International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. , No. (Mon-20..)

http://dx.doi.org/10.12785/ijcds/XXXXXX

MASK DETECTION USING DEEP LEARNING METHODS

M. Mudasar Azeem[1], Inam Ul Haq[2], Muhammad Nauman[3], Muhammad Talha Hashmi[4] and Bilal Shabbir Qaisar[5]

[1,2,3,4,5]*Faculty of Computing, University of Okara, Okara 56300, Pakistan*
[*]*E-mail: hafizmudasar41@gmail.com, inamulhaq@uo.edu.pk, mr.nauman.edu@gmail.com, hashmi7125@gmail.com, bilal.qaisar725@gmail.com*

Abstract: If nothing changes, the COVID-19 pandemic will devastate institutions like the academy around the world, forcing them to lock their doors virtually. SARS-CoV-2 is a coronavirus that causes the severe acute respiratory syndrome. Droplets of contaminated respiratory secretions spread corona virus-2 when an infected person talks, sneezes, or coughs. Close contact with an infected person or exposure to infected surfaces and items speeds up the spread. The only surefire way to keep ourselves safe at this point is to avoid getting infected in the first place. One strategy to prevent exposure to the virus is to wear a facemask that covers the nose and mouth whenever one goes into a public place and to wash hands often or use sanitisers with at least 70% alcohol. As our ability to analyse images has improved, Deep Learning has proven to be an invaluable tool for recognition and classification. The study uses deep learning to determine if a person is correctly wearing a facemask if they are wearing a facemask at all, or if they are not wearing a facemask at all. The gathered dataset consists of 8982 photos with a resolution of 224x224 pixels, and the trained model attained an accuracy rate of between 99.55% and 98.94%. In real time, the system learns to distinguish between three distinct states—not wearing a mask, wearing the wrong mask, and wearing a mask. This research helps prevent infection and stop the spread of the virus.

Keywords: Coronavirus, MobileNet, MobileNetV2, Facemask

1. INTRODUCTION

To make sure people are wearing their masks correctly in public, scientists have tried to build automatic facial mask recognition algorithms. Since the COVID-19 pandemic, several advancements have been made in the field of face mask monitoring. In crowded places, surveillance systems can't see anyone's face because of image processing techniques [1]. In recent years, deep learning techniques have become increasingly popular for application in image analysis and object detection. A great deal of work has already made use of convolutional neural networks (CNNs). Present algorithms are inadequate in two respects when it comes to identifying face masks. Identifying a large number of faces "with mask and without mask" in a single frame of a video or photograph could be challenging. Women in our nation frequently wear half-veils, which are not classified as face masks according to current protocols.

Creating an effective and efficient classification approach is crucial for mobile face mask recognition. Many deep learning models are considered unsuitable for mask detection in the setting of mobile face photos due to the time and money needed for the evaluation process. To address the shortcomings of the existing method, the suggested approach makes use of MobileNet and Depthwise Separable Convolution [2]. Since its introduction in [3] and subsequent widespread use [4], depth-separable convolution (DSC) has become an indispensable tool for image processing classification tasks.

Recently, deep learning has been all the rage in many fields of information processing, including computer vision, text analytics, object recognition, and many more [5]. Much of the object detection research up to now has used models trained on convolutional neural networks. In recent years, convolutional neural networks (CNNs) have gained a lot of traction for many uses, such as image identification [1] and speech synthesis [2]. It is clear from the examples provided above that CNN has a significant capability for extracting information from visual data [6]. The use of CNNs is becoming increasingly popular as a replacement for traditional classification methods in order to better collect picture information while also achieving higher classification results. "The evaluation of many deep neural networks during the assessment step is time-consuming and expensive, which makes them unsuitable for mobile-based facial picture categorization applications. It is reported in this paper that a MobileNet-based face image classification model that makes use of a Depthwise separable convolution technique is developed and applied to the job at hand [1]. While there have been several approaches to solving image processing classification problems, DSC (Depthwise separable convolution) was the first technique to be introduced [3], [7]. Deepwise separable convolution quantizes the convolution in a quantized manner. Depthwise convolutions and 11 pointwise convolutions are two types of convolutions that are frequently distinguished in the literature. Depthwise convolutions are convolutions that are performed at different depths. Instead of applying each filter to all input channels as is the case with standard convolution, the Depthwise convolution layer employs an 11 pointwise convolution to combine the Depthwise convolution results, rather than applying each filter to all input channels as would be the case with standard convolution. By utilizing a depthwise separable convolution method, it is possible to it is feasible to minimize the number of parameters that must be learned, as well as the associated processing costs, significantly.

Many problems hampered the work that had been done before using deep learning methodologies. Face masks have been advised by the World Health Organization and other health agencies to prevent the spread of COVID-19 are a disease that is extremely contagious. People who do not wear facial masks are required to do so in public places, although it is impossible to manually identify those who do not wear them in large crowds.

The current study would propose a Deep Learning Methods for Mask Detection (MD-DLM), which would identify the dataset. The proposed method would be trained on the Mask Detection (MD) dataset. A deep learning technique would be suggested for the improvement of accuracy. The following are the contributions of the present research:

1) Develope a Mask Detection using Deep Learning Methods (MD-DLM) recognized the mask_weared_incorrect, with_mask and without_mask classes.
2) Improve the accuracy of the existing Deep Learning models.

2. LITERATURE REVIEW

As part of object recognition, face mask detection employs image processing algorithms. Classic image processing and deep learning-based image analysis are the two primary kinds of digital image processing that can be distinguished. Deep learning-based techniques to image processing, on the other hand, do not rely on intricate formulas, but rather on models that are designed to replicate the way the human brain functions. Deep Learning models have been employed in the great majority of previous research. Having identified the presence of a face in an image or video, Kaur et al. [1] employ a CNN-based technique to determine whether or not the face has been concealed. When used in a surveillance setting, it has the capability of detecting a mask or a moving face in videos. This approach has excellent accuracy. As part of an effort to recognize face masks in public settings, Bhuiyan and his colleagues created an algorithm called YOLO-v3 [8]. The YOLO-v3 model was trained using a proprietary dataset of photographs classified as "mask and no-mask" by the researchers.

Mata [9] augmented the model's performance by data enrichment. This ROI must be created using a CNN model that is capable of distinguishing between ROIs with and those without a face mask. Face masks may be detected using MobileNetV2 and three different face detector models, which were constructed by Toppo et al. [10] to verify the model's correctness and evaluate its performance. As a result of the model's training, the mask detection method can be implemented on low-power devices more quickly than previously possible. A VGG-16 CNN model created in Keras/TensorFlow and Open-CV was used by Balaji et al. [11] to identify people in government offices who were not wearing masks. Fan et al. [2] suggested two more ways to adjust for the model's small weight.

A novel module of residual contextual awareness for important sections of the face mask. Improved mask discriminating features are discovered using a two-stage synthetic Gaussian heat map regression. Feature engineering and numeric identification both benefit from these approach, according to ablation study. The proposed model outperforms previous models for AIZOO and Moxa3K. For lightweight face picture classification, conventional deep learning techniques have shown to be ineffective since they do not provide a good discriminating feature space, as demonstrated in the studies discussed above.

To better characterize the semantic label dependency and picture label relevance, [12] suggested a system employing CNN and RNN. Experiments conducted on the created framework using publicly available benchmark datasets proved the superior performance of the proposed architecture. In [13] developed, a categorization system employing spectral, spatial features. Hyperspectral datasets are used to conduct experiments, with the findings reported and compared to those of established hyperspectral classification techniques. Texture feature extraction and CNN were used to investigate [14] claim that we can reliably classify the grade of wood boards from their pictures. Classification methods using deep learning were compared to those using standard algorithms. In [15] developed a robust deep face detection system based on R-CNN. Experiments are conducted on the challenging detection benchmarks FDDB and WIDER FACE to demonstrate the strategy's efficacy. Several facial recognition algorithms have been studied in [Benaissaa and Kobayashia]. The authors looked into the feasibility of using facial recognition technology in supermarkets to offer loyalty discounts based on a customer's frequency of store visits. Business analytics rely on the system's collected data.

It is standard practice to train a convolutional neural network with thermal images as an additional set of input images in order to extract more information from the photos. An experiment was carried out by [16] to improve object detection in the thermal image domain using attributes from the RGB image. Similarly, to decrease recorded noise in RGB images, a combination system for image identification in [17] uses merging two individual input photos. Using a hybrid of the Retina and U-Net

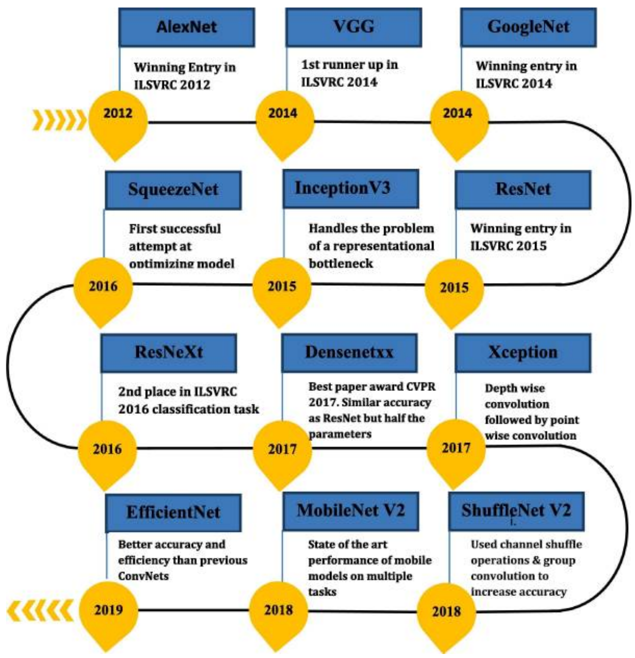


Figure 1. Pre-trained CNN architecture models from 2012 to 2018 [21].

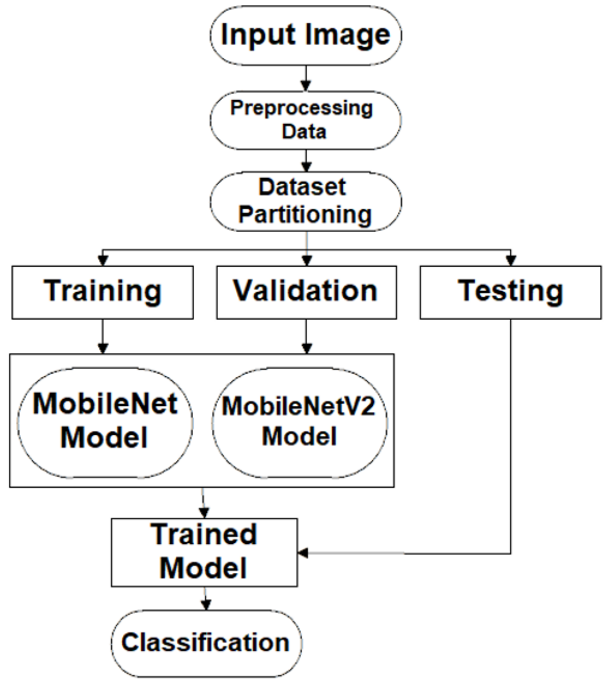


Figure 2. Proposed method flowchart.

networks—the former for semantic segmentation and the latter for a one-stage object detector—was demonstrated in [18] to further improve object detection through exploitation of data. Well-known object detection method RefineNet [19] uses two models for optimum accuracy. Two interdependent modules—one for object identification and one for anchor refining—form the basis of these combinations.

Numerous techniques exist for enhancing the functionality of both one-stage and two-stage detectors [20]. The simplest choice is to clean the training data so that convergence and average accuracy can be achieved faster. An often-used method for obtaining negative samples is the hard negative sampling approach [21]. To improve the accuracy or speed of detection, changing the context information is another option. Pre-trained CNN architecture models from 2012 to 2018 are shown in Figure 1.

3. METHODOLOGY

Artificial intelligence relies on such kinds of self-learning algorithms. Such algorithms are dynamic, changing over time as more data is collected on the project [22]. Technology aimed at fixing these problems is continuously developing. These models of the human brain allow self-learning algorithms to function [23]. Nodes (neurons) in artificial neural networks (ANNs) are connected in layers like human nerve cells. This neuron network is a storage facility, a processor (with positive or negative weighting), and an output device. ANNs holds a lot of promise because of their multi-layered structure and ability to spot subtler patterns. The term "deep learning" [24], [25] describes the types of learning these kinds of networks can be capable of.

This research implements a deep transfer learning system for face mask classification. First, the problem of class imbalance in the dataset is fixed, and variety is generated using pre-processing and several augmentation techniques. The second stage involves automatically extracting features and using a "MobileNet" and "MobileNetV2" model that has already been trained for face mask identification classification. Figure 2 shows a flowchart of the proposed procedure.

A. Dataset

A valid dataset is essential to the achievement of deep learning methods. To complete this study, we are utilizing the following data set.

1) Face Mask Detection Dataset 2020

The most extensive collection of high-quality face mask images made accessible for study can be found in the Face Mask Detection (FMD) dataset archive [26]. There are 8982 images in the dataset, 2994 related to the mask_weared_incorrect class, 2994 associated with the with_mask class, and 2994 related to images without_mask class. Figure 3 shows examples of the various classes. The pictures of the appropriate class were randomly picked from the entire collection. Following that, other techniques for data augmentation, such as rescaling, width shifting, rotation, shear range, horizontal flip, and channel shifting, were implemented.

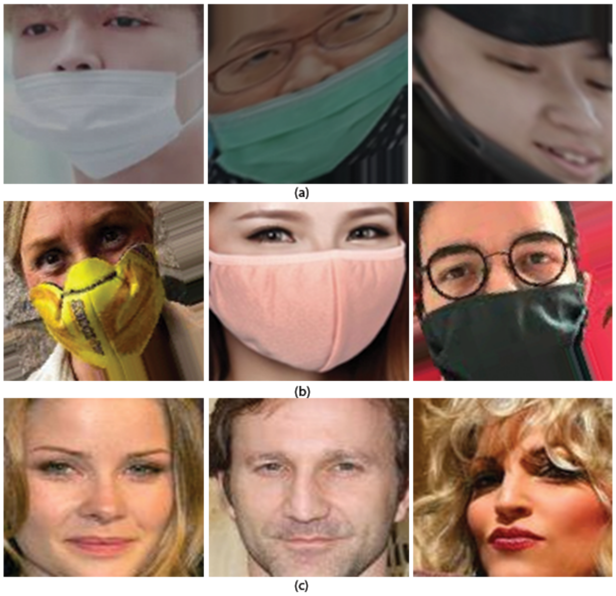


Figure 3. (a) mask_weared_incorrect, (b) with_mask and (c) without_mask classes of FMD dataset.

B. Image Pre-Processing

All input images for the face mask detection dataset are preprocessed to achieve increased consistency in classification results and advanced features. Because of the significant training time required for the CNN method and the potential for over-fitting, a large-scale image dataset was essential.

C. Training, Validation and Testing

Training, testing, and validating sets were available in the FMD-2020 dataset. The training set was used to teach the MobileNetV2 and MobileNet models, while the validation and test sets were used to evaluate the new model's performance. For this reason, we split the data set into a 70%:15%:15% training set, 15% test set, and 15% validation set. Models such as MobileNetV2 and MobileNet were trained using the data above set. The FMD-2020 dataset used 5394 training images, 1794 validation images, and 1794 test images. Summary of the FMD-2020 dataset as shown in Table I.

D. The Proposed Methodology

1) MobileNet Architecture

The MobileNet model is the first mobile computer vision model available in the TensorFlow library, as its name suggests. Convolutions in MobileNet are separated in depth. Compared to a network built with standard convolutions of the same depth, the number of parameters in this method is drastically reduced. A byproduct of this is compact deep neural networks that don't sacrifice performance. Since Google released the source code for the MobileNet class of CNN, we have a great foundation to build our ultra-compact, lightning-fast classifiers. Architecture is shown in Figure 4.

TABLE I. Summary of the FMD-2020 dataset

Split	**Classes**	**Label Samples**	**Total Samples**
Training	mask_weared_incorrect	1798	5394
	with_mask	1798	
	without_mask	1798	
Validation	mask_weared_incorrect	598	1794
	with_mask	598	
	without_mask	598	
Testing	mask_weared_incorrect	598	1794
	with_mask	598	
	without_mask	598	
Total			**8982**

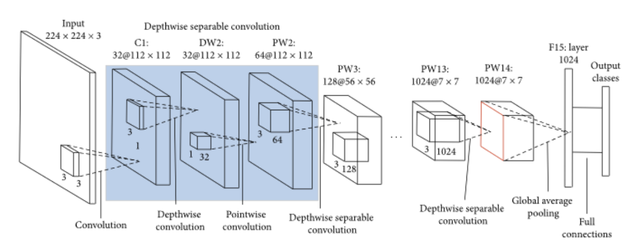


Figure 4. Proposed Method MobileNet Architecture.

2) MobileNetV2 Architecture

The problem of face mask categorization is investigated here using the deep transfer learning MobileNetV2 [27] architecture. The MobileNetV2 concept was chosen due to several contributing reasons. MobileNetV2 is a framework that minimizes the cost of errors while maximizing the speed and memory usage of the execution [27]. Using a compact but expressive system like MobileNetV2 helped reduce the risk of over-fitting due to the short size of the dataset used to train the model. The low memory footprint is a nice bonus, and the quick execution speed makes adjusting and experimenting with the parameters much more manageable. MobileNetV2 borrows heavily on the foundation established by its predecessor, MobileNetV1. Specifically, the depthwise separable convolution, linear bottleneck, and inverted residual are explored to shed light on how these three key ideas come together to describe the MobileNetV2 framework. Architecture is shown in Figure 5.

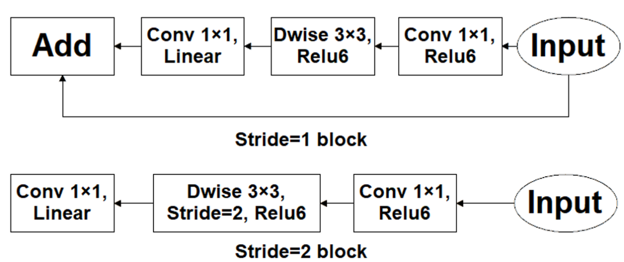


Figure 5. Proposed Method MobileNetV2 Architecture.

E. Evaluation Measures

The proposed method was evaluated on the testing dataset after the training phase. Accuracy, F1 score, precision, and recall were used to verify the architecture's performance. In the following sections, we'll investigate the performance measurements used in this study. True positives (TP), true negatives (TN), false negatives (FN), and false positives (FP) are defined and represented mathematically in the following.

1) Classification Accuracy

The accuracy of a classification system can be evaluated by determining what percentage of its predictions were correct and what percentage were incorrect.

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (1)$$

2) Precision

When analysing the effectiveness of a model, classification accuracy may not always be the most appropriate metric to employ. For instance, this is one of the scenarios where there is a considerable gap in socioeconomic status. It's a safe bet to assume that each sample is of the highest possible quality. If the model isn't picking up any new information, it would be irrational to infer that all components belong to the best class. Therefore, when we talk about accuracy, we refer to the fluctuation in findings you receive while measuring the same object several times with the same tools. The term "precision" refers to one of these statistics and can be defined as follows:

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

3) Recall

Another critical parameter is called recall, and it refers to the percentage of input samples that are of a type that the model can accurately predict. The formula for the recall is as follows:

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

4) F1 Score

The f1 score is a statistic utilised to contrast recall and precision.

$$F1Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (4)$$

4. RESULTS AND DISCUSSION

For training and testing, we took advantage of high-powered Graphics Processing Units (GPUs) hosted on a blank Google Colab [28] Pro account. For this purpose, we employed transfer deep learning models. The proposed FMD-DLM was trained using Sparse Categorical Crossentropy loss functions, and all tests were conducted using the Adam optimizer at a learning rate of 0.0001. The training iterations for this model were 10, with an initial batch size of 8, and the best val_loss models were preserved throughout the process. The MobileNet and MobileNetV2 models recommended the following parameters: 8 batches, 10 epochs, early pausing, and model saving based on val_loss.

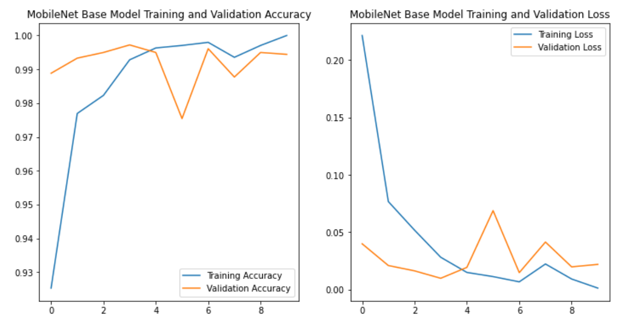


Figure 6. The MobileNet Base Model: (a) Accuracy (b) Loss Graph.

1) We evaluated the performance of the presented MobileNet and MobileNetV2 models on the FMD-2020 dataset using various data augmentation techniques.
2) The proposed MD-DLM is more effective than previous versions in terms of accuracy.
3) The results were compared with state-of-the-art techniques.

A. THE Performance Analysis of the Proposed Mask Detection using Deep Learning Method (MD-DLM)

1) MobileNet Proposed Model Performance on FMD-2020 Dataset

We evaluated and analysed the performance of the MobileNet base model on the FMD dataset. Validation accuracy for the model increased from 98.88% at the end of the first epoch to 99.44% after the most recent epoch. Training accuracy improves from 92.53% after the first epoch to 100% after the last epoch in Figure 6. As seen in Figure 6, MobileNet's validation loss drastically decreased from 4% to 2.21%. Furthermore, similar to the initial loss, the training loss was 22.14% after the first period and 0.14% after finishing training.

The performance of the MobileNet base model performed on an unseen test set is broken down in Table II. The model attained an average accuracy of 99.55% when applied to all classes in the test set; however, MobileNet achieved a precision of 99%, recall of 100%, and F1-score of one on the mask_weared_incorrect class. With a perfect 100% precision, 99% recall, and 99% f1 score, the with_mask class is remarkably outstanding. All three

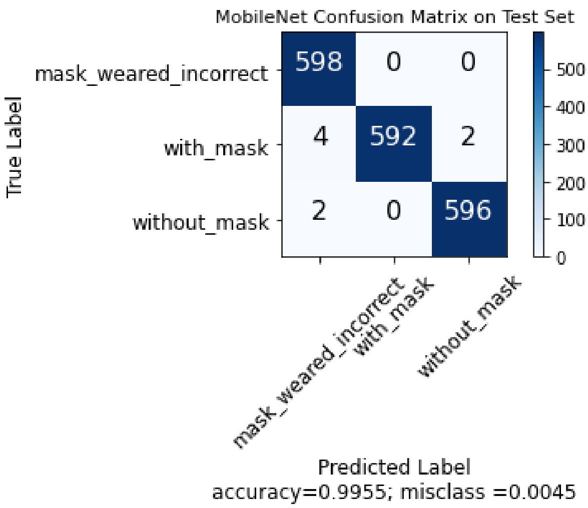


Figure 7. The MobileNet Base Model Confusion Matrix on Test Set.

metrics (f1 score, precision, and recall) achieved perfect scores for the without_mask class.

TABLE II. Precision, Recall, F1 Score, and Accuracy of the MobileNet Base Model.

Performance Measures	Precision	Recall	F1 Score	Accuracy
mask_weared _incorrect	99%	100%	100%	100%
with_mask	100%	99%	99%	98.99%
without_mask	100%	100%	100%	100%
Average Accuracy				**99.55%**

We could visually compare the categorization accuracy of different models using a confusion matrix. Predictions that turned out to be incorrect are represented by rows in the confusion matrix that is not on the diagonal. Darker colours indicated higher classification accuracy in the corresponding MobileNet base model for each class, while lighter colours indicated the existence of misclassified data. Confusion matrices from the test set will be used to assess MobileNet's overall effectiveness (shown in Figure 7). Predictions generated by the MobileNet baseline model were accurate across all image categories, as evidenced by the confusion matrix. Using the default parameters for the MobileNet model, the confusion matrix shows that 99.55% of the data were identified correctly, with only 0.45% of incorrect classifications. Comparing the confusion matrices for the mask_weared_incorrect, with_mask and without_mask samples demonstrate that the MobileNet base model performs admirably.

2) MobileNetV2 Proposed Model Performance on FMD-2020 Dataset

Using the FMD dataset, we analysed and evaluated MobileNetV2's performance as the base model. With each passing epoch, the model's validation accuracy improves, from 95.72% at the end of the first epoch to 99.16% after the most recent epoch. Figure 8 shows that the training accuracy increases from 92.08% after the first epoch to 99.56% after the last epoch. Figure 8 displays the dramatic improvement in validation loss from 14.18% in MobileNetV2 to 4.18%. Furthermore, the training loss was 23.06% after the first period and 1.34% after the concluding training, which is very similar to the original loss.

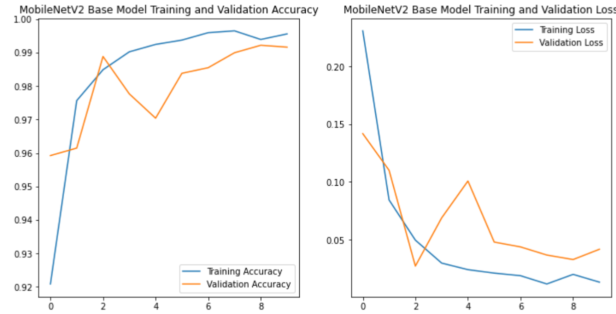


Figure 8. The MobileNetV2 Base Model Confusion Matrix on Test Set.

Table III summarises the results for an unseen test set using the MobileNetV2 baseline model. When applied to all classes in the test set, the model achieved an accuracy of 98.94%, with MobileNetV2 achieving a precision of 99%, recall of 100%, and F1-score of one on the mask_weared_incorrect class. The with_mask class has remarkable excellence: a perfect f1 score of 98%, recall of 97%, and precision of 100%. For the without_mask class, the f1 score, precision, and recall averaged 98%, 100%, and 99%, respectively.

TABLE III. Precision, Recall, F1 Score, and Accuracy of the MobileNet Base Model.

Performance Measures	Precision	Recall	F1 Score	Accuracy
mask_weared _incorrect	99%	100%	100%	100%
with_mask	100%	97%	98%	96.98%
without_mask	98%	100%	99%	99.83%
Average Accuracy				**98.94%**

Using a confusion matrix (as shown in Figure 9), we could examine how effectively various models classified data. The rows of the confusion matrix that are not on the diagonal represent predictions that turned out to be wrong. In the corresponding MobileNetV2 base model, higher classification accuracy for each class was represented by darker colours, while lighter colours indicated misclassified data. MobileNetV2's overall efficacy will be evaluated using confusion matrices from the test set. From what can be seen in the confusion matrix, the predictions made by the MobileNetV2 baseline model are spot-on for every single type of image. The confusion matrix demonstrates that when the

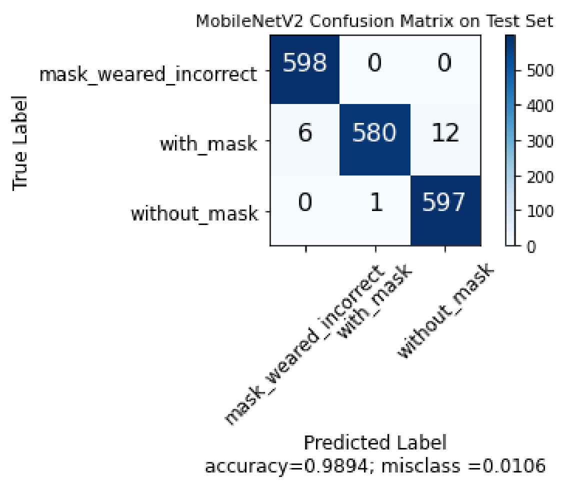


Figure 9. The MobileNetV2 Base Model Confusion Matrix on Test Set.

MobileNetV2 model was trained using the default parameters, 98.94% of the data were correctly identified, with only 1.06% incorrectly classified. The excellent performance of the MobileNetV2 base model is shown by comparing the confusion matrices for the mask_weared_incorrect, with_mask and without_mask samples.

B. Comparison with State-of-the-Art Studies

Researching published works did not work with the MobileNetV2 approach. Since this is the case, we are unable to compare the studies. We evaluate our model against face masks current gold standard for problem detection. A comparison of the accuracy of the proposed technique with that of other recent studies is provided in Table IV.

TABLE IV. Classification Accuracy of Proposed Model on Test Set.

Description	**Model Name**	**Accuracy**
MD-DLM	MobileNet Base Model	**99.55%**
MD-DLM	MobileNetV2 Base Model	**98.94%**

5. CONCLUSION

The research described in this manuscript explored using Convolutional Neural Networks to recognise facemasks in real-time using deep learning techniques. For facemask detection, this method is both reliable and quick. The test findings demonstrate a remarkable accuracy rate in identifying people who are either not wearing a facemask, are wearing the wrong facemask, or are not wearing any mask. The trained model completed the task using the MobileNet and MobileNetV2 models, with individual accuracy results of 99.55% and 98.94%. In addition, the study provides a practical tool for combating the transmission of the COVID-19 virus by determining whether or not a person is donning a facemask.

6. Future Scope

A future development would incorporate physical separation; in this case, the camera would detect if someone is wearing a facemask and, if not, it would measure the distance between them and trigger an alarm. It is suggested to incorporate multiple CNN models and compare each model's performance accuracy during training to enhance the performance of facemask detection and recognition. A new optimizer, improved parameter settings, finer tuning, and models for adaptive transfer learning are also proposed by the authors.

7. Acknowledgment

We all authors acknowledge the IT services, Cloud Servers, LAB facilities, and Technical Supervision provided by the University of Okara.

References

[1] M. R. Bhuiyan, S. A. Khushbu, and M. S. Islam, "A deep learning based assistive system to classify covid-19 face mask for human safety with yolov3," in *2020 11th international conference on computing, communication and networking technologies (ICCCNT)*. IEEE, 2020, pp. 1–5.

[2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[3] X. Fan and M. Jiang, "Retinafacemask: A single stage face mask detector for assisting control of the covid-19 pandemic," in *2021 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 2021, pp. 832–837.

[4] X. Fan, M. Jiang, and H. Yan, "A deep learning based light-weight face mask detector with residual context attention and gaussian heatmap to fight against covid-19," *Ieee Access*, vol. 9, pp. 96 964–96 974, 2021.

[5] A. Gumaei, M. M. Hassan, A. Alelaiwi, and H. Alsalman, "A hybrid deep learning model for human activity recognition using multimodal body sensing data," *IEEE Access*, vol. 7, pp. 99 152–99 160, 2019.

[6] K. Kamal, Z. Yin, M. Wu, and Z. Wu, "Depthwise separable convolution architectures for plant disease classification," *Computers and electronics in agriculture*, vol. 165, p. 104948, 2019.

[7] G. Kaur, R. Sinha, P. K. Tiwari, S. K. Yadav, P. Pandey, R. Raj, A. Vashisth, and M. Rakhra, "Face mask recognition system using cnn model," *Neuroscience Informatics*, vol. 2, no. 3, p. 100035, 2022.

[8] N. Balacheff and J. J. Kaput, "Computer-based learning environments in mathematics," in *International Handbook of Mathematics Education: Part 1*. Springer, 1996, pp. 469–501.

[9] E. Kamir, F. Waldner, and Z. Hochman, "Estimating wheat yields in australia using climate records, satellite image time series and machine learning methods," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 160, pp. 124–135, 2020.

[10] S. Balaji, B. Balamurugan, T. A. Kumar, R. Rajmohan, and P. P. Kumar, "A brief survey on ai based face mask detection system for

public places," *Irish Interdisciplinary Journal of Science & Research (IIJSR)*, 2021.

[11] D. Sage, F. R. Neumann, F. Hediger, S. M. Gasser, and M. Unser, "Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics," *IEEE transactions on image processing*, vol. 14, no. 9, pp. 1372–1383, 2005.

[12] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "Cnn-rnn: A unified framework for multi-label image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2285–2294.

[13] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.

[14] C. Affonso, A. L. D. Rossi, F. H. A. Vieira, A. C. P. de Leon Ferreira *et al.*, "Deep learning for biological image classification," *Expert systems with applications*, vol. 85, pp. 114–122, 2017.

[15] S. K. Addagarla, G. K. Chakravarthi, and P. Anitha, "Real time multi-scale facial mask detection and classification using deep transfer learning techniques," *International Journal*, vol. 9, no. 4, pp. 4402–4408, 2020.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[18] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[19] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," *Advances in neural information processing systems*, vol. 26, 2013.

[20] T. R. Gadekallu, D. S. Rajput, M. P. K. Reddy, K. Lakshmanna, S. Bhattacharya, S. Singh, A. Jolfaei, and M. Alazab, "A novel pca–whale optimization-based deep neural network model for classification of tomato plant diseases using gpu," *Journal of Real-Time Image Processing*, vol. 18, pp. 1383–1396, 2021.

[21] T. R. Gadekallu, M. Alazab, R. Kaluri, P. K. R. Maddikunta, S. Bhattacharya, and K. Lakshmanna, "Hand gesture classification using a novel cnn-crow search algorithm," *Complex & Intelligent Systems*, vol. 7, pp. 1855–1868, 2021.

[22] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Machine learning for predictive modelling based on small data in biomedical engineering," *IFAC-PapersOnLine*, vol. 48, no. 20, pp. 469–474, 2015.

[23] M. Attaran and P. Deb, "Machine learning: the new'big thing'for competitive advantage," *International Journal of Knowledge Engineering and Data Mining*, vol. 5, no. 4, pp. 277–305, 2018.

[24] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and f. Prabhat, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

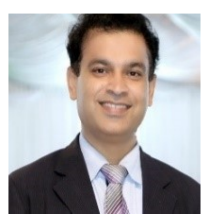
[25] S. K. Sarvepalli, "Deep learning in neural networks: the science behind an artificial brain," *Liverpool Hope University, Liverpool*, 2015.

[26] Kaggle. (Year of publication or last update) Face mask detection. Accessed: December 21, 2023. [Online]. Available: https://www.kaggle.com/datasets/vijaykumar1799/face-mask-detection

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[28] G. Colab. (Year of publication or last update) Google colab. Accessed: December 21, 2023. [Online]. Available: https://colab.research.google.com/

M. MUDASAR AZEEM was born in Okara Pakistan. The author received an MS Computer Science degree from University of Okara in 2023. He is currently working as an Lecturer at the TEVTA. His research interests include Artificial Intelligence, Machine Learning, Image Processing, Healthcare and Wellness.



INAM U. HAQ (M'1979) was born in Okara Pakistan. The author received an MS Computer Science degree from Blekinge Institute of Technology Sweden in 2013 and is now a PhD Scholar from Superior University Lahore. He is currently working as an Assistant Professor at the University of Okara from 2005 to date. His research interests include Artificial Intelligence, Machine Learning, Image Processing, Healthcare and Wellness. He has published 08 research papers as a Principal Author in international journals/conferences. Inam is a member of IEEE, ACM, Parkinson's Disease Foundation, Movement Disorder Society, IEEE Sensors Council, IEEE Electronic Design Automation Council, IEEE Nanotechnology Council, IEEE Biometric Council, IEEE Education Society IEEE Young Professionals, IEEE Collabratec, and many others.

Mr. M Nauman work stands at the forefront of technological progress, shaping the future of AI and its applications. The scientist Muhammad Nauman, Known as Nauman Malik (MS. Scholar), received his master's degree in computer science from the University of Okara, Pakistan. Currently, he is working as a Trainer of Robotics at the University of Central Punjab Okara, Pakistan. His research interests include image processing, machine learning, data science, plant disease detection, and deep learning. Mr. M Nauman has demonstrated a keen ability to push the boundaries of knowledge and provide valuable insights into the intricacies of his field. His commitment to rigorous investigation and scholarly contribution is a testament to his unwavering pursuit of excellence.

MUHAMMAD TALHA HASHMI is born in Okara Pakistan. The Author received an MS Computer Science degree from the University of Okara in 2023. He is currently working as a Manager at the eFaida company. His research interests include Artificial Intelligence, Machine Learning, and Image Processing.

BILAL SHABBIR QAISAR has received his BS degree in Computer Science from the Quaid-I-Azam University, Islamabad in 2019. Now he has received his MS degree from university of Okara in 2023. He is currently working as an Lecturer at the University of Okara. His research interests include Image Processing, Deep Learning, Medical Imaging, Image Classification, Detected Diseases from Plant and Computer Vision.