person and identify the state of the face mask wearing with clear labels can be seen instantly.
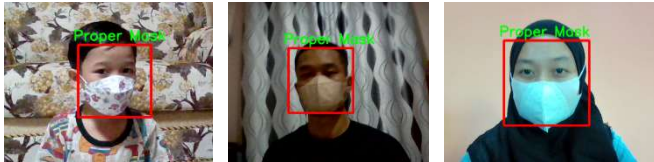

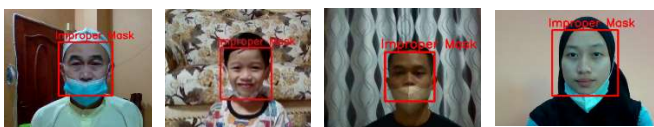Figure 5. Real-time predictions of proper face mask wearing.

Figure 6. Real-time predictions of improper face mask wearing.

IV. CONCLUSION

In order to stop the transmission of any contagious diseases, wearing a face mask has become essential. Numerous countries have made the use of face masks mandatory, especially in enclosed public spaces where there is a significant danger of infection. However, regulation has proven to be exceedingly challenging and labor-intensive. By analyzing three prominent classification models ResNet50, DenseNet121, and MobileNetV2 and the MaskedFace-Net dataset, this research suggests the use of images to support the monitoring of appropriate face mask use during the occurrence of infectious diseases in the future. A valid comparison of the classification models' results in terms of accuracy, precision, and recall becomes possible using the same training and testing datasets. The results after threshold optimization are found significant and taking accuracy as the target evaluation metric for thisresults in a better performance. DenseNet121 provided the highest overall accuracy of 99.8% with an optimum threshold of 0.40, providing 99.7% and 99.9% for precision and recall. DenseNet121 is ideal when the authority wants to ensure that people are wearing their face masks to prevent the transmission of any contagious diseases. The findings indicate that the categorization models are helpful in identifying people who are not appropriately wearing their face masks, and as such, they may be utilized during any infectious respiratory disease outbreak to guarantee adherence to any established rules by the authorities. To enable live and quick detection of people not wearing their face mask in public settings, such as hospitals, airports, and crowded premises which shall be positioned particularly at the entrance of a premises, the approach may be implemented on a real-time video surveillance system for future work. This work is significant to assure compliance and so restrict the spread of disease, and not limited to COVID-19 only but rather to all air-borne diseases.

References

[1] S. M. Jung *et al.*, "Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: Inference using exported cases," *Journal of Clinical Medicine*, vol. 9, no. 2, pp. 1–10, 2020, https://doi.org/10.3390/jcm9020523.

[2] E. A. Nardell, "Transmission and institutional infection control of tuberculosis," *Cold Spring Harbor Perspectives in Medicine*, vol. 6, no. 2, pp. 1–12, 2016, https://doi.org/10.1101/cshperspect.a018192.

[3] R. Zhang, Y. Li, A. L. Zhang, Y. Wang, and M. J. Molina, "Identifying airborne transmission as the dominant route for the spread of COVID-19," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 26, pp. 14857–14863, 2020, https://doi.org/10.1073/pnas.2009637117.

[4] K. Brown and P. A. Leggat, "Human monkeypox: Current state of knowledge and implications for the future," *Tropical Medicine and Infectious Disease*, vol. 1, no. 1, pp. 1–13, 2016, https://doi.org/10.3390/tropicalmed1010008.

[5] V. Singh and M. Sood, "Swine flu - A comprehensive view," *International Journal of Advancements in Research & Technology*, vol. 1, no. 2, pp.1-5, 2012.

[6] C. Matuschek *et al.*, "Face masks: Benefits and risks during the COVID-19 crisis," *European Journal of Medical Research*, vol. 25, no. 1, pp. 1–8, 2020, https://doi.org/10.1186/s40001-020-00430-5.

[7] M. Liao *et al.*, "A technical review of face mask wearing in preventing respiratory COVID-19 transmission," *Current Opinion in Colloid and Interface Science*, vol. 52, p. 101417, 2021, https://doi.org/10.1016/j.cocis.2021.101417.

[8] T. U. Rashid, S. Sharmeen, and S. Biswas, "Effectiveness of N95 masks against SARS-CoV-2: Performance efficiency, concerns, and future directions," *ACS Chemical Health & Safety*, vol. 29, no. 2, pp. 135–164, 2022, https://doi.org/10.1021/acs.chas.1c00016.

[9] L. Y. K. Lee *et al.*, "Practice and technique of using face mask amongst adults in the community: A cross-sectional descriptive study," *BMC Public Health*, vol. 20, no. 1, pp. 1–11, 2020, https://doi.org/10.1186/s12889-020-09087-5.

[10] F. A. Azis, H. Suhaimi, and E. Abas, "Waste classification using convolutional neural network," *ACM International Conference Proceeding Series*, pp. 9–13, 2020, https://doi.org/10.1145/3417473.3417474.

[11] M. A. Humayun, H. Yassin, and P. E. Abas, "Native language identification for Indian-speakers by an ensemble of phoneme-specific, and text-independent convolutions," *Speech Communication*, vol. 139, no. February 2021, pp. 92–101, 2022, https://doi.org/10.1016/j.specom.2022.03.007.

[12] Y. Zhang, J. Gao, and H. Zhou, "Breeds classification with deep convolutional neural network," *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, pp. 145–151, 2020, https://doi.org/10.1145/3383972.3383975.

[13] H. Nguyen, "Fast object detection framework based on Mobilenetv2 architecture and enhanced feature pyramid," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 5, pp. 812–824, 2020.

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

[15] S. Susanto, F. A. Putra, R. Analia, and I. K. L. N. Suciningtyas, "The face mask detection for preventing the spread of COVID-19 at politeknik negeri batam," *Proceedings of the ICAE 2020 3rd International Conference on Applied Engineering*, pp. 1–5, 2020, https://doi.org/10.1109/ICAE50557.2020.9350556.

[16] A. Negi, K. Kumar, P. Chauhan, and R. S. Rajput, "Deep neural architecture for face mask detection on simulated masked face dataset against covid-19 pandemic," *Proceedings of the IEEE 2021 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS'2021*, pp. 595–600, 2021, https://doi.org/10.1109/ICCCIS51004.2021.9397196.

[17] S. E. Snyder and G. Husari, "Thor: A deep learning approach for face mask detection to prevent the COVID-19 pandemic," *Conference Proceedings - IEEE SOUTHEASTCON*, vol. 2021, March 2021, pp. 1-8. https://doi.org/10.1109/SoutheastCon45413.2021.9401874.

[18] J. S. Kaiming He, Xiangyu Zhang, Shaoqing Ren, "Deep residual learning for image recognition kaiming," *Indian Journal of Chemistry - Section B Organic and Medicinal Chemistry*, vol. 45, no. 8, pp. 1951–1954, 2006, https://doi.org/10.1002/chin.200650130.

[19] K. E. Tokarev, V. M. Zotov, V. N. Khavronina, and O. V. Rodionova, "Convolutional neural network of deep learning in computer vision and image classification problems," *IOP Conference Series: Earth and Environmental Science*, vol. 786, no. 1, pp. 1-4, 2021,

https://doi.org/10.1088/1755-1315/786/1/012040.

[20] I. Kwon, G. Jo, and K. S. Shin, "A deep neural network based on resnet for predicting solutions of Poisson–Boltzmann equation," *Electronics (Switzerland)*, vol. 10, no. 21, pp. 1–12, 2021, https://doi.org/10.3390/electronics10212627.

[21] H. Zhang *et al.*, "Deep learning model for the automated detection and histopathological prediction of meningioma," *Neuroinformatics*, vol. 19, no. 3, pp. 393–402, 2021, https://doi.org/10.1007/s12021-020-09492-6.

[22] W. Hariri, "Efficient masked face recognition method during the COVID-19 pandemic," *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 605–612, 2022, https://doi.org/10.1007/s11760-021-02050-w.

[23] S. N. Yahya, A. F. Ramli, M. N. Nordin, H. Basarudin, and M. A. Abu, "Comparison of convolutional neural network architectures for face mask detection," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, pp. 667–677, 2021, https://doi.org/10.14569/IJACSA.2021.0121283.

[24] B. Khasoggi, Ermatita, and Samsuryadi, "Efficient mobilenet architecture as image recognition on mobile and embedded devices," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 1, pp. 389–394, 2019, https://doi.org/10.11591/ijeecs.v16.i1.pp389-394.

[25] S. Habib, M. Alsanea, M. Aloraini, H. S. Al-Rawashdeh, M. Islam, and S. Khan, "An efficient and effective deep learning-based model for real-time face mask detection," *Sensors*, vol. 22, no. 7, pp. 1–13, 2022, https://doi.org/10.3390/s22072602.

[26] A. K. Sharadhi, V. Gururaj, S. P. Shankar, M. S. Supriya, and N. S. Chogule, "Face mask recogniser using image processing and computer vision approach," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 67–73, 2022, https://doi.org/10.1016/j.gltp.2022.04.016.

[27] K. K. Bressem, L. C. Adams, C. Erxleben, B. Hamm, S. M. Niehues, and J. L. Vahldiek, "Comparing different deep learning architectures for classification of chest radiographs," *Scientific Reports*, vol. 10, no. 1, pp. 1–16, 2020, https://doi.org/10.1038/s41598-020-70479-z.

[28] N. Hasan, Y. Bao, A. Shawon, and Y. Huang, "DenseNet convolutional neural networks application for predicting COVID-19 using CT image," *SN Computer Science*, vol. 2, no. 5, pp. 1-11, 2021, https://doi.org/10.1007/s42979-021-00782-7.

[29] B. Batagelj, P. Peer, V. Štruc, and S. Dobrišek, "How to correctly detect face-masks for COVID-19 from visual information?," *Applied Sciences*, vol. 11, no. 5, p. 2070, 2021, https://doi.org/10.3390/app11052070.

[30] 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, p. 100144, 2021, https://doi.org/10.1016/j.smhl.2020.100144.

[31] R. Senthamizh Selvi, D. Sivakumar, J. S. Sandhya, S. Siva Sowmiya, S. Ramya, and S. Kanaga Suba Raja, "Face recognition using Haar - Cascade classifier for criminal identification," *International Journal of Recent Technology and Engineering*, vol. 7, no. 6, pp. 1871–1876, 2019.

[32] X. Feng, X. Gao, and L. Luo, "A resnet50-based method for classifying surface defects in hot-rolled strip steel," *Mathematics*, vol. 9, no. 19, pp. 1–15, 2021, https://doi.org/10.3390/math9192359.

[33] H. H. Hoang and H. H. Trinh, "Improvement for convolutional neural networks in image classification using long skip connection," *Applied Sciences*, vol. 11, no. 5, p. 2092, 2021, https://doi.org/10.3390/app11052092.

[34] C. Ferri, J. Hernández-Orallo, and P. Flach, "Setting decision thresholds when operating conditions are uncertain," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 805–847, 2019,

[35] R. E. Fan and C. J. Lin, "A study on threshold selection for multi-label classification," *Department of Computer Science, National Taiwan University*, pp. 1–23, 2007, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.66.1611&rep=rep1&type=pdf.

[36] A. Monahov, "Model evaluation with weighted threshold optimization (and the 'mewto' R package)," *SSRN Electronic Journal*, no. March, pp. 1-11, 2021, https://doi.org/10.2139/ssrn.3805911.

[37] E. Robles, F. Zaidouni, A. Mavromoustaki, and P. Refael, "Threshold optimization in multiple binary classifiers for extreme rare events using predicted positive data," *CEUR Workshop Proceedings*, vol. 2600, pp. 1-12, 2020.

FATIN AMANINA AZIS *received her B.Eng in General Engineering with concentration in Information and Systems Engineering. She is currently a PhD student in Systems Engineering at Universiti Brunei Darussalam, Brunei. Her research interests are data science, machine learning, waste management, green technologies, and sustainability.*

HAZWANI SUHAIMI *received her B.Eng in Chemical Engineering, MSc in Advanced Chemical Engineering with IT and management and Ph.D in Chemical Engineering from Loughborough University. She is currently a lecturer in Chemical and Process Engineering at Universiti Brunei Darussalam. Her research areas are tissue engineering, membrane science, ankle injury, rehabilitation, sustainability, computer modeling and simulation.*

PG EMEROYLARIFFION ABAS *received his B.Eng in Information Systems Engineering and received his Ph.D in Communication Systems from Imperial College, London. He is now working as a Senior Assistant Professor in System Engineering, Faculty of Integrated Technologies, Universiti Brunei Darussalam. His present research interests are data analysis, IoT and Sensor Network, photonics, and energy system.*