

A COMPARATIVE ANALYSIS OF MACHINE LEARNING APPROACHES FOR AUTOMATED FACE MASK DETECTION DURING COVID-19

Junaed Younus Khan* & Md Abdullah Al Alamin*

junaedyounus.khan@ucalgary.ca, mdabdullahal.alamin@ucalgary.ca
University of Calgary

Abstract— The World Health Organization (WHO) has recommended wearing face masks as one of the most effective measures to prevent COVID-19 transmission. In many countries, it is now mandatory to wear face masks, specially in public places. Since manual monitoring of face masks is often infeasible in the middle of the crowd, automatic detection can be beneficial. To facilitate that, we explored a number of deep learning models (i.e., VGG1, VGG19, ResNet50) for face-mask detection and evaluated them on two benchmark datasets. We also evaluated transfer learning (i.e., VGG19, ResNet50 pre-trained on ImageNet) in this context. We find that while the performances of all the models are quite good, transfer learning models achieve the best performance. Transfer learning improves the performance by 0.10%–0.40% with 30% less training time. Our experiment also shows these high-performing models are not quite robust for real-world cases where the test dataset comes from a different distribution. Without any fine-tuning, the performance of these models drops by 47% in cross-domain settings.

1. INTRODUCTION

The novel coronavirus disease (COVID-19) has created a global health crisis taking 3.7M lives and affecting ~172M people. This is a highly contagious disease (more infectious than influenza and Ebola [7, 8]) and mostly transmitted through respiratory droplets. Hence, several precautionary measures are recommended to prevent its transmission e.g., social distancing, wearing mask, cleaning hands. Although vaccines are available, the worldwide mass vaccination is not yet done. It is also possible to be affected even after getting vaccinated since vaccines are not 100% effective. While vaccinated people are less likely to have severe complications after getting COVID-19, a few exceptions are not unheard of¹. Moreover, due to genetic mutation, the virus is continuously evolving. The most recent strain “Omicron” has

already become a great concern for global health. Hence, precautionary measures are still crucial. In fact, wearing mask is considered the most effective measure according to WHO. Several studies ask for mass mask wearing since they found it to be highly beneficial to prevent COVID-19 [5, 6].

It is found that crowded places spread the virus most [1]. In many countries, government has made it mandatory to wear face mask in public places and public transportation. However, many people still show reluctance towards wearing masks due to negligence and lack of awareness². Hence, it is necessary to locate unmasked people, make them aware and ensure mask wearing. Since, it is practically infeasible do it manually in the middle of a crowd or a public place, automated detection of face mask is probably the only way for this, and therefore currently receiving great attention from the research community [9, 2, 4, 12, 14].

Most existing works in this direction suffer from any of these two limitations- **(1) Lack of Models Explored.** The number of evaluated models or techniques is limited, **(2) Lack of Datasets Used.** They did not evaluate the models on different dataset. To fill this gap, we explored a variety of models from three different perspectives (i.e., Baseline 1VGG, well-known VGG19 & ResNet50, and transfer learning). We also evaluated the concept of domain adaptation in this context. In particular, we addressed two research questions.

RQ1. How accurately different models can detect face mask? We compared the performance of different models i.e., 1VGG, VGG19, ResNet50, and transfer learning (VGG19 & ResNet50 pretrained on ImageNet) in mask detection task using two different datasets (Simulated and Real). It will help us to have a complete picture of the possibilities and challenges of automated face mask detection. In our study, the best performing model is ResNet50 with transfer learning which achieves 100% & 99.45% accuracy in simulated & real dataset respectively.

RQ2. How effective the concept of domain adaptation is in face mask detection? We evaluated domain adaptation by using simulated and real images as source and target domain respectively to check model transferability from simu-

*The authors contribute equally to this paper.

¹<https://globalnews.ca/news/8279892/covid-deaths-fully-vaccinated/>

²<https://www.bbc.com/news/health-58979626>

lated to real in this context. One motivation of doing so is to explore the performance of the models that are trained on simulated data but later deployed/tested on a different (real) dataset. We observed that currently there is a shortage of real images for incorrectly worn mask in the existing benchmark datasets (most of them are simulated). Hence, this RQ will help us to evaluate the feasibility of detecting incorrectly masked images in real world using simulated data. Our experiment shows that model performance severely drops (by 47%) in such cross domain setting. However, the performance quickly improves when the trained model is fine-tuned using real data. This signifies that models that are trained on simulated dataset may not perform well on target real world dataset without some fine-tuning.

We believe that our comparative analysis will help the research community to explore further in this context and to select the most appropriate mask detection method.

Replication Package. Our code is shared at <https://github.com/JunaedYounusKhan51/FaceMaskDetection>

2. DATASET

We studied two datasets that are used for various other studies [4, 12].

2.1. Simulated Facemask dataset (SFMD)

The dataset is published by Cabani et al. [3]. It contains high resolution (1024×1024) 67K simulated samples of people wearing mask correctly and incorrectly. The original images are collected from Flickr dataset [10]. For our study, we randomly selected 6,442 masked images from SFMD and 6,442 unmasked images from the Flickr dataset (Table 1). Some examples are available in Figure 1.

2.2. Real Facemask dataset (RFMD)

This is the biggest dataset of masked face for this problem [17]. It consists of 90K images of 525 different people without masks and 6,442 images of real masked people. Among the masked images, many are from the same 525 people and some are different. Figure 2 shows some samples from this dataset. To keep a balanced dataset, we used 6,442 masked and 6,442 unmasked images as described in Table 1.

2.3. Train, Test and Validation split

As mentioned, each of our datasets now contain 12,884 images (6,442 masked and 6,442 unmasked). We split the dataset into train and test set maintaining 90:10 ratio. We further divided the train portion to prepare train and validation set in 88:12 ratio. Hence, for both dataset, 12,884 images were divided into 10,203 train, 1,291 validation, and 1,288 test images.

Table 1. Overview of the facemask datasets

Dataset	Original		Our Study	
	With Mask	Without Mask	With Mask	Without Mask
Real Facemask	6442	90000	6442	6442
Simulated Facemask	67049	65534	6442	6442



Fig. 1. Simulated Masked Face Dataset (SMFD) Samples

3. EXPERIMENTAL SETUP

In this section, we describe our experimental setup. An overview of our methodology is showed in Figure 3.

3.1. Data Augmentation

We used different data augmentation techniques such as rotation (20°), smearing (0.2), flipping (horizontal & vertical), zooming (range 0.1) as showed in Figure 4. Our dataset images were of different sizes (e.g., 1024×1024 , 128×128). For data augmentation we used batch size of 128 and converted image dimension to 128×128 .

3.2. Models

Baseline. As a baseline model, we used a basic VGG block. First, we used a normalized layer of the size of the input images (e.g., $128 \times 128 \times 3$), then we used 32 filters of size 3×3 followed by a max pooling layer. For the simplicity of this model, we did not use any Dropout for regularization.

VGG19. It is a CNN that is 19 layers deep. We trained it from scratch. We used Keras's API to get the VGG architecture and changed the final output layer from 1000 to 2.

ResNet50. We also experimented with ResNet50 architecture, a CNN containing 50 layers. Similar to VGG19, we trained ResNeT50 model from scratch.

Transfer Learning. To evaluate transfer learning, we used pre-trained VGG19 and ResNet50 (on ImageNet). We discarded the top layer and froze the base model. Then we flattened the output and used a dense layer with Softmax activation.

3.3. Domain Adaptation

Domain adaptation is the ability to apply a model trained in one domain (source) to a different but similar domain (target). We evaluated domain adaptation in the context of mask detection where we considered the simulated dataset as the



Fig. 2. Real Masked Face Dataset (RMFD) Samples

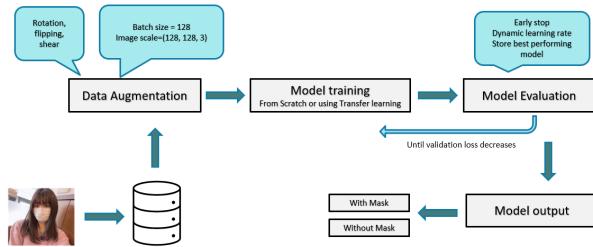


Fig. 3. Overview of our research methodology

source domain and the real dataset as the target domain. We trained our best model on the simulated data and test it on real data without retraining. Later, we fine-tuned the model on real dataset and observed the change in performance.

3.4. Training Process

We initialized each model with Adam Optimizer and initial learning rate of 0.0001. As this is a binary classification, we used categorical cross entropy as loss function and accuracy as a metric to be observed. We started the training with 300 epochs and defined four callbacks to prevent over-fitting. 1) We stored the best model after each epoch. 2) We stopped the training if the validation loss does not decrease for three epochs. 3) We half the learning rate after 10 epochs. 4) We stop the training if the validation accuracy was 100%. All models are trained on TALC cluster of University of Calgary with NVIDIA Tesla T4 GPU.

3.5. Evaluation Metrics

We reported the performances using standard evaluation metrics. Accuracy (A) is the ratio of correctly predicted instances out of all the instances. Precision (P) is the ratio between the number of correctly predicted instances and all the predicted instances for a given class. Recall (R) is the ratio of the number of correctly predicted instances and all instances for that class.

$$A = \frac{TP + TN}{TP + FP + TN + FN}, P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}$$



Fig. 4. Samples from augmented dataset

Table 2. Performance of the Deep learning models

Model	Simulated Dataset			Real Dataset		
	A	P	R	A	P	R
Baseline-1VGG	99.8	99.8	99.8	98	98	98
VGG19	99.8	99.8	99.8	99.2	99.2	99.2
ResNet50	99.6	99.6	99.6	96.8	96.8	96.8
TL-VGG19	99.7	99.7	99.7	98.6	98.6	98.6
TL-ResNet50	100	100	100	99.45	99.45	99.45

4. RESULT

In this section, we answer two research questions:

RQ1. How accurately different models can detect face mask?

RQ2. How effective the concept of domain adaptation is in face mask detection?

4.1. Performance of ML models (RQ1)

Table-2 shows the summary of the result of the experiment. From the table, we see that all the models perform moderately well in both datasets, however, the best performing model is ResNet50 with transfer learning. It achieves an accuracy of 100% in the simulated dataset and 99.45% in the real dataset.

4.2. Domain Adaptation Result (RQ2)

For this RQ, we did not consider any transfer learning model (that is already pre-trained on ImageNet) so that we can figure out the actual adaptation capability of model in the context of face mask detection. Hence, we trained VGG19 (our best performing non-transfer learning model) on the simulated dataset (source) and test it on the real dataset (target). It achieves 52% accuracy (i.e., 47% performance drop). Note that the performance of this same model is 99.2% when it is trained and tested on the real dataset (Table 2). Then we took the model (trained on source dataset) and fine-tuned it on using target



Fig. 5. TL-ResNet50 mispredicted samples on RMFD



Fig. 6. Difficult/incorrectly labeled test samples on RMFD

dataset. We find that the performance has quickly increased ($\sim 99\%$) after the fine tuning.

4.3. Error Analysis

We find that most models perform very good in both datasets. Figure 5 shows the 7 incorrectly predicted samples by our model. From our misclassification analysis and data exploration, we can see there are some samples in real test data which are either incorrectly labeled or very difficult to predict (Figure 6). This shows, any model which is performing more than 99% is one of the best possible models for this dataset.

5. DISCUSSION

5.1. Time Complexity Analysis of Different Models

Besides performance, training time is an important concern for any ML model. Hence, we conducted a time complexity analysis of different models. Figure 7 shows the performance and required training time for every model on real dataset. We can see that transfer learning (e.g., TL-ResNet50) can achieve high performance (99.45%) with very short training time (< 1000 s). Nearest competitor in terms of training time, ResNet50, achieves 96.8% accuracy in 1500s.

5.2. Implications of the Study

All models performed very well, even the simple baseline achieved $>98\%$ in both datasets. It has two implications—(1) Binary classification of masked vs unmasked is probably a simple problem (at least for used datasets), (2) Even a simple model like VGG1 can be successfully used for some problems with proper data augmentation. Another lesson could be

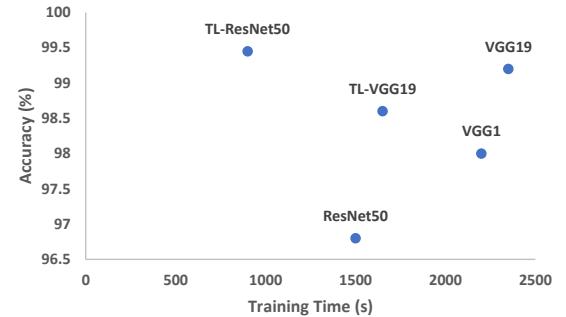


Fig. 7. Time and performance analysis of different models

the effectiveness of transfer learning. We can see how transfer learning improved the performance with less training time.

5.3. Limitations

The unmasked images of the real dataset mainly focused on the face only, but the masked images sometimes have whole body and multiple people as well. This phenomenon can have some manipulative effect on the model performance. Also, we did not make sure not to have same people in train and test set. However, as all the 525 persons of unmasked images are also present in masked images and the images are moderately diverse as well, there is little chance of bias here.

6. RELATED WORK

ML algorithms can help prevent the spread of the COVID-19 pandemic in many ways such as forecasting [15], diagnosis [13], exploring potential drugs for re-purpose [11]. Recently, several studies have also focused on using ML techniques for automated face mask detection. Some of them demonstrated solution to detect face mask in real time camera images[14, 9, 2]. Some proposed solution to detect if face mask is worn correctly or not [16]. Chowdary et al. proposed a transfer learning approach by fine-tuning InceptionV3 for face mask detection [4]. They used a different version of simulated dataset which contains around 700 masked and 700 without masks images. Loey et al. proposed a hybrid model where they used a pre-trained Resnet50's 2nd last layer as the feature vector for classical ML algorithms such as SVM, decision trees [12]. Both of these studies report more than 99% accuracy.

7. CONCLUSION

Our study shows that the current DL models for image classification are powerful for mask detection. Specially, performances of pre-trained transfer learning models are promising considering their time effectiveness. Future works can focus

on detecting mask in real time and incorrectly worn mask. We also see that model's performance drops significantly on real data when it is trained on simulated data. This indicates that the models are not robust for such real life usage.

8. REFERENCES

- [1] Who coronavirus disease (COVID-19) dashboard. Available: <https://covid19.who.int/>. [Online; accessed 1-December-2021].
- [2] K. Bhambhani, T. Jain, and K. A. Sultanipure. Real-time face mask and social distancing violation detection system using yolo. In *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, pages 1–6. IEEE, 2020.
- [3] 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*, 2020.
- [4] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal. Face mask detection using transfer learning of inceptionv3. In *International Conference on Big Data Analytics*, pages 81–90. Springer, 2020.
- [5] B. J. Cowling, K.-H. Chan, V. J. Fang, C. K. Cheng, R. O. Fung, W. Wai, J. Sin, W. H. Seto, R. Yung, D. W. Chu, et al. Facemasks and hand hygiene to prevent influenza transmission in households: a cluster randomized trial. *Annals of internal medicine*, 151(7):437–446, 2009.
- [6] S. E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, Y. Kuang, E. Kostelich, and A. B. Gumel. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the covid-19 pandemic. *Infectious Disease Modelling*, 5:293–308, 2020.
- [7] C. for Disease Control, Prevention, et al. Similarities and differences between flu and covid-19. *Retrieved July, 13:2020*, 2020.
- [8] M. Francis. Just how contagious is covid-19? this chart puts it in perspective. *Popular Science [Internet]. Popular Science*, 2020.
- [9] M. Inamdar and N. Mehendale. Real-time face mask identification using facemasknet deep learning network. Available at SSRN 3663305, 2020.
- [10] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014.
- [11] J. Y. Khan, M. T. I. Khondaker, I. T. Hoque, H. R. Al-Absi, M. S. Rahman, R. Guler, T. Alam, and M. S. Rahman. Toward preparing a knowledge base to explore potential drugs and biomedical entities related to covid-19: automated computational approach. *JMIR Medical Informatics*, 8(11):e21648, 2020.
- [12] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic. *Measurement*, 167:108288, 2021.
- [13] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi. Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. *Medical image analysis*, 65:101794, 2020.
- [14] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth. Ssdmv2: A real time dnn-based face mask detection system using single shot multibox detector and mobilenetv2. *Sustainable cities and society*, 66:102692, 2021.
- [15] F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, B.-W. On, W. Aslam, and G. S. Choi. Covid-19 future forecasting using supervised machine learning models. *IEEE access*, 8:101489–101499, 2020.
- [16] J. Tomás, A. Rego, S. Viciano-Tudela, and J. Lloret. Incorrect facemask-wearing detection using convolutional neural networks with transfer learning. In *Healthcare*, volume 9, page 1050. Multidisciplinary Digital Publishing Institute, 2021.
- [17] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, et al. Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*, 2020.