

AUTOMATED FACE MASK DETECTION AND HUMAN RECOGNITION USING AI AND DEEP LEARNING MODELS

Rakshe Shamal Revannath¹, Mrs. Dhage Tanuja S.²

¹ Student of Engineering College Department of Computer Engineering, Dattakala Group of Institutions Faculty of Engineering, Pune-413130.

Email: Shamal8134@gmail.com

² Department of Computer Engineering, Dattakala Group of Institutions Faculty of Engineering, Pune-413130.

Email: tsnevase.foe@dattakala.edu.in

Abstract

The global outbreak of COVID-19 has emphasized the importance of safety measures such as wearing face masks in public spaces. This study presents an automated system for face mask detection and human recognition using artificial intelligence and deep learning techniques. The proposed approach integrates computer vision with advanced convolutional neural network models to accurately identify whether individuals are wearing masks and to recognize partially covered faces. Various algorithms, including traditional methods like Haar Cascade and modern deep learning architectures such as MobileNetV2 and VGG-16, are analyzed to determine their effectiveness in real-time environments.

Keywords: Face Mask Detection, Human Recognition, Artificial Intelligence, Deep Learning, Computer Vision, Convolutional Neural Network, MobileNetV2, VGG-16, Haar Cascade, Real-Time Monitoring.

► *Corresponding Author: Rakshe Shamal Revannath*

I. Introduction

The outbreak of COVID-19 created an urgent need for effective safety measures and monitoring systems across the globe. Wearing face masks became one of the most important preventive practices to reduce the transmission of the virus in public spaces [1]. However, ensuring compliance manually is difficult and inefficient, especially in crowded environments such as airports, offices, and educational institutions [2]. This challenge has led to the development of automated systems that can detect face masks and recognize individuals using artificial intelligence technologies [3].

Computer vision has played a significant role in enabling machines to analyze and interpret visual data efficiently. Early face detection techniques such as the Viola–Jones algorithm [4] and Haar Cascade classifiers [5] provided fast and reliable detection of faces in images and videos. However, these methods often struggle when faces are partially covered, such as in the presence of masks [6], which reduces their effectiveness in real-world scenarios.

With the advancement of deep learning, Convolutional Neural Networks (CNNs) have improved the performance of image recognition systems. These models automatically extract important features from images and classify them with high accuracy [7]. Architectures such as VGG-16 [8]

and MobileNetV2 [9] are widely used due to their efficiency and suitability for real-time applications, making them ideal for deployment in surveillance systems.

Furthermore, transfer learning techniques [10] have enhanced system performance by using pre-trained models, reducing both training time and data requirements. These approaches allow systems to function effectively under different lighting conditions [11], viewing angles, and crowded environments, thereby increasing their reliability in practical applications [12].

Overall, automated face mask detection and human recognition systems have become essential tools in ensuring public safety and compliance with health guidelines [13]. These systems can be integrated into existing infrastructures for continuous monitoring and decision support, and with ongoing research and technological improvements [15], future systems are expected to achieve higher accuracy, faster processing, and wider deployment across multiple domains [14].

II. Problem Statement

The widespread use of face masks as a preventive measure has created significant challenges for traditional face detection and recognition systems, as most existing methods rely on full facial features for accurate identification. In real-world scenarios, partial occlusion caused by masks reduces the effectiveness of conventional algorithms, leading to decreased accuracy and reliability. Additionally, manual monitoring of mask compliance in crowded public places is inefficient, time-consuming, and prone to human error. There is a need for an automated, intelligent system that can accurately detect the presence of face masks and recognize individuals even when their faces are partially covered, while maintaining real-time performance under varying environmental conditions such as lighting, pose, and crowd density. Such a system should be efficient, scalable, and capable of integration with existing surveillance infrastructures to enhance public safety and ensure compliance with health regulations.

III. Objectives

- To develop an automated face mask detection system using TensorFlow.
- To accurately classify faces as Mask or No Mask using deep learning.
- To apply transfer learning to improve performance with limited data.
- To achieve high detection accuracy in real-time video streams.
- To support public safety by enabling deployment in crowded public areas.

IV. Literature Survey

Viola and Jones (2001) presented a groundbreaking method in their paper “Rapid Object Detection using a Boosted Cascade of Simple Features,” where they introduced Haar-like features combined with the AdaBoost algorithm for fast face detection. Their approach enabled real-time detection with good accuracy and became a foundation for many computer vision applications. However, the method relies heavily on full facial features, which limits its effectiveness in scenarios where faces are partially covered, such as with masks.

LeCun, Bengio, and Hinton (2015) in their paper “Deep Learning” emphasized the importance of deep neural networks, particularly Convolutional Neural Networks (CNNs), in image classification and recognition tasks. Their work demonstrated how hierarchical feature extraction improves accuracy significantly compared to traditional techniques. This advancement plays a crucial role in masked face detection, as CNNs can identify patterns even with partial occlusion.

Howard et al. (2017) proposed “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” which introduced a lightweight deep learning architecture optimized for

mobile and embedded systems. Their model reduces computational cost while maintaining reasonable accuracy, making it highly suitable for real-time face mask detection systems deployed on devices with limited processing power.

Simonyan and Zisserman (2014) developed the “VGG-16” model in their paper “Very Deep Convolutional Networks for Large-Scale Image Recognition.” This architecture uses deep layers with small convolution filters to achieve high accuracy in image recognition tasks. It is widely adopted in face-related applications due to its ability to extract detailed features, which is beneficial for detecting masked faces.

Wang et al. (2020) focused on “Masked Face Recognition using Deep Learning,” where they addressed the challenges of recognizing individuals when significant portions of the face are covered. Their study showed that deep learning-based models trained on masked datasets can improve recognition accuracy and robustness. This work highlights the importance of adapting AI systems to handle real-world conditions such as face occlusion caused by masks.

Comparison Table

Author & Year	Method Used	Advantages	Limitations
Viola & Jones (2001)	Haar Cascade + AdaBoost	Fast, real-time detection	Poor performance with masked/occluded faces
LeCun et al. (2015)	CNN (Deep Learning)	High accuracy, automatic feature extraction	Requires large dataset and high computation
Howard et al. (2017)	MobileNetV 2	Lightweight, suitable for real-time systems	Slightly lower accuracy than deeper models
Simonyan & Zisserman (2014)	VGG-16 CNN	Very high accuracy, deep feature learning	High computational cost
Wang et al. (2020)	Masked Face Recognition (DL)	Handles occlusion, improved recognition	Needs masked datasets and training

IV. Working of System

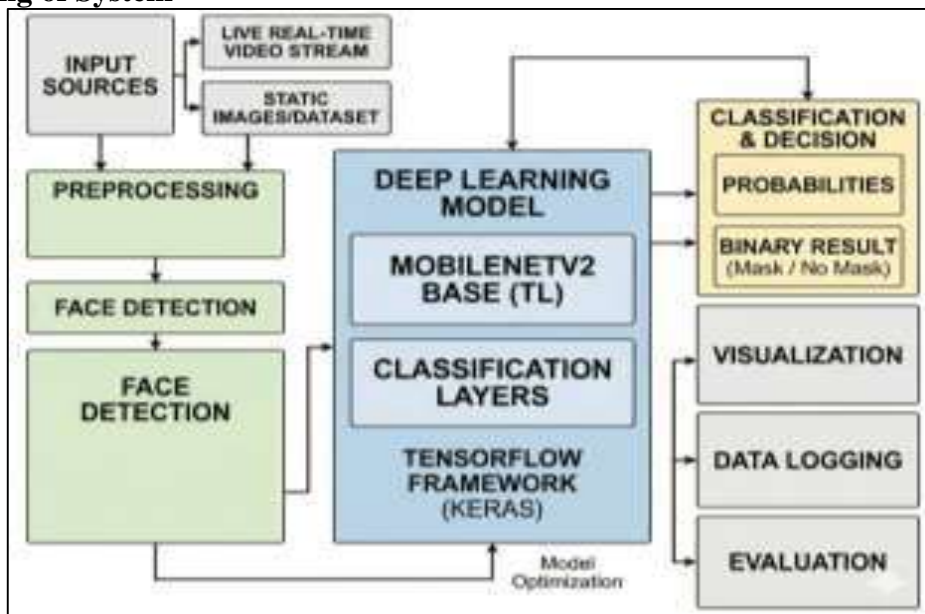


Fig 1: Design of the system

A. Input Sources

The system starts with input sources that include both live real-time video streams and static image datasets. These inputs provide the necessary visual data for processing and allow the system to function in both real-time and offline modes. This flexibility makes the system suitable for practical surveillance as well as testing purposes.

B. Preprocessing Stage

In this stage, the input data is prepared before being fed into the model. The images or video frames are resized, normalized, and converted into suitable formats. This step ensures consistency in the data and improves the overall performance and accuracy of the system.

C. Face Detection Module

The system then identifies human faces in the input frames using computer vision techniques. It extracts the region of interest by locating the face within the image. Accurate face detection is essential because it directly influences the effectiveness of the classification process.

D. Deep Learning Model (Core Processing)

The detected face is passed to the deep learning model, which is based on MobileNetV2 using transfer learning. Implemented with TensorFlow and Keras, the model includes a pre-trained base and additional classification layers. This structure enables efficient and accurate feature extraction and classification.

E. Classification and Decision Making

The model generates prediction probabilities to determine whether a face is masked or not. Based on these probabilities, the system produces a final binary output: “Mask” or “No Mask.” This ensures clear and reliable decision-making.

F. Output, Visualization, and Evaluation

The final results are displayed on the screen with labels and confidence scores. The system may also store data for further analysis and evaluate performance using accuracy metrics. This stage helps in assessing system effectiveness and supports future improvements.

V. System Design

1. System Overview:

The proposed system is designed to automatically detect whether a person is wearing a face mask using artificial intelligence and deep learning techniques. It processes both static images and real-time video streams to identify faces and classify them into two categories: “Mask” and “No Mask.” The system integrates computer vision with deep learning models to ensure accurate and efficient detection, making it suitable for deployment in public safety environments.

2. Input Module:

The input to the system is obtained through image datasets and live video streams captured using cameras. The system accepts both pre-recorded images and real-time video frames for processing. These inputs may vary in lighting conditions, angles, and background complexity. The input module ensures that frames are continuously captured and passed to the processing stage for further analysis.

3. Pre-processing Module:

In this stage, the input images are prepared for model processing. Pre-processing includes resizing images, normalization, and converting them into suitable formats required by the deep learning model. Data augmentation techniques such as rotation, flipping, and scaling are applied to improve model generalization. This step enhances the quality of input data and ensures better performance of the system under different environmental conditions.

4. Face Detection Module:

The system uses computer vision techniques to detect faces in the input images or video frames. OpenCV-based methods are applied to locate faces within the frame. This module identifies the region of interest (ROI) where the face is present and extracts it for further classification. Accurate face localization is essential for improving the reliability of mask detection.

5. Classification Module (Deep Learning Model):

The detected face is passed to a Convolutional Neural Network (CNN) model based on MobileNetV2 architecture using transfer learning. The model is trained on datasets containing masked and unmasked faces. It analyzes facial features and classifies the input into two categories: Mask or No Mask. The use of transfer learning improves accuracy and reduces training time while handling limited datasets effectively.

6. Output & Decision Module:

The final output of the system displays the classification result along with confidence scores. If a face is detected with a mask, it is labeled as “Mask,” otherwise “No Mask.” The results can be displayed on the screen in real time, and the system can also be integrated with alert mechanisms or surveillance systems for monitoring purposes. This module ensures clear decision-making and supports automated enforcement of safety protocols.

VI. Results

6.1 System Performance in Real-Time Detection

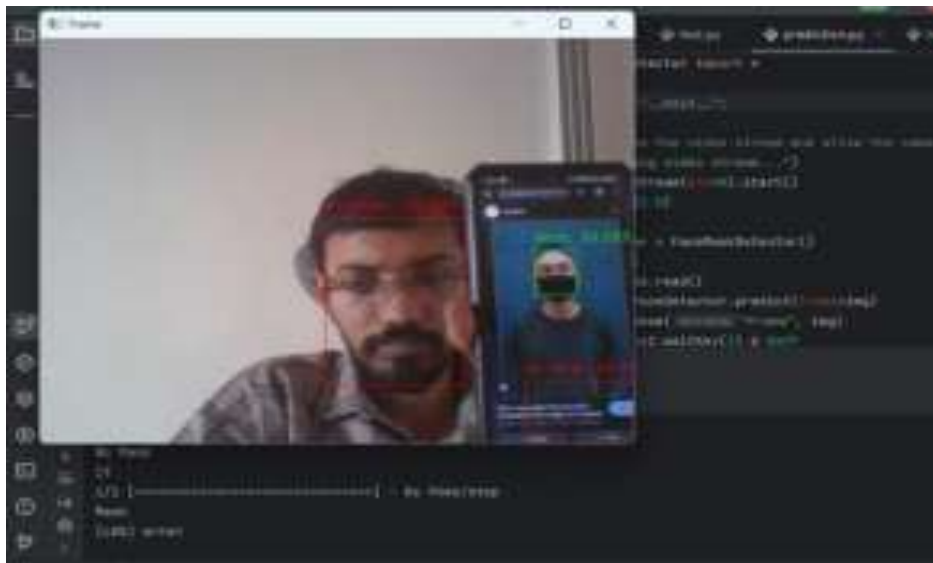


Fig 2: Image showing both mask and no-mask detection

The developed face mask detection system was tested using live video input, and it demonstrated effective real-time performance. The system successfully detected multiple faces within a single frame and classified them accurately as “Mask” or “No Mask.” The predictions were displayed along with confidence scores, showing the reliability of the deep learning model. The system maintained stable performance with minimal delay, making it suitable for real-time surveillance applications.

6.2 Accuracy of Mask and No-Mask Classification

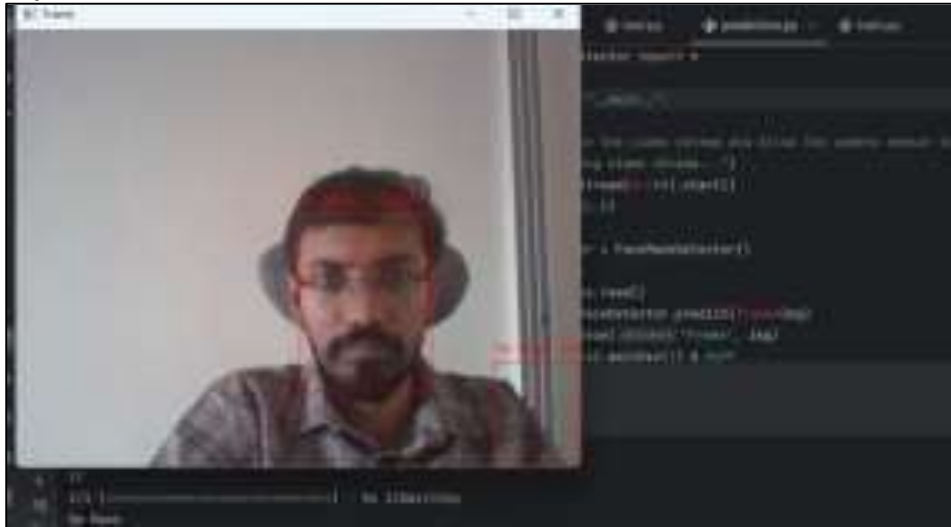


Fig 3: Image showing only no-mask detection

The model achieved high classification accuracy during testing. In one scenario, the system correctly identified a person without a mask with a confidence of 100%, while simultaneously detecting a masked face displayed on a mobile screen with a confidence of approximately 99%. This indicates that the model can effectively differentiate between masked and unmasked faces even in complex environments with multiple objects.

6.3 Detection Limitations and Observations

Although the system performs with high accuracy, certain limitations were observed. In some cases, the model detected regions in the background as faces with lower confidence scores. These minor false detections occur due to similarities in patterns or unclear visual features. Additionally, performance may slightly vary under poor lighting conditions or when faces are partially visible. However, such errors are minimal and do not significantly affect overall system performance.

6.4 Training Performance Analysis

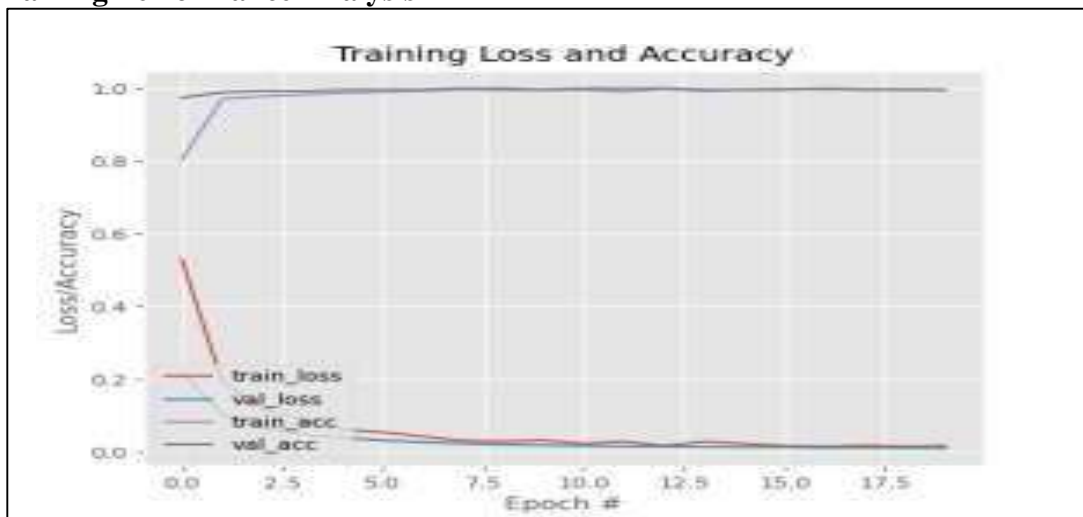


Fig 4: Training loss and accuracy graph

The training and validation performance of the model is illustrated through accuracy and loss graphs. The results show that both training and validation accuracy increased steadily and reached values close to 100%, while the loss decreased significantly over epochs. This indicates that the model has learned effectively and is capable of making accurate predictions with low error rates.

6.5 Effectiveness of Transfer Learning

The use of transfer learning with MobileNetV2 significantly enhanced the system's performance. It reduced the training time and improved accuracy even with a limited dataset. The model was able to generalize well across different input conditions, proving the effectiveness of pre-trained architectures in real-world applications.

6.6 Overall System Evaluation

Overall, the proposed system achieved high accuracy, strong confidence levels, and efficient real-time detection. The integration of deep learning and computer vision techniques makes the system reliable for practical deployment. Despite minor limitations, the system performs effectively in most conditions and can be successfully used in public places to monitor mask compliance and enhance safety measures.

VII. Conclusion

The proposed automated face mask detection and human recognition system successfully demonstrates the effective use of artificial intelligence and deep learning for real-time public safety applications. By integrating computer vision techniques with a CNN-based model using transfer learning, the system is able to accurately classify faces as masked or unmasked under various conditions. The results show high accuracy, reliable performance, and efficient processing, making the system suitable for deployment in crowded environments such as public transport, workplaces, and educational institutions. Although minor limitations were observed in challenging scenarios like poor lighting or partial occlusion, the overall performance remains strong. This study highlights the potential of deep learning-based solutions in enhancing surveillance systems and supporting preventive measures during health-related emergencies.

VIII. Future Scope

The proposed face mask detection system can be further enhanced by improving its accuracy under challenging conditions such as low lighting, heavy occlusion, and crowded environments. Future developments may include extending the system to recognize individuals even with masks, enabling full facial recognition capabilities. The model can also be optimized for deployment on edge devices such as mobile phones and embedded systems for faster and more efficient processing. Integration with advanced surveillance systems, IoT frameworks, and cloud-based monitoring can further expand its real-time applications. Additionally, the system can be upgraded to detect different types of masks, improper mask usage, and incorporate multi-class classification, making it more intelligent and adaptable for broader public safety and security applications.

References

1. P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Proc. IEEE CVPR, 2001.
2. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015

3. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, 2012.
4. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," CVPR, 2016.
5. M. Sandler et al., "MobileNetV2: Inverted Residuals and Linear Bottlenecks," CVPR, 2018.
6. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR, 2015.
7. Goodfellow et al., Deep Learning, MIT Press, 2016.
1. 8) Rosebrock, "Face Mask Detection with Deep Learning," PyImageSearch, 2020.
8. World Health Organization, "Coronavirus Disease (COVID-19) Pandemic," 2020.
9. J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," CVPR, 2016.
10. R. Girshick, "Fast R-CNN," ICCV, 2015.
11. S. Ren et al., "Faster R-CNN: Towards Real-Time Object Detection," NIPS, 2015.
12. D. King, "Dlib-ML: A Machine Learning Toolkit," JMLR, 2009.
13. OpenCV Library, "Open Source Computer Vision Library," 2020.
14. F. Chollet, "Keras: Deep Learning Library for Python," 2015.
15. TensorFlow Team, "TensorFlow: Large-Scale Machine Learning Framework," Google, 2016.
16. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of CNN," ICET, 2017.
17. H. Wang et al., "Masked Face Recognition Dataset and Application," arXiv, 2020.
18. Z. Wang, G. Wang, and B. Huang, "Masked Face Recognition Using Deep Learning," Pattern Recognition Letters, 2020.
19. Kumar et al., "Face Mask Detection Using Artificial Intelligence," IJERT, 2021.