International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(13s) 2026 pp. 482-493
https://doi.org/10.70917/ijcisim-2026-4085
Article

Real-Time Mask-wearing Detection on Edge Devices via Lightweight Convolutional Neural Networks

Myneni Madhu Bala[1*], Sreelakshmi Doma[2], Mohammad Riyaz Belgaum[3], Muniraju Naidu Vadlamudi[4], Mahdi H. Miraz[5,6,7]

[1]Department of CSE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India, E-mail: madhubala_m@vnrvjiet.in
[2]Department of ECE, Institute of Aeronautical Engineering, Hyderabad, India, E-mail: d.sreelakshmi@iare.ac.in
[3]Faculty of Computer Studies, Arab Open University-Bahrain, A'ali, Kingdom of Bahrain, Email: mohammad.riyaz@aou.org.bh
[4]Department of CSE, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad, India, E-mail: munirajunaidu@klh.edu.in
[5]School of Computing and Data Science, Xiamen University Malaysia, Malaysia, E-mail: m.miraz@ieee.org
[6]Faculty of Arts, Science and Technology, Wrexham University, UK, E-mail: m.miraz@ieee.org
[7]Faculty of Computing, Engineering and Science, University of South Wales, UK, E-mail: m.miraz@ieee.org

Abstract: Real-time mask-wearing compliance monitoring with autonomous surveillance plays a key role in decision-making in industrial, healthcare, and educational systems. Existing face mask-wearing monitoring systems often rely on computationally intensive or cloud-based models, making them unsuitable for accurate, low-latency, real-time deployment on resource-constrained mobile and edge devices, particularly for multiclass mask-wearing compliance detection. In this regard, the primary focus is on an efficient autonomous edge vision model with a lightweight architecture. The proposed lightweight edge vision framework addresses the limitations of existing cloud-dependent and computationally intensive face mask-wearing monitoring systems by enabling accurate, low-latency, real-time multiclass compliance detection on resource-constrained mobile and edge devices. It consists of real-time data pipeline design, MobileNetV2-based model development, hyperparameter optimization, and real-time performance evaluation. It is implemented with a lightweight CNN architecture optimized for resource-constrained environments, which classifies into three categories: correctly masked, unmasked, and improperly masked. This is demonstrated by deploying the model on real-time devices, such as mobile devices and camera modules, that capture video streams, thereby addressing challenges such as varied lighting conditions and facial orientations. The proposed system achieves a high accuracy of 99.35% while maintaining low latency, making it suitable for public health surveillance in crowded settings. The findings highlight the potential of edge-based AI in enhancing compliance with safety protocols in public spaces.
Keywords: Face mask-wearing detection, public health surveillance, MobileNetV2, Deep learning in mobile devices, COVID-19 compliance, PPE.

1. Introduction

In response to the growing demand for AI surveillance systems that rely on on-device vision applications in mobile and edge devices. This study evaluates MobileNetV2, a lightweight convolutional neural network architecture efficacy in the practical application of face mask-wearing compliance monitoring and warning in public spaces. The motivation for intelligent AI surveillance systems and on-device vision applications lies in their potential to enhance efficiency, privacy, and real-time responsiveness. It emphasizes the need for an automated diagnostic tool that continuously monitors whether individuals in public spaces are wearing masks. This tool, powered by artificial intelligence, can significantly contribute to public health measures and industrial restrictions by providing a streamlined and effective means of screening and enforcing mask-wearing protocols. The use of MobileNetV2 as the underlying technology aligns with the practical requirements of real-world scenarios, ensuring a balance between accuracy and computational efficiency. This technology not only enhances monitoring capabilities but also improves the overall effectiveness of public health measures, playing a crucial role in combating the spread of the coronavirus. Several recent studies have explored face mask-wearing detection, object detection, public surveillance, and intelligent real-time vigilance systems in public spaces.

The overview of related works on face mask-wearing screening reveals that Fan *et al.* [1] proposed an advanced, high-performance mask-wearing screening system to differentiate between good and poor mask-wearing coverage. RetinaFaceMask also contributed a new dataset with annotations. Chowdary *et al.* utilized the InceptionV3 architecture for face mask-wearing detection, incorporating image augmentation to enhance the dataset and transfer learning to improve efficiency. Sharma *et al.* proposed a CNN variant using VGG16 for mask-wearing identification, emphasizing the potential for further optimization. Vijitkunsawat *et al.* and Nowrin discovered various ML approaches (KNN, SVM, and CNN) to detect face masks, and concluded that CNN outperforms all. Sengupta *et al.* propose a novel outbreak response system that integrates AI and edge computing, utilizing YOLO for efficient mask-wearing detection and HRNetV2 for effective social distancing measures in public or workplace settings, thereby enhancing citizen-centric tracking and tracing of individuals who violate safety protocols. Lee *et al.* evaluate the performance of vision transformers, including DeiT, LeViT, and MobileViT, on embedded platforms. The accuracy and inference time of these models are evaluated using the ImageNet dataset on edge devices. Among all the LeViT, it showed the best performance in CPU-based edge devices. Loey *et al.* proposed a fusion approach for mask-wearing detection using both computer vision (ResNet50) and supervised learning (SVM, decision trees) algorithms. Singh *et al.* developed models using Faster CNN and YOLOv3, with the former exhibiting slightly better precision. While YOLOv3 demonstrated faster detection, both models lacked an alert system for improper mask-wearing and displayed limited information on face mask-wearing coverage. Jiang *et al.* introduced the SE YOLOv3 face mask-wearing detection model, comparing it to the traditional model in terms of accuracy. Baluprithviraj developed a smart door to screen people for mask-wearing compliance. Ramadhan *et al.* evaluated various face mask-wearing detection models utilizing the MaskedFace-Net dataset. The results of this study reveal that EfficientNetB4 stands out with the top performance in automated mask-wearing detection, showcasing its effectiveness in comparison to other evaluated models. Kumar *et al.* utilize a Caffe-MobileNetV2 model to detect faces in photos and real-time videos, emphasizing the facial features such as eyes, ears, nose, and forehead. The approach combines Caffe for face identification and MobileNetV2 for face mask-wearing recognition in real-time videos, including photos and scenes. Su *et al.* introduce an FMD method that combines Transfer Learning (TL) and Deep Learning (DL). The methodology is built upon YOLOv3, leveraging EfficientNet for feature extraction. The loss function is used to enhance face mask-wearing accuracy while minimizing the number of network parameters. The masks are classified into unqualified types (headscarves, masks (luffa and cotton)) and qualified types (one-use, medical, and N95). Naseri *et al.* utilized a technology fusion with IoT, an optimization function called Single Shot Multi-Box Detector to enhance face detection efficiency, ResNet for classification of different types, and MobileNet for learning rate optimization. Himeur *et al.* explore the application of advanced technologies, such as deep learning and computer vision, for ensuring safety in smart city public spaces. The study highlights the growth of research by emphasizing deep-learning techniques. Other notable technological advancements include augmented reality-based solutions for pose tracking with weak supervision. A visual analysis system that detects network anomalies through incremental learning will ensure robustness in dynamic environments . Facial emotion detection with a limited feature set for metaverse avatars and video games . Both spatial and temporal information at multiple scales in attention networks improve human action recognition in virtual reality (VR) settings . A novel attention pyramid transformer, designed to enhance image processing tasks such as image classification, segmentation, and enhancement, is employed to improve the computational efficiency and performance of the proposed model . A deep learning-based model is a tool for medical imaging used to detect anomalous pathological features in affected kidneys by glomerulonephritis . The synthesis of face sketches from real images with improved visual quality [24] was introduced using a deep learning model. The complement of the problem above has gained popularity in the present day as mask face detection with motion detection . Deep learning techniques are established to provide secure authentication in digital systems. This study encompasses existing datasets, recent advancements, the application of CNNs, and benchmarking results, and proposes future research directions to enhance safety in public areas of smart cities. Several studies have stated high-accuracy face mask-wearing detection models, which rely on cloud computing, binary classification, or computationally intensive architectures. There is a gap to focus on a lightweight multiclass mask-wearing compliance monitoring model deployed directly on mobile and edge devices with comprehensive hyperparameter evaluation.

By recognizing the significance of real-time face mask-wearing compliance monitoring in community gatherings, we proposed a methodology that employs MobileNetV2 for the implementation of an autonomous tool for real-time face mask-wearing compliance classification and notification. It is chosen for its suitability in resource-constrained environments, making it an ideal candidate for deployment on mobile and edge devices. This study focuses on integrating artificial intelligence, particularly MobileNetV2, as a face mask-wearing compliance monitoring and alert system, which holds promise in addressing the challenges associated with enforcing mask-wearing in public settings across diverse settings, ensuring the adaptability and applicability of on-device vision applications in diverse settings, including hospitals, schools, workplaces, airports, homes, malls, and transit stations. This wide applicability enhances safety measures by automatically detecting individuals who are not wearing their masks or wearing them improperly.

Our primary contributions are summarized as follows:

- Development of a lightweight, real-time face mask-wearing compliance monitoring system using MobileNetV2, optimized for deployment on mobile and edge devices. The proposed model supports efficient on-device vision processing with low computational overhead.
- Comprehensive hyperparameter optimization, including evaluation across different epoch settings (100, 200, and 300), enabling robust performance tuning and a deeper understanding of the model's generalization capabilities under varying training conditions.
- Achieving high detection accuracy with overfitting mitigation, as demonstrated by a peak accuracy of 99.35% across three mask-wearing classification categories: Mask, No_Mask, and Improper_Mask. The training strategy, which includes dropout regularization, effectively addresses overfitting and enhances real-world reliability.

The methodology, including the proposed real-time facemask-wearing compliance monitoring framework, is discussed in Section 2. The architecture of on-device model design, hyperparameter regulation, and evaluation measures is presented in Section 3. The experimental discoveries, along with performance analysis of hyperparameter tuning, are presented in Section 4. The conclusion is given in Section 5.

2. Methodology

2.1. Use Case and Dataset Overview

The rapid transmission of the COVID-19 virus through respiratory droplets has raised global concern and led to widespread disruptions. To contain the outbreak, many governments implemented unplanned lockdowns, resulting in significant economic and social consequences. In India, the second wave of COVID-19 resulted in approximately 43 million confirmed cases and over 0.52 million deaths. The country's high population density poses additional challenges to controlling the spread of the virus, thereby necessitating stringent safety protocols, such as social distancing, mask usage, regular hand sanitization, and vaccination. Sethi *et al.* highlight that consistent face mask usage can reduce daily mortality by 34–58%. To ensure broad compliance, especially in densely populated areas, surveillance systems are essential. These systems can combine individual accountability with automated, intelligent monitoring. As of January 2024, the emergence of the JN.1 sub-variant as the most prevalent global strain has reignited the urgency of preventive measures. India reported 605 new COVID-19 cases in a single day, while the U.S. experienced a surge, prompting a renewed emphasis on vaccination and the wearing of masks in public.

In this paper, we evaluate an edge vision model tailored for mobile and edge platforms using the MobileNetV2 architecture. The objective is to detect and classify face mask usage into three categories: Mask, No Mask, and Improperly Masked in real-time using camera feeds. The model is trained on a diverse dataset and integrated into a live video stream processing pipeline, making it suitable for deployment on embedded systems, such as smartphones or edge AI devices. The proposed approach incorporates essential stages for model development, including data collection, preprocessing, training, validation, and deployment. Figure 1 illustrates the schematic architecture of the proposed real-time face mask-wearing compliance monitoring system through surveillance, highlighting the real-time classification based on mask positioning.

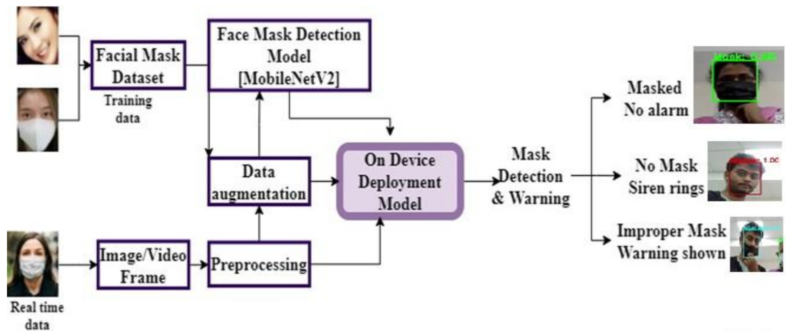


Figure 1. Proposed Edge Vision framework for Real-Time Mask-wearing Compliance Monitoring System through AI Surveillance.

2.2. Data Pipeline

The data pipeline was designed to perform an end-to-end image (vision) data process that spans real-time data collection, preprocessing, augmentation, and labelling.

2.2.1. Data Description

The proposed framework was evaluated using real-time face mask-wearing data collected from surveillance (live-edge sources), such as mobile and IP camera streams, as well as publicly available datasets. The primary dataset used for training and testing is the Kaggle Face Mask-wearing detection dataset. , which contains labeled images of faces with and without masks across various angles (frontal, side) and environmental settings. This dataset is organized into two primary classes: masked and unmasked faces, with 1,915 and 1,918 samples, respectively. Additionally, video frames captured from live camera feeds were incorporated for real-time validation, including Improper-mask images. , ensuring the model performs reliably under dynamic and unconstrained conditions.

2.2.2. Image Pre-processing

All images undergo a standardized preprocessing pipeline before training. First, each image is resized to 224×224 pixels, matching the input dimensions required by MobileNetV2. This resizing helps reduce computational cost and ensure consistency across samples. Images are then converted to numerical arrays, enabling them to be processed by deep learning models. Additionally, class label encoding is applied to convert categorical labels (e.g., "Mask", "No_Mask") into a numerical format compatible with the training process. Figure 2 shows examples of preprocessed input frames.

Figure 2. Real-time image samples before and after preprocessing

2.2.3. Data Augmentation

Data augmentation techniques are applied to enhance the model's generalization ability and mitigate potential overfitting resulting from limited data. These techniques simulate real-world variations by introducing transformations such as flipping, zooming, rotation, shifting, and shearing. The data augmentation includes rotation up to 20 degrees, zoom up to 15%, width and height shifts up to 20%, shear up to 15%, horizontal flipping enabled, and the flip mode set to "nearest."

2.2.4. Model Architecture

The face mask-wearing compliance monitoring model is built on the MobileNetV2 architecture, chosen for its lightweight design and efficiency on edge devices. The model uses average pooling to downsample feature maps and reduce spatial dimensions, improving detection speed while maintaining accuracy. The proposed framework was configured to support single-face and multi-face detection in crowded environments, enabling the identification and categorization of mask usage as Mask, No Mask, or Improper Mask. An integrated alert system provides real-time audio or visual cues for non-compliance.

3. Architecture of the Edge Vision Model

The proposed edge vision model architecture comprises various computing elements, including edge devices such as mobile phones and camera modules embedded with a lightweight, trained model. This architecture enables real-time face mask-wearing detection and categorization without dependence on cloud-based inference. Figure 3 illustrates the architectural overview of the proposed MobileNet V2-based edge vision system.

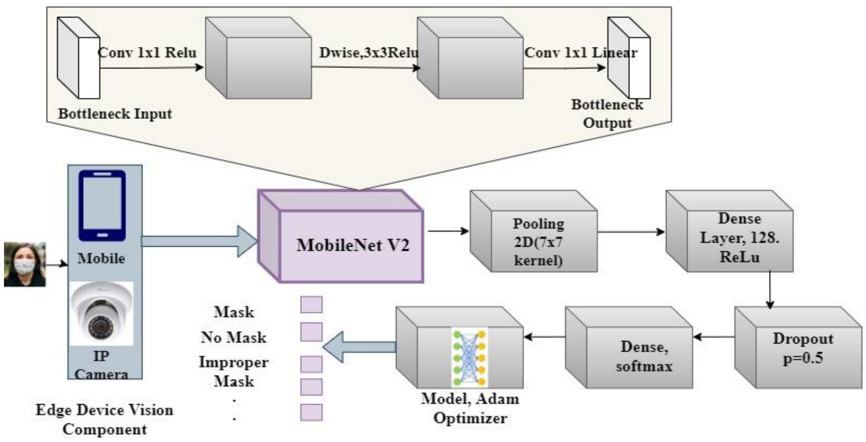


Figure 3. Architecture of the proposed MobileNet V2-based edge vision model.

In the proposed architecture, MobileNet V2 was used as the base model due to its lightweight, efficient design, which makes it well-suited for edge vision tasks. It is an efficient convolutional neural network (CNN) architecture optimized for real-time applications on resource-constrained edge devices. The architecture comprises inverted residual blocks with linear bottlenecks, where each block reduces the number of parameters by using depth-wise separable convolutions. These bottlenecks serve as efficient transformation units that compress and expand feature dimensions, thereby improving computational efficiency while preserving model accuracy. The model begins with a convolutional layer featuring 32 filters, followed by a series of 19 inverted residual bottleneck layers designed to refine feature representations progressively. During initialization, pre-trained weights from ImageNet are loaded into MobileNetV2, enabling the model to benefit from transfer learning and accelerating convergence during training. The detailed configuration of the model architecture is summarized in Table 1.

Table 1. MobileNetV2 design parameters

Layer	Type	Parameters / Configuration	Output Shape (approx.)
input_tensor	Input Layer	Shape = (224, 224, 3)	(224, 224, 3)

MobileNetV2	Base Model	Weights = 'imagenet', include_top=False	Depends on the architecture
average_pooling2d	AveragePooling2D	pool_size = (7, 7)	(1, 1, 1280)
flatten	Flatten	Name = 'flatten'	(1280, 1)
dense_1	Dense Layer	Units = 128, activation = 'relu'	(128,)
dropout	Dropout	Rate = 0.5	(128,)
dense_output	Dense Layer (Output)	Units = 3, activation = 'softmax'	(3,)

The 2D average pooling layer operates by sliding a fixed-size rectangular window across the input feature map and computing the mean value within each window. Each convolutional unit can be interpreted as a detector of a specific pattern or feature. The average pooling layer aggregates these patterns, yielding an averaged representation that highlights the general presence of features rather than the most prominent ones. The flatten layer converts a multidimensional input tensor (typically 2D or 3D) into a 1D vector, which is required for the subsequent fully connected (dense) layers. The dense layer, also known as a fully connected layer, is a fundamental building block in deep neural networks. It connects each neuron in the current layer to every neuron in the preceding layer, allowing for comprehensive feature integration across the network. Dropout is a regularization technique used in deep learning to prevent overfitting. It randomly deactivates a fraction of neurons during training, forcing the model to learn more robust, generalized features rather than relying on specific neurons. The model optimization was achieved through the Adam optimizer. A stochastic optimization algorithm is well-suited for large datasets and models with sparse gradients.

The model's performance was converged by tuning the key hyperparameters during training and optimization.

Table 2 presents a list of hyperparameters for optimizing model performance. Out of which, a few parameters, such as input and output size, are fixed at the training stage on the base model as 224×224 pixels with RGB channels, ensuring consistency in data representation. A learning rate of 1×10^{-4} is employed with a decay schedule proportional to the number of training epochs, facilitating gradual and stable convergence. Training is performed in batches of 32 samples to strike a balance between computational efficiency and training stability. A dropout rate of 30% is applied during training to overcome the overfitting problem. The Adam optimizer, known for its adaptive learning capabilities, is used to update model weights, while the categorical cross-entropy loss function guides the model in multi-class classification tasks.

Table 2. List of Hyperparameters and values considered for model performance evaluation

S. No	Hyperparameter	Values considered
1	Input image size	224x224 RGB image
2	Output image size	224x224 RGB image
3	Learning Rate (LR)	le-4 having DR (decay rate) = learning rate/no. of epoch
4	Batch size	32
5	Dropout rate value (evaluated)	0.2, 0.5, 0.8 and 0.9
6	Optimizer	ADAM
7	Loss function for multi-class	Categorical Cross-Entropy

4. Results with Discussions

The proposed real-time face mask-wearing compliance through autonomous surveillance at an edge vision application with MobileNetV2 architecture was implemented in two phases: design and construction of a vision data pipeline and a lightweight vision model that improves the reliability of edge vision applications under realistic and challenging scenarios commonly encountered in public spaces, such as varying lighting conditions, different camera angles, and diverse facial appearances. The evaluation of the model demonstrated the practicality of deploying edge vision applications for face-mask-wearing compliance monitoring and alerting using MobileNetV2. The model was developed using the Kaggle Open-Source RFMD dataset, which contains 1,915 face images with masks and 1,918 without masks. Various model parameters, including the number of epochs, dropout rates, and training sample sizes, were explored. Model evaluation considered metrics including accuracy, loss during training and deployment phases, F1-score, precision, recall, and support across different epoch values. Additionally, real-time validation on diverse test cases is thoroughly discussed to demonstrate the model's effectiveness in practical scenarios. The model was trained for 300 epochs using a high dropout rate of 0.90 while maintaining 80% of the dataset for training. The aim was to assess the impact of such a high dropout rate on accuracy and to evaluate its effectiveness in preventing overfitting. Table 3 presents the key evaluation metrics of the proposed model, trained with 300 epochs and a dropout rate of 0.9, on both the training and validation datasets.

Table 3. Evaluation of the model on training and validation data for 300 epochs and a dropout rate of 0.9

Epoch Number	T_Loss	T_Accuracy	V_Loss	V_Accuracy
60	0.045	0.960	0.038	0.990
120	0.031	0.975	0.034	0.992
180	0.028	0.978	0.030	0.993
240	0.025	0.980	0.028	0.994
300	0.023	0.982	0.026	0.993

Table 4 presents the evaluation of the model's performance across various mask-wearing positions after training for 300 epochs with a dropout rate of 0.90. The results demonstrate that the previously observed overfitting issues have been effectively addressed. The proposed network architecture is designed not only to memorize training data but also to generalize well, enabling more accurate identification of different classes. Moreover, the evaluation highlights the model's optimal fit, confirming the successful mitigation of overfitting and improved generalization across diverse mask-wearing positions. Table 5 presents a comparison of dropout value and accuracy, along with observations on its impact on overfitting.

Table 4. Evaluation of face-mask-wearing compliance monitoring model performance

Epochs	Dropout	Class Label	Precision	Recall	F1-score	Support
300	0.2	Mask	0.90	1.00	0.95	383
		Unmasked	1.00	0.90	0.95	385
		improperly masked	0.90	1.00	0.95	383
200	0.8	Mask	0.91	0.89	0.90	575
		Unmasked	0.89	0.91	0.90	575
		improperly masked	0.91	0.89	0.90	575
100	0.5	Mask	0.89	0.90	0.90	575
		Unmasked	0.91	0.88	0.89	575
		improperly masked	0.89	0.90	0.90	575

Table 5. Comparison of Dropout Values and Accuracy

Dropout Value	Accuracy	Observations
0.2	100.00%	Highly Overfitting Observed
0.5	99.92%	Moderate Overfitting Observed

0.8	99.85%	Overfitting reduced considerably
0.9	**99.35%**	**The model shows a good fit.**

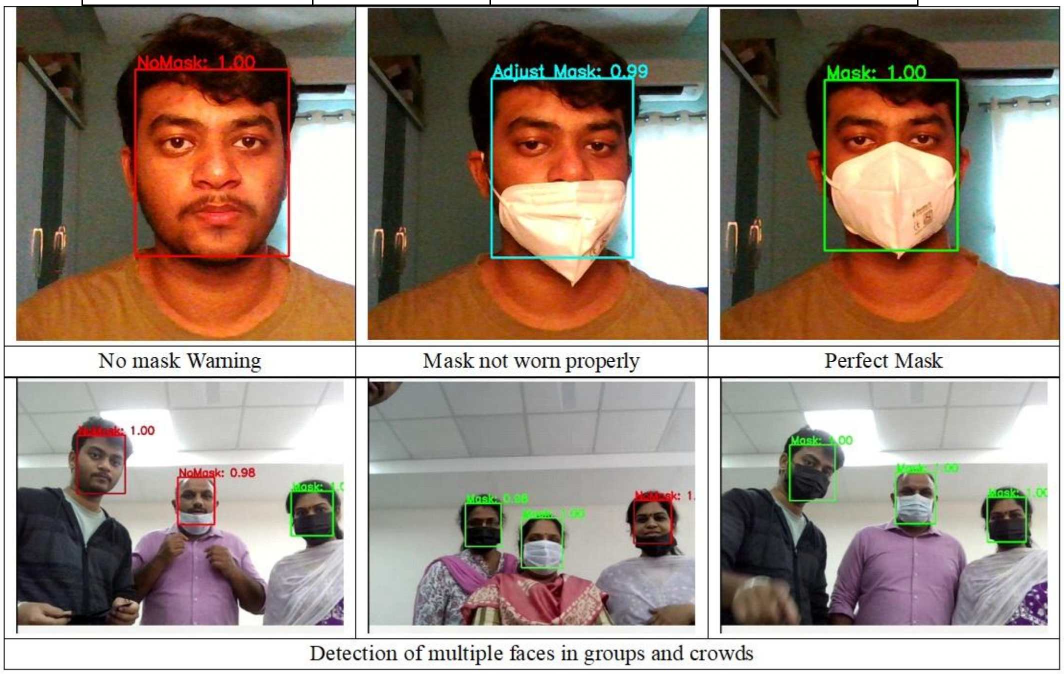


Figure 4. Real-time face mask-wearing compliance monitoring evaluation with multiple faces in groups and a crowd.

The real-time edge face detection model has been evaluated using various test cases to assess its performance, accuracy, and efficiency in different public-area scenarios. Figure 4 depicts a real-time edge evaluation with multiple faces in groups and crowds. The test cases considered for evaluation are:

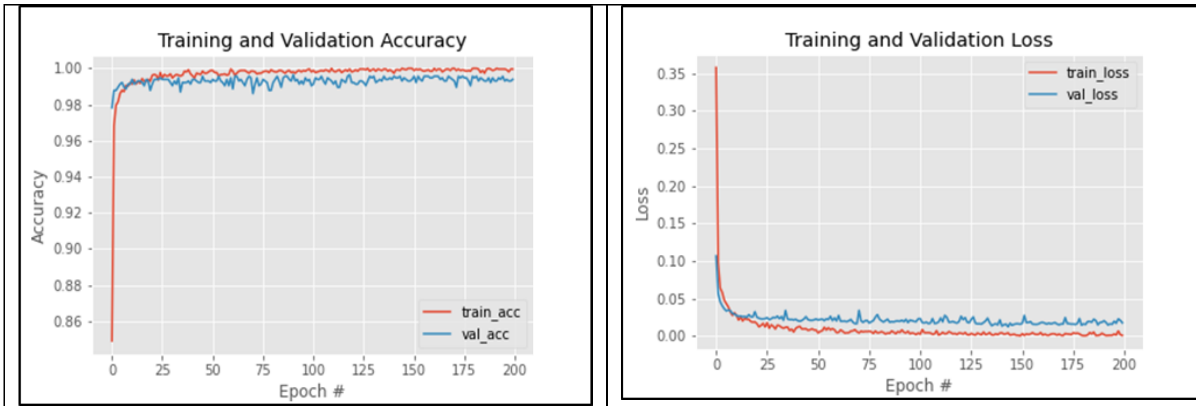
Case 1: No Mask: People not wearing masks are being alerted by an alarm system.

Case 2: Improper Mask: Using percentage accuracy, the system will alert the person based on the mask-wearing position, such as when it is worn below the nose or lips.

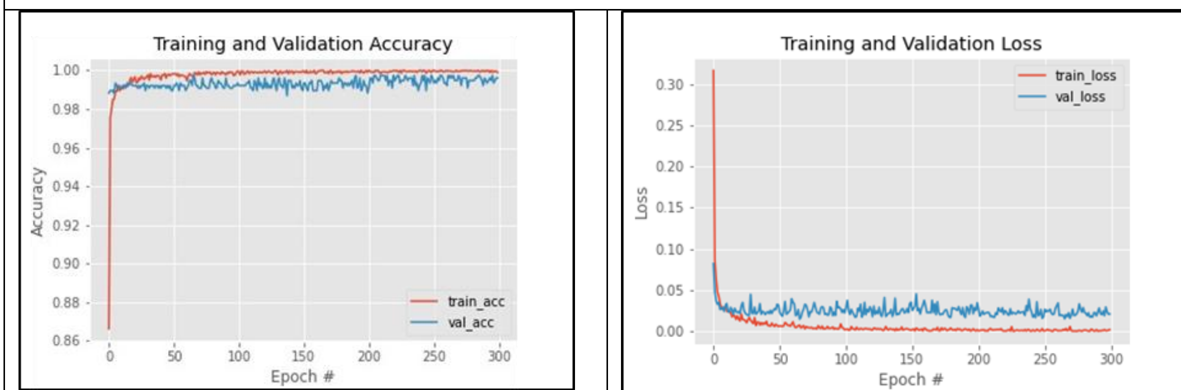
Case 3: Perfect Mask: In this scenario, the model gives no alert and allows the person to pass freely, as the mask is worn perfectly and protocol is being followed.

Case 4: Detection in Groups and Crowds: Many scenarios where people are found meeting can be put under surveillance and detected for violating the protocol.

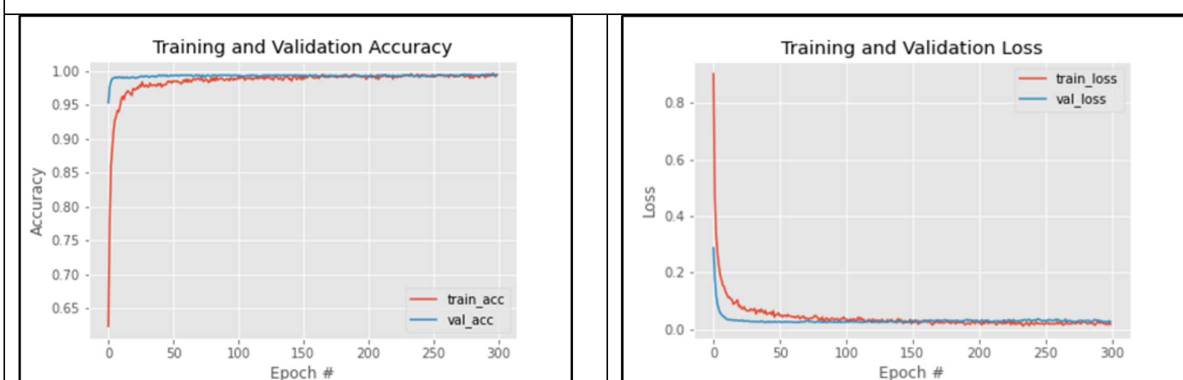
Thus, in real-time scenarios of a wider use case, including public areas such as malls, theaters, hospitals, clinics, bus stops, and airports, the primary source of contamination is often identified. Thus, the current model not only successfully detects multiple faces but also identifies faces in a crowd. Figure 5 illustrates the training and validation phase model accuracy/loss rate at variant epochs (200 and 300) and with a dropout of (0.2, 0.8, and 0.9).



Training and Validation Accuracy/Loss for 200 epochs with dropout of 0.8.



Training and Validation Accuracy/Loss for 300 epochs for dropout of 0.2



Training and Validation Accuracy/Loss for 300 epochs with dropout of 0.90

Figure 5. Model evaluation under variant epochs with dropout rate conditions

A comparison of the evaluation of various face mask-wearing compliance monitoring and warning models' performance is summarized in Table 6. Findings contribute valuable information to the broader discourse on leveraging edge computing for real-time visual processing in public health and safety applications. The design is suitable for mobile and edge devices, making them powerful tool for on-device vision applications. This paves the way for advancements in real-world applications that prioritize efficiency, privacy, and public safety.

Table 6. Comparison of the evaluation of the face mask-wearing detection model performance.

Ref. No.	Model	No. of Classes	Accuracy
[30]	VGG-16 CNN-based model	2 classes (not wearing /wearing a mask)	96.00%
[31]	CNN	2 classes (not wearing /wearing a mask)	94.20%
[32]	CNN	2 classes (not wearing /wearing a mask)	91.11%
[8]	ResNet50	2 classes (masked and unmasked face)	99.00%
[13]	Caffe-MobileNetV2	2 classes (With and without mask)	99.64%
[10]	Efficient-Yolov3	2 classes (Qualified and Unqualified mask types)	97.84%
Proposed	MobileNetV2	3 classes (With_Mask, No_Mask and Improper_Mask)	99.35%

5. Conclusion

The proposed AI surveillance framework enables real-time, on-device face mask-wearing compliance monitoring for mobile and edge devices. It is demonstrated with appropriate position finding and warning using MobileNetV2, a lightweight deep learning architecture well-suited for deployment in real-time systems. The proposed multiclass model achieved a remarkably high accuracy of 99.35% across three practical classes: With Mask, No Mask, and Improper Mask. Its high accuracy, efficiency, and adaptability contribute significantly to the broader landscape of public health monitoring and safety compliance. The improved accuracy, efficiency, and adaptability make it a valuable tool in the wider landscape. Overall, this contributes to the ongoing efforts to leverage technology to automate adherence to protocols, particularly in the context of face mask-wearing compliance.

Funding: There is no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this work is freely available and can be downloaded from the following link: https://www.kaggle.com/datasets/andrewmvd/face-mask-detection.

Conflicts of Interest: There are no conflicts of interest present in this work.

References

1. Fan X, Jiang M. RetinaFaceMask: A Single Stage Face Mask Detector for Assisting Control of the COVID-19 Pandemic. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE; 2021, p. 832–7. https://doi.org/10.1109/SMC52423.2021.9659271.
2. Jignesh Chowdary G, Punn NS, Sonbhadra SK, Agarwal S. Face Mask-wearing detection Using Transfer Learning of InceptionV3. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12581 LNCS, 2020, p. 81–90. https://doi.org/10.1007/978-3-030-66665-1_6.
3. Sharma A, Gautam R, Singh J. Deep learning for face mask-wearing detection: a survey. Multimed Tools Appl 2023;82:34321–61. https://doi.org/10.1007/s11042-023-14686-6.
4. 4] Vijitkunsawat W, Chantngarm P. Study of the Performance of Machine Learning Algorithms for Face Mask-wearing detection. 2020 - 5th International Conference on Information Technology (InCIT), IEEE; 2020, p. 39–43. https://doi.org/10.1109/InCIT50588.2020.9310963.

5. Nowrin A, Afroz S, Rahman MdS, Mahmud I, Cho Y-Z. Comprehensive Review on Facemask-wearing detection Techniques in the Context of Covid-19. IEEE Access 2021;9:106839–64. https://doi.org/10.1109/ACCESS.2021.3100070.
6. Sengupta K, Srivastava PR. HRNET: AI-on-Edge for Mask-wearing detection and Social Distancing Calculation. SN Comput Sci 2022;3:157. https://doi.org/10.1007/s42979-022-01023-1.
7. Lee M, Lee S, Kim T. Performance evaluation of efficient vision transformers on embedded edge platforms. IEMEK Journal of Embedded Systems and Applications 2023;18:89–100.
8. Loey M, Manogaran G, Taha MHN, Khalifa NEM. A hybrid deep transfer learning model with machine learning methods for face mask-wearing detection in the era of the COVID-19 pandemic. Measurement 2021;167:108288. https://doi.org/10.1016/j.measurement.2020.108288.
9. Singh S, Ahuja U, Kumar M, Kumar K, Sachdeva M. Face mask-wearing detection using YOLOv3 and faster R-CNN models: COVID-19 environment. Multimed Tools Appl 2021;80:19753–68. https://doi.org/10.1007/s11042-021-10711-8.
10. Jiang X, Gao T, Zhu Z, Zhao Y. Real-Time Face Mask-wearing detection Method Based on YOLOv3. Electronics (Basel) 2021;10:837. https://doi.org/10.3390/electronics10070837.
11. Baluprithviraj KN, Bharathi KR, Chendhuran S, Lokeshwaran P. Artificial Intelligence based Smart Door with Face Mask-wearing detection. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), IEEE; 2021, p. 543–8. https://doi.org/10.1109/ICAIS50930.2021.9395807.
12. Ramadhan MV, Muchtar K, Nurdin Y, Oktiana M, Fitria M, Maulina N, et al. Comparative analysis of deep learning models for detecting face mask. Procedia Comput Sci 2023;216:48–56. https://doi.org/10.1016/j.procs.2022.12.110.
13. Kumar BA, Bansal M. Face Mask-wearing detection on Photo and Real-Time Video Images Using Caffe-MobileNetV2 Transfer Learning. Applied Sciences 2023;13:935. https://doi.org/10.3390/app13020935.
14. Su X, Gao M, Ren J, Li Y, Dong M, Liu X. Face mask-wearing detection and classification via deep transfer learning. Multimed Tools Appl 2022;81:4475–94. https://doi.org/10.1007/s11042-021-11772-5.
15. Naseri RAS, Kurnaz A, Farhan HM. Optimized face detector-based intelligent face mask-wearing detection model in IoT using deep learning approach. Appl Soft Comput 2023;134:109933. https://doi.org/10.1016/j.asoc.2022.109933.
16. Himeur Y, Al-Maadeed S, Varlamis I, Al-Maadeed N, Abualsaud K, Mohamed A. Face Mask-wearing detection in Smart Cities Using Deep and Transfer Learning: Lessons Learned from the COVID-19 Pandemic. Systems 2023;11:107. https://doi.org/10.3390/systems11020107.
17. Sun R, Yu X, Feng H, Wang F, Zhang X. Motion-robust mask face presentation attack detection via dual-stream texture-rPPG network. Vis Comput 2025;41:4517–32. https://doi.org/10.1007/s00371-024-03675-x.
18. Hu X, Bao X, Wei G, Li Z. Human-pose estimation based on weak supervision. Virtual Reality & Intelligent Hardware 2023;5:366–77. https://doi.org/10.1016/j.vrih.2022.08.010.
19. Tian X, Wu Z, Cao J, Chen S, Dong X. ILIDViz: An Incremental Learning-Based Visual Analysis System for Network Anomaly Detection. Virtual Reality & Intelligent Hardware 2023;5:471–89. https://doi.org/10.1016/j.vrih.2023.06.009.
20. Bellenger D, Chen M, Xu Z. Facial emotion recognition with a reduced feature set for video game and metaverse avatars. Comput Animat Virtual Worlds 2024;35. https://doi.org/10.1002/cav.2230.
21. Xiao Z, Chen Y, Zhou X, He M, Liu L, Yu F, et al. Human action recognition in immersive virtual reality based on multi-scale spatio-temporal attention network. Comput Animat Virtual Worlds 2024;35. https://doi.org/10.1002/cav.2293.
22. Lin X, Sun S, Huang W, Sheng B, Li P, Feng DD. EAPT: Efficient Attention Pyramid Transformer for Image Processing. IEEE Trans Multimedia 2023; 25:50–61. https://doi.org/10.1109/TMM.2021.3120873.
23. Ali SG, Wang X, Li P, Li H, Yang P, Jung Y, et al. EGDNet: an efficient glomerular detection network for multiple anomalous pathological feature in glomerulonephritis. Vis Comput 2025;41:2817–34. https://doi.org/10.1007/s00371-024-03570-5.
24. Li P, Sheng B, Chen CLP. Face Sketch Synthesis Using Regularized Broad Learning System. IEEE Trans Neural Netw Learn Syst 2022;33. https://doi.org/10.1109/TNNLS.2021.3070463.
25. Hu Y, Xu Y, Zhuang H, Weng Z, Lin Z. Machine Learning Techniques and Systems for Mask-Face Detection—Survey and a New OOD-Mask Approach. Applied Sciences (Switzerland) 2022;12. https://doi.org/10.3390/app12189171.
26. Turhal U, Günay Yılmaz A, Nabiyev V. A new face presentation attack detection method based on face-weighted multi-color multi-level texture features. Vis Comput 2024; 40:1537–52. https://doi.org/10.1007/s00371-023-02866-2.
27. Sethi S, Kathuria M, Kaushik T. Face mask-wearing detection using deep learning: An approach to reduce risk of Coronavirus spread. J Biomed Inform 2021;120:103848. https://doi.org/10.1016/j.jbi.2021.103848.
28. Face Mask-wearing detection, https://www.kaggle.com/datasets/andrewmvd/face-mask-detection, Accessed on 02 January 2025. n.d.
29. Facemask real-time stream test data, https://github.com/Madhubala2024/faces-with-mask, Accessed on 02 January 2025. n.d.
30. Nanda SK, Ghai D, Pande S. VGG-16-Based Framework for Identification of Facemask Using Video Forensics, 2022, p. 673–85. https://doi.org/10.1007/978-981-16-6285-0_54.
31. Lad AM, Mishra A, Rajagopalan A. Comparative Analysis of Convolutional Neural Network Architectures for Real Time COVID-19 Facial Mask-wearing detection. J Phys Conf Ser 2021; 1969:012037. https://doi.org/10.1088/1742-6596/1969/1/012037.

32. Hussain GKJ, Priya R, Rajarajeswari S, Prasanth P, Niyazuddeen N. The Face Mask-wearing detection Technology for Image Analysis in the Covid-19 Surveillance System. J Phys Conf Ser 2021;1916:012084. https://doi.org/10.1088/1742-6596/1916/1/012084.