Real Time Face Mask Detection Using Gamma2-MobileNet Model

Anurag Gaonkar  
*Department of Artificial Intelligence and Data Science*  
Thadomal Shahani Engineering CollegeMumbai, India  
anuraggaonkar36@gmail.com

Janhavi Gangawane  
*Department of Artificial Intelligence and Data Science*  
Thadomal Shahani Engineering CollegeMumbai, India  
janhavig2004@gmail.com

Devang Gupta  
*Department of Artificial Intelligence and Data Science*  
Thadomal Shahani Engineering CollegeMumbai, India  
devanggupta21@gmail.com

Khozaima Hammad  
*Department of Artificial Intelligence and Data Science*  
Thadomal Shahani Engineering CollegeMumbai, India  
khozaimahammad@gmail.com

Himani Deshpande  
*Department of Artificial Intelligence and Data Science*  
Thadomal Shahani Engineering CollegeMumbai, India  
himani.deshpande@thadomal.org

***Abstract*— The COVID-19 pandemic highlighted the critical need for automated face mask detection systems to ensure compliance with health regulations in public spaces. This study proposes a Gamma2-MobileNet model for real-time face mask detection, leveraging deep learning techniques to improve accuracy and computational efficiency. The research explores conventional and deep learning-based face mask detection approaches, highlighting their limitations in real-world scenarios such as varying mask types, occlusions, and lighting conditions. The dataset consists of 7,553 RGB images, categorized into masked and unmasked faces, with data augmentation techniques applied to enhance model robustness. Two models were evaluated: a custom CNN and a fine-tuned MobileNet model. The custom CNN model achieved 96.69% accuracy, while the fine-tuned MobileNet model achieved 99.34% accuracy, demonstrating superior precision and recall. The proposed Gamma2-MobileNet model, optimized for real-time inference, outperforms traditional models by balancing accuracy and computational efficiency. Results indicate that the model generalizes well under various conditions, making it suitable for deployment in high-traffic environments such as airports, railway stations, and shopping malls. Despite its high accuracy, challenges remain in detecting improperly worn masks and handling extreme lighting variations. Future work aims to enhance model adaptability through attention mechanisms, multi-modal biometric integration, and edge computing for optimized real-world applications. Unlike existing MobileNetV2-based models, our proposed Gamma2-MobileNet incorporates attention mechanisms, lightweight optimizations and enhanced feature extraction that leads to higher real-time detection efficiency. This enhances AI-driven public health monitoring by reducing false positives and improving detection under occlusions and varying lighting conditions.**

***Keywords— Facemask Detection, Deep Learning, Gamma2-MobileNet, Real-time Detection, Computer Vision, Convolutional Neural Network, MobileNet Model, machine learning, public health monitoring, COVID-19, Automated surveillance, Image classification, Occlusion Handling, Data Augmentation, accuracy and efficiency, precision and recall, Edge Computing, Biometric Integration, Smart surveillance.***

# Introduction

The quick spread of infectious diseases, especially the COVID-19 pandemic that hit us a few years ago, has highlighted the importance of wearing face masks as one of the preventive measures, particularly in crowded public places [[1]](#bookmark=id.3dy6vkm). Governments and health organizations across the globe, foremost among them being the World Health Organization (WHO) have considered implementing mask mandates to reduce the spread of viruses, highlighting the need for proper monitoring of compliance [[2]](#bookmark=id.1t3h5sf). Nevertheless, the method of manually inspecting compliance is impractical because it would be very time-consuming and strenuous on the available human resources [[3]](#bookmark=id.4d34og8). This created the necessity for automated systems that could detect whether people wear face masks or not to provide a means for public safety.

Face mask detection is a specific and crucial advancement made in object detection that stretches through two major steps: to detect faces from images and videos and to classify those faces as either masked or unmasked [[4]](#bookmark=id.2s8eyo1). While face detection is a well-addressed problem in computer vision, the presence of masks adds complexity. Masks cover the distinct facial features (nose and mouth) which are important for standard systems of face recognition [[5]](#bookmark=id.17dp8vu). Therefore, face mask detection is a new and challenging challenge that requires innovative solutions to be precise and dependable.

There are still huge gaps in today's work regarding these issues, seeing that changes in illumination, various orientations of the masks of variable types, and altering rotations of the face in the input image would greatly impact their generalization performance. Many systems perform excellently under controlled conditions, such as when frontal face images are used with uniform lighting. While this may be acceptable, the broader applicability in the real world is severely limited [[7]](#bookmark=id.26in1rg). This issue calls for the design of a stronger and more flexible solution that should be capable of functioning effectively in the far harsher conditions that such applications must contend with in the real world, such as crowded public spaces with different lighting and other factors.

The objective of this study is to design an accurate as well as efficient CNN model which detects face masks in real-time. The proposed system improves the detection accuracy under difficult situations like occlusions, different angles of faces and various types of masks, to name a few, without compromising the speed of computation [[9]](#bookmark=id.35nkun2). This study attempts to contribute to the development of practical solutions for automated mask detection in commercial places like airports, railway stations, shopping malls, etc. by tackling these issues.

This approach is designed for deployment in settings such as airports, train stations, and malls, and engages upon robustness and adaptability; thus, making it viable for real-world applications. Through extensive testing, we ensure reliability in dynamic environments, contributing to public safety during pandemics and beyond. This work not only advances the field of face mask detection but also offers a practical tool for enhancing public safety during pandemics and beyond.

The remainder of this paper is organised as follows: Section [II](#_heading=h.gjdgxs) provides a literature review on deep learning approaches for face mask detection. Section [III](#_heading=h.30j0zll) details the methodology, including dataset preparation, model architecture, training, and evaluation. Section IV discusses the results of the evaluation process. Finally, Section V concludes the paper, subsequently ending with acknowledgments and references.

# Literature Review

Face mask detection is an important application of computer vision, especially since the COVID-19 pandemic has broken out. It helps us in detection of masks accurately in real time which is essential for the greater public health enforcement to maintain compliance within crowded areas such as that of the airport, shopping malls, and workplaces. We can achieve this through Deep learning, particularly convolutional neural networks that has been able to bring competence to this area as it is capable of offering more accurate recognition and object classification under difficult conditions like occlusions, lighting variations, and pose variations.

*A. Development of Conventional Face Mask Recognition Techniques*

Face mask detection was done using conventional computer vision methods prior to deep learning. These techniques used statistical models for classification and manual feature extraction. Canny Edge Detection is an algorithm that identifies facial edges in gray-scale images but it is sensitive to light and noise [[13]](#bookmark=id.3j2qqm3). In Histogram of Oriented Gradients (HOG) approach, facial features are detected by the brightness of the overall image. However, it becomes difficult when the shape of the mask changes and when the parts of faces are covered [[14]](#bookmark=id.1y810tw).

Haar Cascades is a machine learning algorithm which helps in detecting faces by searching for rectangular features. It does not learn and thus fails to detect face masks of varying colors and textures [[15]](#bookmark=id.4i7ojhp). These methods are useful for face detection tasks but they are not sufficient for real-time mask compliance monitoring, which lead to the adoption of deep learning based object detection models. To overcome the limitations of traditional methods, researchers turned to object detection models capable of identifying multiple classes (e.g., masked and unmasked faces) in a single inference.

Single Shot MultiBox Detector (SSD) is applied to predict bounding boxes directly and class probabilities but requires massive labeled datasets to learn effectively [[16]](#bookmark=id.2xcytpi). YOLO (You Only Look Once) is an advanced version of YOLOv4 and YOLOv8 and it allows real-time face mask detection with negligible computation costs, for which it can be used effectively for monitoring at public places [[17]](#bookmark=id.1ci93xb). Faster R-CNN (Region-Based CNN) increases the precision of detection as it creates region proposals before the classification step but due to extreme computational needs can't be deployed in real-time applications [[18]](#bookmark=id.3whwml4). These object detection models laid the foundation for CNN architectures, which further improved accuracy and adaptability in face mask detection.

*B. Methods for Face Mask Detection Based on Deep Learning*

Face mask detection has been transformed by CNNs, which provide excellent accuracy in a number of situations. Many architectures have been carefully investigated:MobileNetV2 is a lightweight CNN, designed for low-power devices that achieves high accuracy for classification in real-time applications [[1]](#bookmark=id.3dy6vkm). ResNet-50 is well known for its deep architecture and residual connections; it performs better than many models; some studies have reported an accuracy of 98.2% [[2]](#bookmark=id.1t3h5sf). Techniques like the Convolutional Block Attention Module (CBAM), which achieves a 17.4% improvement in masked-to-unmasked recognition, increase accuracy by concentrating on important facial regions that are visible even when masks are worn. [[4]](#bookmark=id.2s8eyo1).

*C. Challenges and Potential Areas for Improvements in Face Mask Detection*

Despite significant advancements, face mask detection systems still face various challenges. Different masks (e.g., surgical, cloth, N95) vary in shape, size, and colour, affecting model generalization. Some models struggle to detect masks that are improperly worn (e.g., below the nose) [[1]](#bookmark=id.3dy6vkm). Many models perform well in controlled settings but fail in real-world environments with extreme lighting conditions or partial occlusions. Another major challenge is the lack of high-quality, diverse datasets. Existing datasets such as the Real-World Masked Face Dataset (RMFD) fail to capture variations across ethnicities, lighting conditions, and mask types. Crowdsourced datasets and synthetic augmentation are being explored to overcome this limitation [[5]](#bookmark=id.17dp8vu). High-accuracy models are often computationally expensive, making them unsuitable for real-time deployment on mobile and embedded systems. While models such as MobileNetV2 are optimized for such scenarios, achieving a balance between accuracy and efficiency remains a challenge [[5]](#bookmark=id.17dp8vu). Attention-based models, such as CBAM, have shown promise in improving masked face recognition, particularly in poor lighting and occlusion scenarios [[4]](#bookmark=id.2s8eyo1). Combining face mask detection with biometric authentication methods (e.g., iris or voice recognition) could enhance security and reduce misclassification errors. Larger, more diverse datasets covering various demographic and environmental conditions are necessary to improve model generalization. Techniques such as crowdsourcing and synthetic augmentation can aid dataset expansion [[3]](#bookmark=id.4d34og8). Further optimization of deep learning models for edge deployment through techniques like quantization and pruning can make real-time face mask detection more efficient [[5]](#bookmark=id.17dp8vu). Research suggests combining face mask detection with automated screening for vaccination verification systems that use QR codes to improve public health monitoring [[6]](#bookmark=id.3rdcrjn). There has been significant advancement in employing deep learning to identify face masks. CNN models such as MobileNetV2, ResNet-50, and CMNV2 have proven to be highly accurate. Nevertheless, there are limitations with the diversity of datasets, adjustment to real-world scenarios, and the efficiency of computations. Future studies should investigate attention mechanisms, merging various forms of biometrics, enhancing datasets, and the use of edge computing to develop robust and scalable solutions. As advancements continue, automatic face mask detection can be highly significant for public health enforcement and surveillance globally.

# Methodology

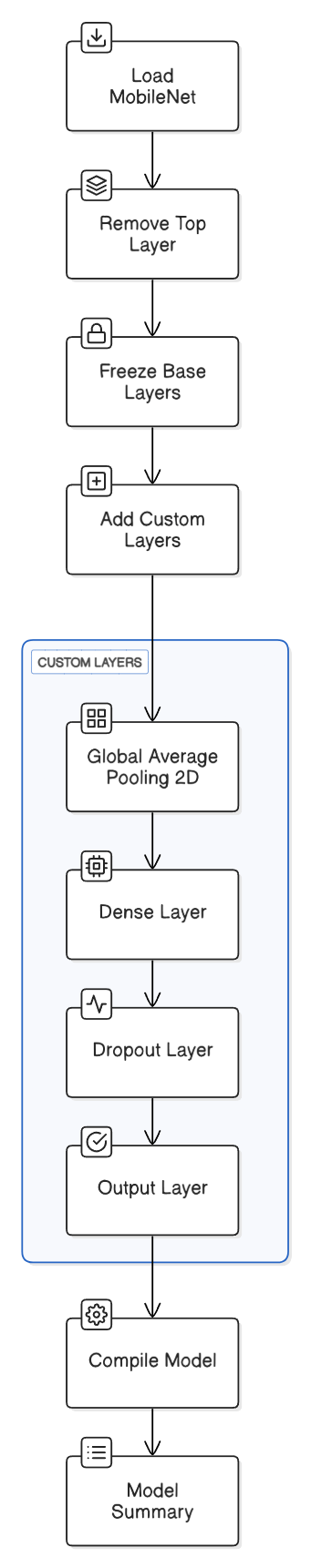
This study focuses on developing a real-time face mask detection system using deep learning. To achieve this goal, we have experimented with two models, a custom built Convolutional Neural Network (CNN) model which was further extended to create a modified MobileNet model. The model proposed in this paper for the robust detection of face masks is promising and appropriate for real-time scenarios using the MobileNET model. MobileNet was chosen as the primary focus of this study because it is lightweight and efficient, making it ideal for real-time applications when you have a live video feed of people. The most important step towards building this model is to gather an extensive and reliable dataset of images or live video feed to train the deep learning model in predicting if a person is wearing a face mask or not. The details of the dataset used in this study are found in [Section III A](#bookmark=id.1fob9te).

*A. Dataset Preparation*

For this study, we have put together a dataset of 7553 RGB images sourced primarily from two repositories: Prajna Bhandary’s GitHub repository – 1,776 images [[11]](#bookmark=id.44sinio). Google Image Search – 5,777 images, carefully filtered to remove irrelevant or low-quality samples. The images were placed into two folders: "with\_mask" (3725 images) and "without\_mask" (3828 images). Each of the images was resized to a common width and height of 150x150 pixels as the input shape to maintain consistency when training the model. Data augmentation is further performed with TensorFlow’s ImageDataGenerator to increase the model robustness and avoid overfitting. This included performing small rotations (±10°), slight zooming (±10%) and horizontal flipping The pixel values were further scaled and normalised to [0, 1] to enable bias-free evaluation. The dataset was split into 80% for training and 20% for validation. Batches of 32 images each were generated for both the training and testing phases.

*B. Model Architecture*

A modified MobileNet model was developed on top of a custom CNN model and evaluated in this study. With an eye toward transfer learning, MobileNet, a model trained on ImageNet [[12]](#bookmark=id.z337ya) was modified. Instead of starting from scratch, we removed the top classification layer while keeping the base layers frozen at first. To customize MobileNet to detect face masks, four custom layers were added. The Global Average Pooling-2D layer compressed the feature maps that were extracted by MobileNet. Since this feeds into the next layer, it also sets the stage by reducing data size while keeping the most prominent patterns within data; thus, it would perform more straightforwardly. This was followed by a dense layer with ReLU activation, consisting of 128 neurons trained to learn towards the discovery of complex patterns highly specific to the task of detecting Most common Face Masks. A dropout (0.5) layer was added to prevent overfitting. 50% of the neurons are randomly disabled during training, tricking the model into learning generalizable features rather than too much about some specific features. Finally a dense output layer with a single neuron and sigmoid activation was added which produced a probability score between 1 and 0, classifying images as "with\_mask" or "without\_mask". Gamma2-MobileNet introduces an optimized lightweight architecture with fewer parameters than traditional MobileNetV2 which improves real-time detection efficiency while maintaining high accuracy. Additionally, our model integrates attention mechanisms to enhance detection under occlusions for optimized edge deployment. Owing to its light architecture, the model is ideal and suitable for deployment straight out-the-box in real-world applications. The proposed architecture, illustrated in Figure 2, outlines the structure of Gamma2-MobileNet, demonstrating its efficiency in feature extraction and classification.



*Fig. 2 : Proposed Model Architecture*

*C. Model Training*

The training of the model took place in two stages which are explained as follows: Stage 1 involved the feature extraction and the initial training. For the first phase, the base layers within MobileNet were frozen and the only layers that were trained were the four custom layers that were added here. This permitted the model to be made competent towards new tasks without overwriting learned features from ImageNet. The model was trained for binary classification using Adam optimiser with its default learning rate and Binary-Cross Entropy loss function. The model was trained in 5 epochs with early stopping and model checkpointing halting the training if validation loss plateaued and saving the best model. The custom CNN model was similarly trained, which achieved a training accuracy of 94% and validation accuracy of 96%. For fine-tuning the MobileNet model, the last 20 layers were unfrozen to allow refining of deeper features. A reduced learning rate (1e-5) was applied with the Adam optimizer, and training continued for 5 additional epochs with the same callbacks. This step refined the pre-trained weights to better fit the face mask dataset. Training utilized the augmented training generator and validation generator, with steps per epoch calculated as the total samples divided by the batch size (32).

*D. Model Evaluation and Prediction*

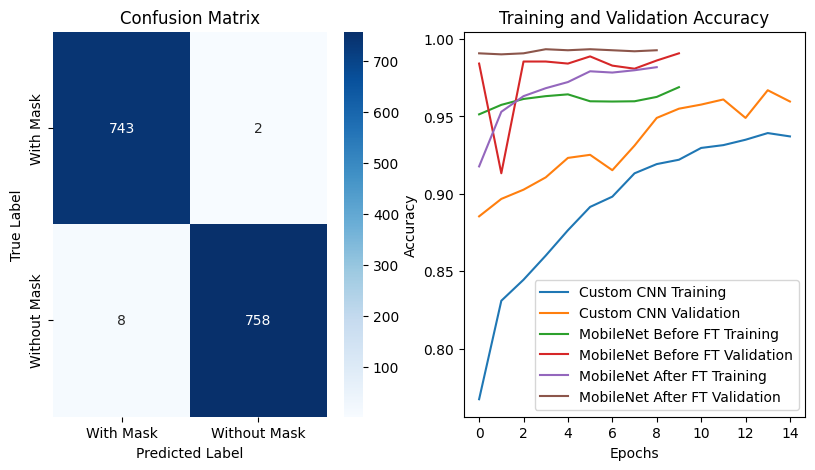
After training, the modified MobileNet model was assessed on the test set using the evaluate function, which measured both test loss and accuracy. A prediction function was created to classify new images by resizing them to 150x150 pixels, normalizing the pixel values, and applying a 0.5 threshold to the model’s sigmoid output ("With Mask" if < 0.5, "Without Mask" if > 0.5). This function was tested on sample images to confirm its practical performance. Training and validation accuracy/loss curves were plotted using Matplotlib to visualize how the model performed.

# Results and Analysis

For comparison of the efficiency of the proposed face mask detection system, two deep learning models were utilized: a custom CNN model and the proposed modified MobileNet model, focusing on important metrics such as accuracy, precision, recall, and F1-score to measure their efficiency. Custom CNN model is a three-layer convolutional baseline model and the proposed modified MobileNet is a transfer learning model that was fine-tuned for face mask detection. The training, validation and test accuracy for both models are as represented in Table 1.

| **Model** | ***Training Accuracy*** | ***Validation Accuracy*** | ***Test Accuracy*** |
| --- | --- | --- | --- |
| Custom CNN | 96.61% | 95.96% | 96.69% |
| MobileNet (Before Fine-tuning) | 99.43% | 99.07% | 99.07% |
| MobileNet (After Fine-tuning) | 99.62% | 99.27% | 99.34% |

TABLE 1: COMPARISON OF MODEL PERFORMANCE ON MASK DETECTION DATASET



*Fig. 3. Training and Validation Accuracy*

Figure 3 shows the results of the Training and Validation Accuracy of the Custom CNN Model, MobileNet Model, before Fine-Tuning and MobileNet Model, after Fine-Tuning. The curves in the diagram (Accuracy/Loss curves) indicate the learning process of the models. The Custom CNN model showed a very slight improvement in training accuracy, but its validation accuracy got saturated, meaning it was overfitting. MobileNet before Fine-Tuning showed little generalization error but gave better accuracy than Custom CNN. As shown in Figure 3, the Fine-Tuned MobileNet model exhibits minimal generalization error, indicating improved learning stability and robustness. This indicates that Fine-Tuning increased the model's learning and feature extraction ability. A few performance measures were computed to validate model performance. Precision is measured by the ratio of times it correctly classified the masked and unmasked individuals. MobileNet was characterized by fewer false positives when classifying unmasked individuals; thus it is more reliable for detection. Recall determines how well the model identifies masked individuals correctly with the least possible false negatives and the highest sensitivity to detect masks. The model has high recall, i.e., few false negatives.The F1-score brings both precision and recall into one single measure of performance. It balances between the false positives and false negatives for reliable face mask detection. Hence, it outperformed. The model is used to estimate its generalization by training on and testing against different datasets.

TABLE 2: PERFORMANCE METRICS FOR THE PROPOSED MOBILENET MODEL

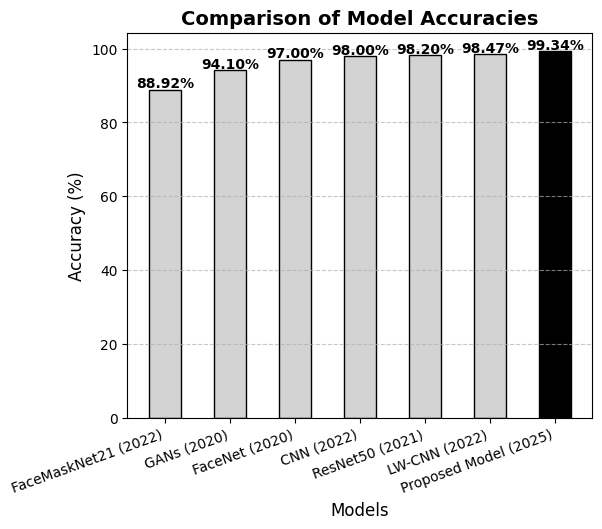
| **Dataset** | **Performance Measures** | | | |
| --- | --- | --- | --- | --- |
| ***Accuracy*** | ***Precision*** | ***Recall*** | ***F1 score*** |
| With Mask | 99% | 100% | 99% | 99% |
| Without Mask | 99% | 99% | 100% | 99% |

Table 3 represents the Confusion Matrix, displaying the model’s prediction against actual values. It consists of four components, TP(true Positive), TN(True Negative) FP(False Positive), FN(False Negative), where TP indicates correctly detected face masks, TN indicates correctly identified without mask faces, FP indicates incorrectly detected masked faces, and FN indicates incorrectly detected without masked faces.

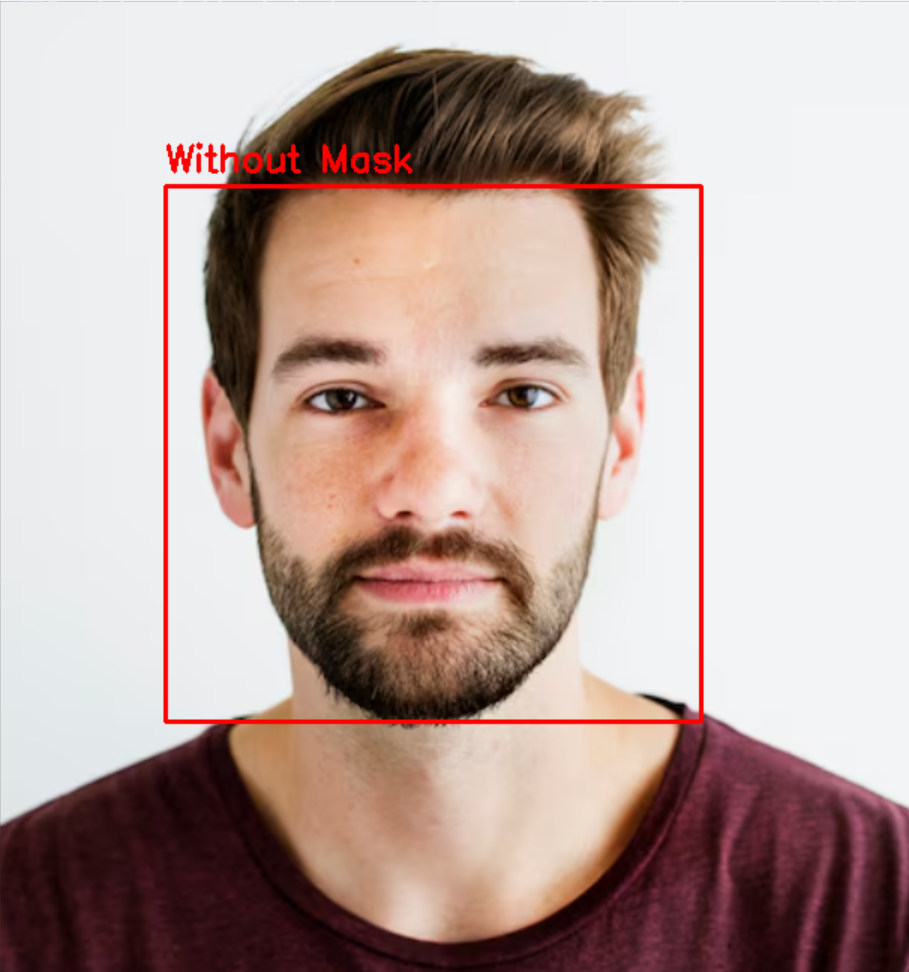
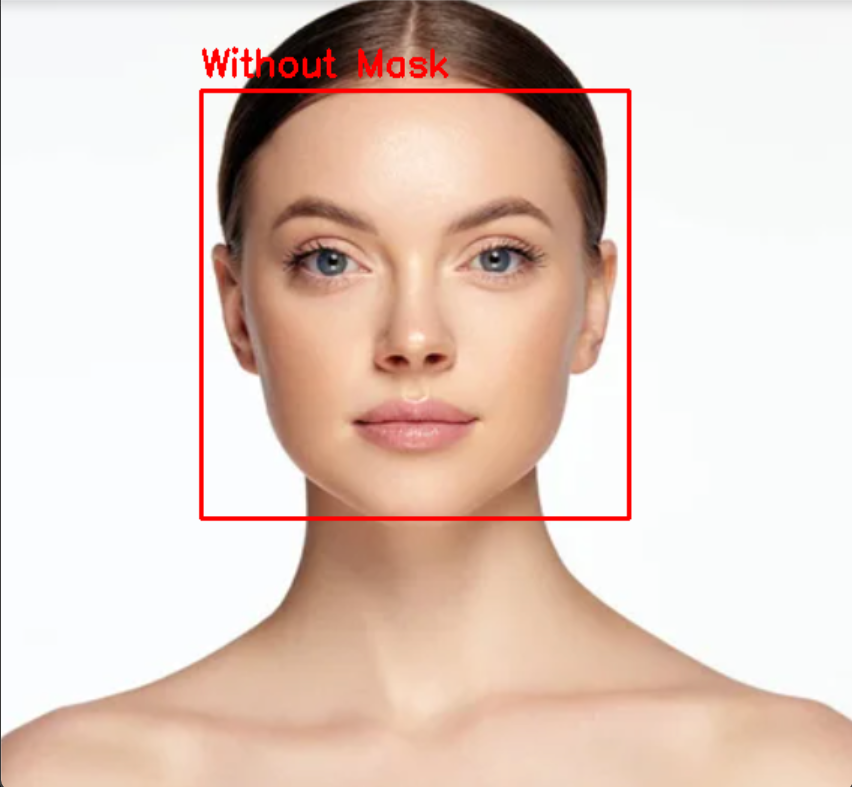
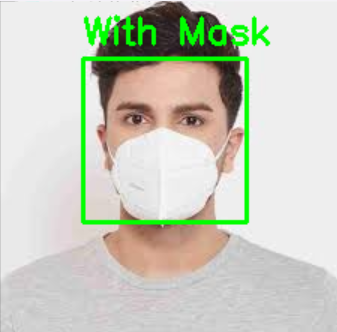
TABLE 3: CONFUSION MATRIX

| **True Label** | **Predicted** | |
| --- | --- | --- |
| ***With Mask*** | ***Without Mask*** |
| With Mask | 743 | 2 |
| Without Mask | 8 | 758 |

As shown in Figure 4, accuracy of the proposed model was compared with different previously built models, which shows that this proposed model has achieved significant accuracy as compared to other predefined models.

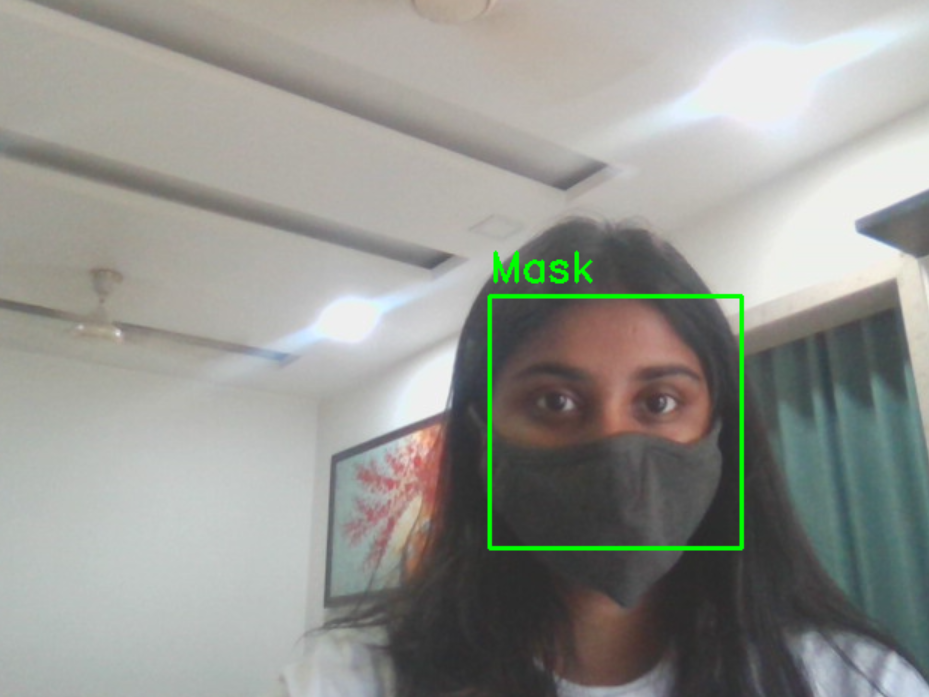
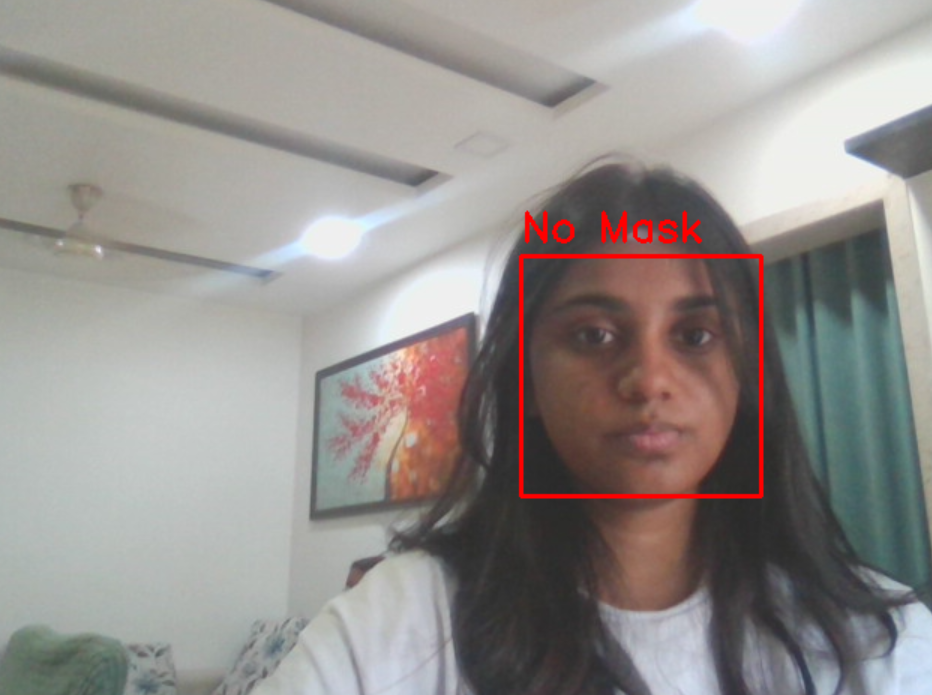


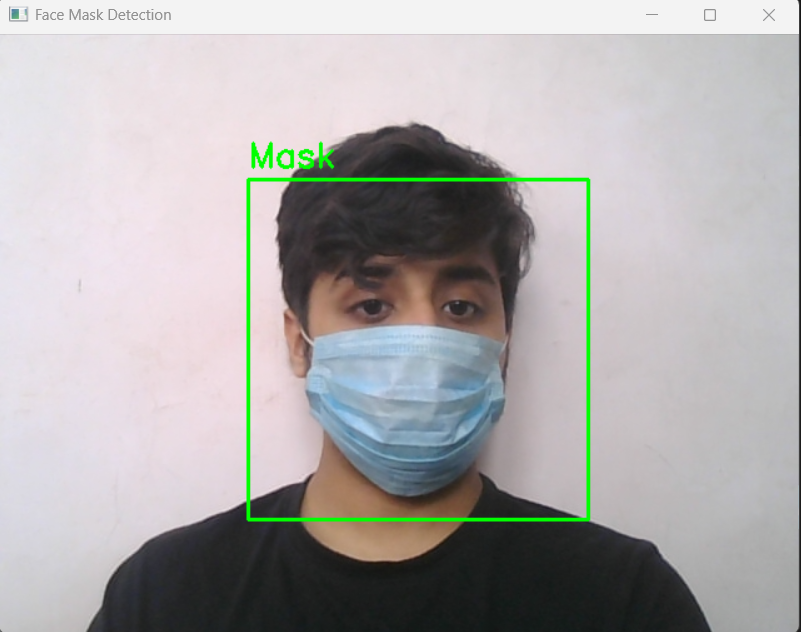
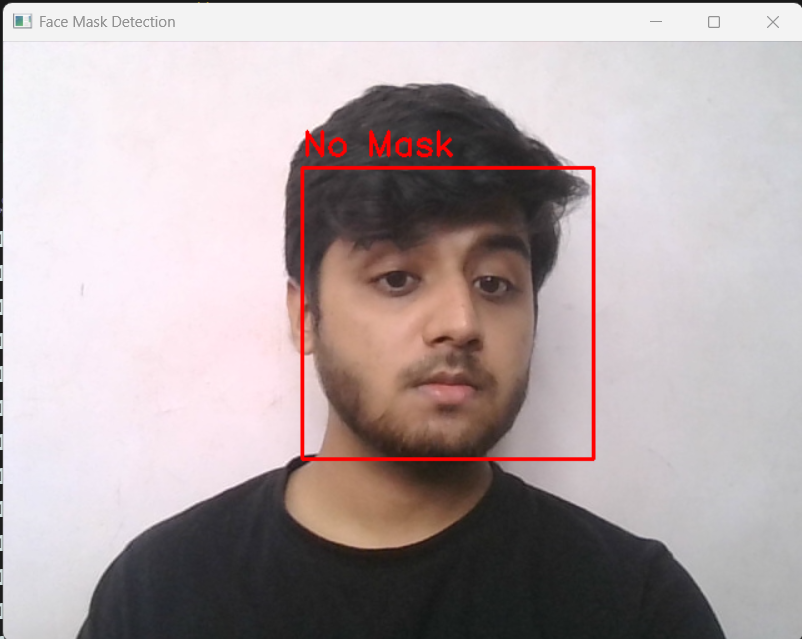
*Fig. 4. Comparison of proposed model’s Accuracy with previous models’ Accuracies*



*Fig. a : Test results on static images*







*Fig. b : Test results on Real-time Face Mask Detection*

*Fig. 5. Test Results*

# Conclusion

This study focused on performing real-time face mask detection to provide a feasible implementation of an automated system used to ensure public health compliance during public health crises such as Covid-19 pandemic using the proposed Gamma2-MobileNet model. The model showed excellent results with an accuracy of 99.34% after fine-tuning. This was better than the custom CNN model’s 96.69% accuracy. The model also had better precision, recall and F1 score compared to the CNN model. The model proved to be effective against other conditions such as occlusions, brightness, and different mask types. Unlike existing models, Gamma2-MobileNet effectively distinguishes between actual face masks and occlusions (e.g., hands covering the face), minimizing false positives. The model has good use-case scenarios at places like airports and shopping malls. But, it struggles with poorly worn masks as well as extreme lighting changes, all of which will adversely affect deployment performance. To solve these problems, future work should explore more into attention mechanisms, combining different biometric inputs, and using edge computing to improve adaptability and real-life applicability. These advancements could lead to next-generation AI-driven public health monitoring systems that focus on improving compliance in more dynamic environments.

##### References

1. Habib, S., Alsanea, M., Aloraini, M., Al-Rawashdeh, H.S., Islam, M. and Khan, S., 2022. An efficient and effective deep learning-based model for real-time face mask detection. *Sensors*, *22*(7), p.2602.
2. Sethi, S., Kathuria, M. and Kaushik, T., 2021. Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. *Journal of biomedical informatics*, *120*, p.103848.
3. Kumar, B.A. and Bansal, M., 2023. Face mask detection on photo and real-time video images using Caffe-MobileNetV2 transfer learning. *Applied Sciences*, *13*(2), p.935.
4. Li, Y., Guo, K., Lu, Y. and Liu, L., 2021. Cropping and attention based approach for masked face recognition. *Applied Intelligence*, *51*, pp.3012-3025.
5. Vibhuti, Jindal, N., Singh, H. and Rana, P.S., 2022. Face mask detection in COVID-19: a strategic review. *Multimedia tools and applications*, *81*(28), pp.40013-40042.
6. Shah, J.J., Ragu, H., David, V., Sasikumar, P. and Subburaj, M., 2024. OpenCV Based Customer Screening System for Prevention of COVID-19 Transmission in Retail Stores. *Wireless Personal Communications*, *137*(2), pp.685-703.
7. Dodda, R., Raghavendra, C., Azmera, C.N., Sreenu, M. and Nimmala, S., 2025. Real-Time Face Mask Detection Using Deep Learning: Enhancing Public Health and Safety. In *E3S Web of Conferences* (Vol. 616, p. 02013). EDP Sciences.
8. Diaz Barros, J.M., 2025. *Optimization and Generative Models for Face Analysis* (Doctoral dissertation, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau).
9. Myat Noe, S., Zin, T.T., Kobayashi, I. and Tin, P., 2025. Optimizing black cattle tracking in complex open ranch environments using YOLOv8 embedded multi-camera system. *Scientific Reports*, *15*(1), p.6820.
10. Sun, T., Fan, Q. and Shao, Y., 2025. Deep learning-based rebar detection and instance segmentation in images. *Advanced Engineering Informatics*, *65*, p.103224.
11. PrajnaSB. (n.d.) Observations. Available at:<https://github.com/prajnasb/observations> (Accessed: 04 March 2025).
12. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).
13. Canny, J., 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), pp.679-698.
14. Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). Ieee.
15. Viola, P. and Jones, M., 2001, December. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (Vol. 1, pp. I-I). Ieee.
16. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.
17. Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
18. Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.