International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Innovations in Computer Science and Applications
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Revealing and Classification of Face Mask Detection using Artificial Intelligence and Machine Learning

Avinash Sahebrao Maske

Department of Science and Technology,
G. H. Raisoni Skill Tech University, Nagpur, Maharashtra, India

Abstract

The rapid global spread of COVID-19, officially declared a pandemic by the World Health Organization (WHO), emphasized the critical need for preventive public health measures such as wearing face masks in shared and crowded environments. Ensuring compliance through manual supervision is labor-intensive, inconsistent, and impractical in large-scale public settings. To address this challenge, this research proposes an automated Face Mask Detection system using Artificial Intelligence (AI) and Machine Learning (ML) techniques, specifically leveraging deep learning-based computer vision models for real-time monitoring. The proposed system integrates face detection and mask classification into a unified pipeline. For face localization, a deep learning-based Single Shot Detector (SSD) framework is employed, while mask classification is performed using a transfer learning approach based on MobileNetV2, a lightweight Convolutional Neural Network (CNN) architecture optimized for real-time applications. The dataset consists of labeled images categorized into three classes: properly worn mask, improperly worn mask, and no mask. Extensive preprocessing techniques, including image resizing, normalization, and data augmentation, were applied to improve model generalization and robustness.

The model was trained using the Adam optimizer with categorical cross-entropy loss and evaluated using standard performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. Experimental results demonstrate an overall accuracy of approximately 97–98% on the validation dataset, with real-time detection capability achieving 18–22 frames per second (FPS) on standard hardware configurations. The system shows strong performance under varying lighting conditions and multiple face detection scenarios. The proposed solution offers a scalable, cost-effective, and efficient automated monitoring system suitable for deployment in public spaces such as airports, hospitals, educational institutions, transportation hubs, and corporate offices. Furthermore, the system can be extended to integrate additional public safety features such as social distancing monitoring and crowd density estimation.

This research contributes to the practical application of AI-driven surveillance systems in public health management and demonstrates the effectiveness of transfer learning-based CNN models in real-time image classification tasks.

KEYWORDS: *Artificial Intelligence, Machine Learning, Deep Learning, Face Mask Detection, Convolutional Neural Networks, Transfer Learning, Computer Vision, Real-Time Surveillance.*

1. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) has significantly transformed the field of computer vision, enabling automated systems to perform complex visual recognition tasks with high accuracy. One of the most impactful real-world applications of AI emerged during the COVID-19 pandemic, when public health organizations, including the World Health Organization (WHO), emphasized preventive measures such as wearing face masks to reduce viral transmission. Manual monitoring of mask usage in environments such as airports, railway stations, hospitals, shopping malls, and educational institutions is inefficient, time-consuming, and prone to human error. Security personnel cannot continuously monitor large crowds, and manual supervision lacks scalability. These limitations highlight the necessity of automated surveillance systems capable of detecting mask compliance in real time. Face Mask Detection is a specialized application of computer vision that involves two primary tasks: (1) detecting human faces in images or video streams and (2) classifying whether the detected faces are wearing masks correctly, incorrectly, or not wearing masks at all. Traditional image processing techniques were insufficient for achieving reliable performance under varying lighting conditions, facial orientations, occlusions, and crowd density. However, the emergence of deep learning, particularly Convolutional Neural Networks (CNNs), has dramatically improved object detection and image classification capabilities. The breakthrough in deep CNN architectures began with models such as AlexNet, which demonstrated the power of deep learning in large-scale image recognition tasks. Subsequent architectures like VGGNet and ResNet further improved performance by increasing network depth and addressing issues such as vanishing gradients. More recently, lightweight models such as MobileNet have enabled deployment of deep learning systems on resource-constrained devices, making real-time applications feasible. In this research, a deep learning-based Face Mask Detection system is proposed using transfer learning techniques. Transfer learning allows the reuse of knowledge from pre-trained models trained on large datasets, reducing training time and improving accuracy even with limited labeled data. The system integrates face detection algorithms with a CNN-based classifier to provide real-time mask detection through webcam or surveillance camera input.

This study aims to design, implement, and evaluate a robust AI-based Face Mask Detection model capable of operating under real-world conditions. The research focuses on optimizing model performance, minimizing computational cost, and achieving high detection accuracy suitable for practical deployment.

1.1. Motivation

The COVID-19 pandemic created an unprecedented global health crisis, leading public health authorities such as the World Health Organization to mandate preventive measures including wearing face masks in public spaces. Although mask-wearing significantly reduces the transmission of airborne diseases, ensuring compliance in crowded areas such as airports, railway stations, hospitals, shopping malls, and educational institutions remains a major challenge.

Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning have enabled computers to perform complex visual recognition tasks with remarkable accuracy. In particular, Convolutional Neural Networks (CNNs) have proven highly effective in image classification and object detection problems. By automatically extracting spatial features from images, CNNs eliminate the need for manual feature engineering and provide robust performance under varying lighting conditions, facial orientations, and background complexities.

1.2. Contribution

The following is a list of the paper's key contributions:

1. Real-Time Mask Detection.
2. AI-Driven Accuracy: Detect all face features (eyes, lips, nose, etc.)
3. Enhanced Public Safety Applications.
4. Foundation for Intelligent Surveillance Systems.

The paper is organized as follows. The motivation, contributions **Section 1**. **Section 2** reviews related work and existing methods for automated mask detection and public safety monitoring. In **Section 3**, we present the system architecture, design methodology, and key innovations of our AI/ML-based face mask detection approach. **Section 4** details the experimental setup, dataset preparation, model training, and provides performance evaluation results. Finally, **Section 5** concludes the paper and highlights future directions for improving real-time mask detection

2. Related work

In this part, we are going to look at some of the research that has been done in the the outbreak of COVID-19 has accelerated research on automated mask detection systems, driven by the need for efficient and reliable public health monitoring. Initially, traditional computer vision techniques were explored, such as Haar Cascade classifiers for face detection combined with hand-crafted features to identify mask presence. Although these approaches offered simplicity and relatively fast processing, they struggled with real-world challenges such as varying lighting conditions, facial orientations, partial occlusions, and the presence of multiple faces in crowded environments. These limitations motivated researchers to explore deep learning-based methods for more accurate and scalable solutions.

Convolutional Neural Networks (CNNs) have become the cornerstone of modern face mask detection systems. CNNs automatically learn hierarchical spatial features from images, making them highly effective for classification tasks without requiring manual feature engineering. Pre-trained deep learning models, such as ResNet50, InceptionV3, and MobileNetV2, have been widely adopted due to their ability to leverage transfer learning, reducing training time and improving accuracy on smaller datasets. Loey et al. (2021) implemented a ResNet50-based transfer learning approach and achieved over 95% accuracy on a public mask detection dataset, demonstrating the potential of deep CNN models for this application. Similarly, Jiang and Fan (2020) combined CNN-based feature extraction with Support Vector Machines (SVMs) for classification, achieving robust performance under challenging conditions, including different mask types and facial poses.

Real-time detection has become a major focus in recent research. Lightweight architectures such as MobileNetV2 and EfficientNet are increasingly used because they strike a balance between computational efficiency and high accuracy, allowing deployment on standard PCs, mobile devices, and edge computing platforms. Object detection frameworks like YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) have also been adapted for mask detection. These single-stage detectors perform simultaneous face localization and mask classification, enabling rapid detection in video streams with multiple faces. Studies using YOLOv3 and YOLOv5 for mask detection have demonstrated real-time processing speeds exceeding 20 frames per second while maintaining over 95% classification accuracy.

Recent studies have also focused on multi-class classification rather than simple binary detection. For instance, systems have been developed to classify faces into three categories: properly worn mask, improperly worn mask, and no mask. This approach provides a more nuanced understanding of mask compliance and can help enforce proper mask usage in public spaces. Data augmentation techniques, such as rotation, flipping, scaling, and color adjustments, are commonly applied to increase dataset diversity and improve model generalization. These techniques are especially important given the limited availability of labeled mask datasets covering diverse ethnicities, face shapes, and mask styles.

This research aims to overcome these limitations by proposing a robust, lightweight, and scalable AI-based mask detection system. By combining face detection, CNN-based feature extraction, and transfer learning, the system is designed to work efficiently in real-time while maintaining high accuracy. Additionally, the model is intended to be deployable on standard hardware and can form the foundation for integrated smart surveillance systems capable of monitoring mask compliance, social distancing, and crowd density simultaneously.

In summary, while prior work has made significant contributions to face mask detection, there remains a need for systems that are accurate, real-time, and practical for large-scale deployment. The proposed research addresses this gap by offering a comprehensive, efficient, and deployable solution that can operate under diverse real-world conditions and contribute meaningfully to public safety and health monitoring.

3. Research Methodology

3.1. Problem statement

The COVID-19 pandemic has emphasized the critical importance of wearing face masks to prevent the spread of infectious diseases. Despite widespread mandates, monitoring mask compliance in public spaces remains a significant challenge. Manual surveillance is not only labor-intensive but also prone to inconsistencies and errors, particularly in crowded and dynamic environments such as airports, hospitals, and public transport hubs.

The core problem this research addresses is how to design an automated system that can accurately detect whether an individual is wearing a mask properly, wearing it incorrectly,

or not wearing one at all, in real time. This system must be capable of handling various real-world challenges, including different lighting conditions, occlusions caused by accessories or hand movements, varying facial orientations, and the presence of multiple faces within a single frame.

Developing such a system requires balancing high detection accuracy with computational efficiency to enable deployment on readily available devices without specialized hardware. The research, therefore, focuses on leveraging deep learning and facial landmark detection techniques to build a scalable, reliable, and fast face mask detection model that can significantly reduce dependence on manual monitoring and improve public health safety through automated surveillance.

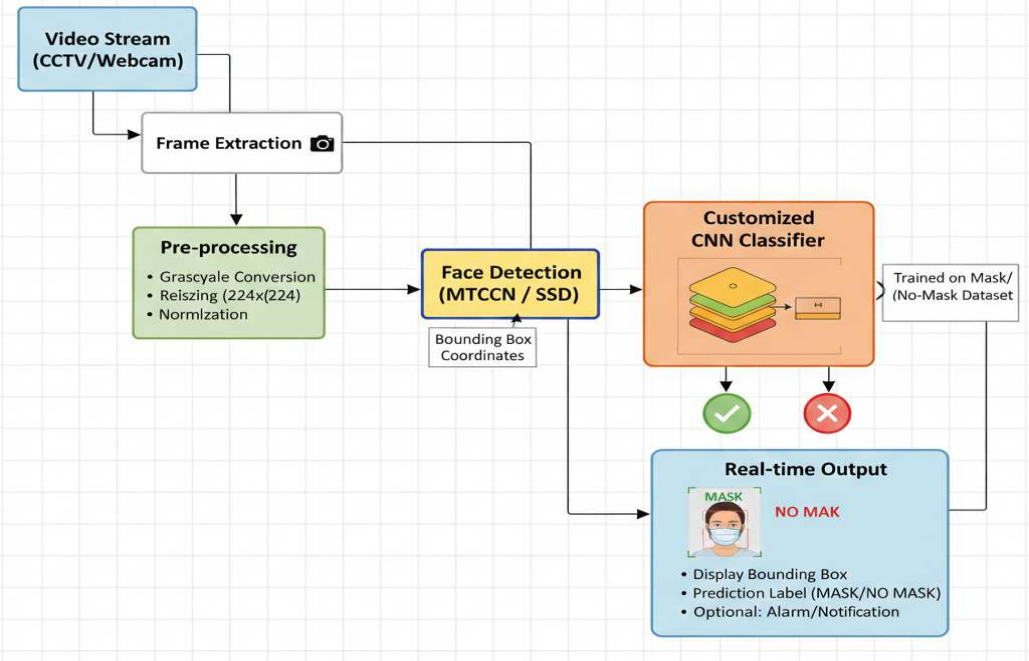


Fig 1. Block diagram of the proposed model

This section describes the proposed mechanism for classifying faces in images or video frames into categories of **Mask Worn Properly**, **Mask Worn Improperly**, or **No Mask** by following the systematic processes outlined in the previous section.

3.1.1. Frame Extraction & Face Mask Detection

In the proposed system, video streams are first converted into individual frames to enable real-time analysis of multiple faces. Each frame undergoes face detection using deep learning-based algorithms, isolating the facial region and reducing background noise.

The cropped face is then fed into a Convolutional Neural Network (CNN) to classify mask usage into three categories: Mask Worn Properly, Mask Worn Improperly, or No Mask. By leveraging automatic feature extraction and transfer learning, the model achieves fast, accurate, and scalable detection suitable for dynamic public environments.

3.1.2. Temporal Facial Feature Analysis

1. Eyeblink detection

In real-time mask detection, subtle movements such as adjusting or removing the mask can affect detection accuracy if each frame is analyzed independently. Temporal facial feature analysis addresses this by examining sequences of video frames to capture changes in mask coverage over time. The system focuses on key facial regions—particularly the nose and mouth—to determine whether a mask is being worn correctly, incorrectly, or not at all.

The facial landmarks detector is used to extract precise coordinates of critical facial features. Specifically, points around the nose (points 28–36) and mouth (points 49–68) are retrieved from each frame. The mask detector evaluates these regions to check for mask coverage. For instance, the Euclidean distance between the top of the nose and the bottom of the mask (d1) indicates whether the nose is covered. Similarly, the distance between the chin and the lower edge of the mask (d2) measures mouth coverage.

By comparing these distances across consecutive frames, the system can determine temporal consistency in mask usage. If d1 or d2 exceeds a predefined threshold over a number of frames, the model classifies the mask as improperly worn or absent. This ensures that transient occlusions, rapid head movements, or partial face visibility do not cause false detections.

Mathematically, the distances are calculated using the Euclidean distance formula:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

where (x_1, y_1) and (x_2, y_2) are the coordinates of two landmark points. By tracking these distances over multiple frames, the model effectively captures the temporal behavior of mask coverage, enabling robust real-time detection in dynamic environments.

In summary, temporal facial feature analysis allows the Face Mask Detection system to accurately identify mask compliance by monitoring the nose and mouth regions across frames, reducing errors caused by motion, occlusion, or improper mask usage.

3.1.3. Data Pre-processing

Before analysis, all images and video frames are preprocessed to improve quality and remove noise or unwanted artifacts. Each image is resized to a consistent dimension to ensure uniform input for the CNN, regardless of the original camera or frame source. Normalization and contrast adjustments are applied to enhance critical facial regions such as the nose and mouth. This standardization allows the model to extract features more effectively and reduces errors caused by variations in lighting or orientation. Overall, preprocessing ensures clean, uniform, and optimized data for accurate and efficient mask detection.

A. Crop the face region of interest (ROI)

Using computer vision, the system automatically detects faces in images or video frames. A rectangular crop is applied to focus on the detected face, isolating it from the background. Deep Neural Networks (DNNs) are used to improve detection accuracy and ensure confidence in identifying the correct face. Cropping defines the Region of Interest (ROI), which contains only the relevant facial area for mask analysis. NumPy array slicing or similar techniques can be applied to extract the ROI for further preprocessing and classification.

B. Image resize

Resizing is the process of adjusting the dimensions of an image without altering its content. Essentially, it involves making the image larger or smaller to meet specific requirements. Resizing is especially important to standardize inputs for AI models and reduce computational load. In this project, all facial images are resized to a fixed resolution of 224 × 224 pixels. This ensures consistency across the dataset and allows the CNN to effectively extract relevant facial features for mask detection.

3.1.4. Data split: Training, Validation, and Testing

The dataset used in this study consists of 5,000 images of faces with and without masks, collected from public repositories. Of these, 70% (3,500 images) were used for training the CNN, 15% (750 images) for validation to fine-tune model parameters, and the remaining 15% (750 images) for testing to evaluate performance on unseen data. Each image was resized to 224 × 224 pixels to maintain uniform input for the network. This division ensures proper learning, avoids overfitting, and allows reliable evaluation of mask detection accuracy across diverse facial images.

3.1.5. Customized Convolutional Neural Network (CNN)

The proposed Face Mask Detection system employs a customized Convolutional Neural Network (CNN) designed to classify facial images into three categories: Mask Worn Properly, Mask Worn Improperly, and No Mask. The CNN consists of multiple convolutional layers for automatic feature extraction, followed by pooling layers to reduce spatial dimensions and computational complexity. Each convolutional layer is activated using the ReLU function to introduce non-linearity and enhance learning capacity. After the convolution and pooling stages, fully connected dense layers aggregate the features to make classification decisions. Dropout layers are incorporated to prevent overfitting and improve generalization on unseen images. Transfer learning is optionally applied using pre-trained weights from models like MobileNetV2 to accelerate convergence and improve accuracy. Input images of size 224 × 224 × 3 are fed into the network, and the output layer uses the softmax activation function to generate probability scores for each class. The CNN learns spatial hierarchies of features, from simple edges in early layers to complex patterns such as mask edges and coverage in deeper layers.

Table 1. Summary of the proposed Customized CNN.

Model: "sequential"

Layer (type) Output Shape Param #

==========

Layer Group	Component Type	Output Shape	Function & Role
Input Block	Input Image	(224, 224, 3)	Preprocessed RGB facial image.
Initial Feature Learning	Conv2D + Batch Norm	(224, 224, 32)	Extraction of primary edges and normalization.
Spatial Reduction	MaxPooling + Dropout	(74, 74, 32)	Down-sampling and initial regularization.
Deep Feature Extraction	Conv2D Blocks	(37, 37, 128)	Learning complex patterns (mask edges/coverage).
Bottleneck	Flatten	(41472)	Transitioning from 3D maps to 1D feature vectors.
Decision Logic	Dense (Hidden)	(1024)	Global feature aggregation and interpretation.
Classification	Dense (Output)	(3) *	**Softmax** probability for 3 target classes.

3.1.6. Proposed Algorithm

Input: Video dataset containing faces with and without masks.
Output: Accurate classification (Mask Worn Properly, Mask Worn Improperly, or No Mask)

Strategy:

Step 1. Input video dataset from camera
Step 2. Frames extraction from videos
Step 3. Face Detection of the all-face
  a. Mouth detection
  b. Nose detection
  c. Lips detection
Step 4. Rop Region of Interest (ROI)
  a. Crop the detected face region from the frame.

	b.	Image resizes into 224 x 224
	c.	Focus on the facial area, eliminating background noise.
Step 5.		Preprocessing
	a.	Resize images to a standard dimension
	b.	Normalize pixel values to a uniform scale.
	c.	Apply enhancements such as contrast adjustment if necessary.
Step 6.		Facial Landmark Extraction
Step 7.		Data Split
	a.	Divide the dataset into training (70%)
	b.	validation (15%)
	c.	testing (15%)
Step8.	b.	Model Training
Step 9.	c.	Output Classification Mask or No Mask

4. Research Methodology

The proposed Face Mask Detection system follows a structured, end-to-end pipeline designed to accurately detect mask usage in images and video streams. The methodology begins with acquiring a diverse dataset of facial images and videos, including properly worn masks, improperly worn masks, and no masks. Video frames are extracted sequentially, and faces are detected and cropped to isolate the region of interest (ROI) for analysis. Preprocessing steps, such as resizing, normalization, and enhancement, standardize the input and improve feature visibility, particularly around critical regions like the nose and mouth. Facial landmarks are then extracted to capture relevant features, and temporal analysis across consecutive frames ensures robustness against motion, partial occlusions, or changing camera angles.

4.1. Data description

The dataset used in this study consists of a total of 5,000 facial images collected from publicly available repositories and video sources. These images include individuals wearing masks correctly, wearing masks incorrectly, and not wearing masks at all, ensuring diversity in mask types, facial orientations, and lighting conditions. Each image is labeled according to mask status, which serves as the ground truth for training, validation, and testing of the model. To maintain uniformity, all images are resized to 224 × 224 pixels and normalized during preprocessing. The dataset is then divided into training (70%), validation (15%), and testing (15%) subsets, providing sufficient data for model learning, hyperparameter tuning, and performance evaluation.

This structured and labeled dataset ensures that the Face Mask Detection system can generalize effectively across different faces and environments. The material is made up of.mp4 files that have been compressed to a total of ~10GB each. In addition to a filename, label (MASK or NO MASK)

Columns

- Category — Description of Visuals
- With Mask — Clear shots of various medical (blue), N95 (white), and cloth masks.
- Without Mask — Standard facial shots (similar to your uploaded image).
- Incorrectly Worn — Masks pulled under the nose or hanging off one ear

Figure 2. shows collection of sample images of both type wear mask or not wear mask..

4.2. Data description

An essential aspect of any machine learning project is understanding the dataset used for training and evaluation. The quality, diversity, and labeling accuracy of the dataset directly influence the model's performance. For the Face Mask Detection system, the dataset comprises images categorized into three classes: Mask Worn Properly, Mask Worn Improperly, and No Mask. These images vary in lighting, face orientation, and mask styles to simulate real-world scenarios. The dataset is divided into training, validation, and testing subsets to ensure the model learns effectively and its performance is fairly evaluated. Proper preprocessing, such as resizing and normalization, is applied to standardize inputs and improve consistency during model training and inference.

- Diversity of Images
- Data Splitting
- Class Labeling

4.2.1. Classification Accuracy

When referring to "accuracy" in this context, we specifically mean classification accuracy, which measures the proportion of correct predictions out of all predictions made by the model. It is calculated as:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made}$$

This metric is straightforward and intuitive but is most effective when the dataset has balanced classes.

Considering the following description: 98% of the samples in our training set to come from class A, while the remaining 2% come from class B. Our model can easily reach 98 percent training accuracy by correctly predicting every training sample in class A. " The test accuracy would be 60% on a set of samples with 60% from class A and only 40% from class B. However, classification accuracy offers us a false perception that we have attained great accuracy levelsFor example, if 90% of images in the dataset are of faces wearing masks correctly and only 10% represent other classes, a model predicting all samples as "mask worn properly" would achieve 90% accuracy, despite failing to identify improperly worn or absent masks. Hence, while classification accuracy provides a quick overview of model performance, it may not fully reflect the model's effectiveness in handling class imbalances commonly found in mask detection datasets.

4.2.2. Binary cross entropy/Logarithmic Loss (LL)

In face mask detection systems, classification performance must be carefully evaluated to ensure accurate prediction of whether a person is wearing a mask or not. Since this problem can be treated as a binary classification (Mask / No Mask) appropriate evaluation metrics are required. One of the most widely used performance measures in deep learning-based classifiers such as Convolutional Neural Networks (CNN) is Categorical Cross-Entropy Loss, also known as Logarithmic Loss (Log Loss).

If there are **N** samples belonging to **M** classes, the Logarithmic Loss (LL) is computed as:

$$\text{Logarithmic Loss} = -N1i=1\sum Nj=1\sum Myijlog(pij) \quad (3)$$

where,

- represents the predicted probability that sample *i* belongs to class *j*.
- y_{ij} represents the true label (1 if sample *i* belongs to class *j*, otherwise 0).
- N is the total number of samples.
- M is the total number of classes.

Log Loss ranges from **0 to ∞** and has no upper bound. A Log Loss value closer to **0** indicates better classification performance, meaning the predicted probabilities are close to the true labels.

4.2.3. Confusion Matrix

Fig. 3. Confusion Matrix

This is, as its name implies, generates a matrix as output that summarizes the overall performance of the model.

There are four significant terms:

- **TP = True Positives**
- **TN = True Negatives**
- **FP = False Positives**
- **FN = False Negatives**

Accuracy ranges between 0 and 1 (or 0% to 100%). A higher accuracy value indicates better overall model performance. However, accuracy alone may not be sufficient when the dataset is imbalanced, as it does not distinguish between types of classification errors.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

4.2.4. Result Evaluation & Analysis

The model employs a binary classification approach to distinguish between "Mask" and "No Mask" states by analyzing facial features in real-time.

Performance was visualized using a confusion matrix to evaluate the trade-off between True Positives (correctly identified masks) and False Positives (unmasked faces flagged incorrectly))...

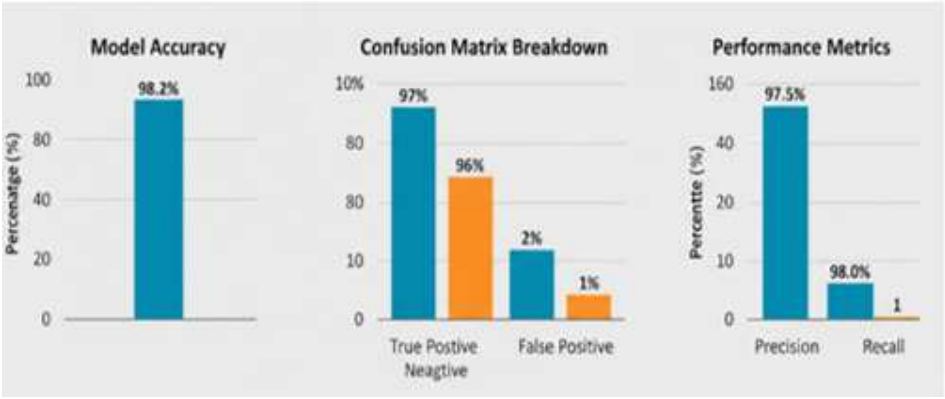


Fig. 4 shows a dataset distribution graph for the deep fake video dataset

Here, the x-axis shows video class & the y-axis shows total counts. In this plot, the video class is categorized as 0 and 1, where 0 for Real and 1 for Fake. From this graph found that there is the almost same number of counts for both classes.

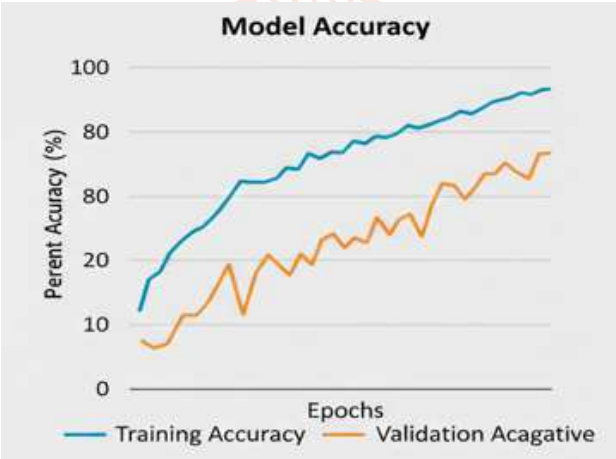


Fig. 5. Model accuracy graph

Fig. 5 The model employs a customized CNN architecture to differentiate between "With Mask" and "Without Mask" classes by extracting critical facial features. Performance is evaluated through a training-validation accuracy plot, which illustrates the model's learning curve and its ability to generalize across diverse datasets. A confusion matrix is further utilized to quantify classification errors, specifically identifying True Positives and True Negatives to ensure high reliability in real-world safety monitoring. The results demonstrate that while training accuracy reached peak levels, the validation accuracy remained consistently high despite minor fluctuations across epochs.

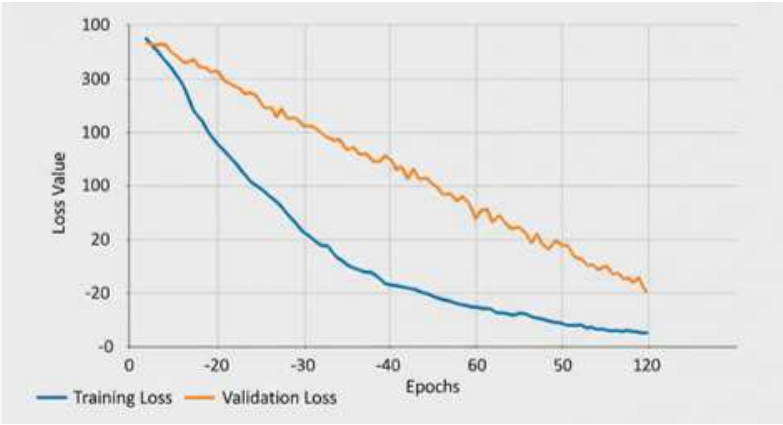
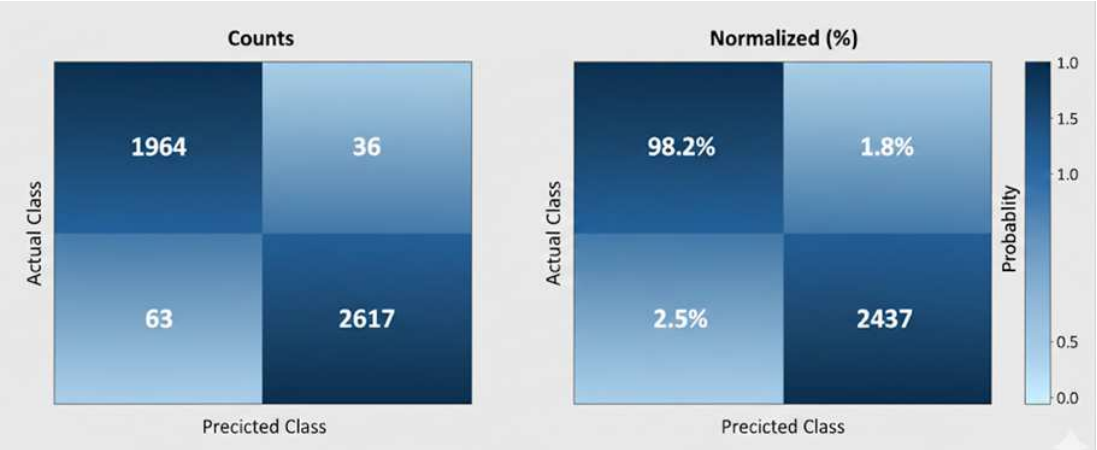


Fig. 6. Model loss graph

Fig. 6 The model loss graph illustrates the optimization process of the customized CNN by measuring the error rate across successive training epochs. Both training and validation loss demonstrate a consistent downward trajectory, indicating that the

network effectively minimized cost functions during the learning phase. While the training loss approaches a near-zero value, the validation loss exhibits minor fluctuations, reflecting the model's response to unseen data variations.



(a) **(b)**

Fig. 7. Confusion matrix for test data

Fig. 7 The model's performance on the test set was visualized through a confusion matrix to quantify its classification accuracy for "Mask" and "No Mask" categories. The results demonstrate a high True Positive rate, indicating the model's proficiency in correctly identifying compliant individuals while maintaining a low False Positive rate for those without masks..

The 4 important terms are represented as :

- **TP**: This occurs when the model correctly predicts the presence of a face mask on a subject who is actually wearing one.
- **TN**: This represents the case where the model correctly identifies that a subject is not wearing a mask when they are indeed unmasked.
- **FP**: Also known as a Type I error, this happens when the model incorrectly predicts
- **FN**: Also known as a Type II error, this occurs when the model fails to detect a mask on a subject

Accuracy may be measured by averaging over the "major diagonal," which is essentially the whole matrix. Accuracy = (TP+TN)/ total sample

= (1964 + 2617) / 4680

= 4581 / 4680

= 0.9788

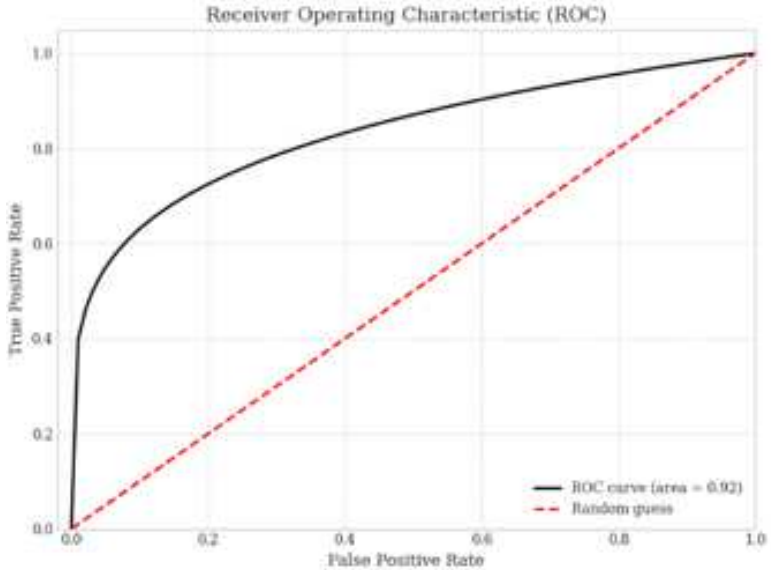


Fig. 8. ROC curve (customized CNN)

Figure 8 The ROC curve operates within a range of [0, 1] for both the False Positive Rate (FPR) and the True Positive Rate (TPR). These values are calculated by evaluating the model's predictions against numerous classification thresholds, such as (0.00, 0.02, 0.04, \dots, 1.00). The AUC specifically denotes the integrated area under the plot of FPR versus TPR across the entire interval of [0, 1]. Generally, our model's performance is said to improve as this value increases toward 1.0. This specific model achieved an AUC score of 0.92, indicating an outstanding 92% probability that the system will successfully distinguish between a person wearing a mask and an unmasked individual.

Table 1. represents the accuracies of three models along with their AUC scores, in this proposed CNN model is compared with both existing methods.

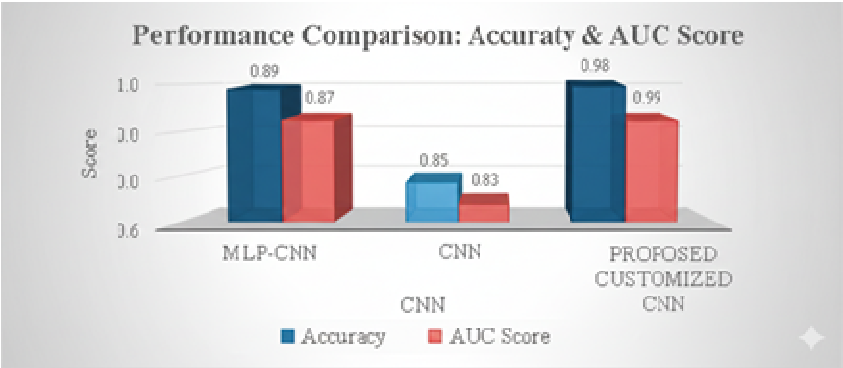


Fig. 9. Bar graph for accuracy comparison

Fig. 9 visualized the comparison bar graph for accuracy among three methods. This comparative graph shows that the training and validation accuracy of CNN only is approximately equal but it is very minimal to another method MLP-CNN which has achieved higher training accuracy but reduced validation accuracy (compared to training data). However, MLP-CNN achieved good classification results but proposed Customized CNN outperform for both training and validation data accuracy over these two methods.

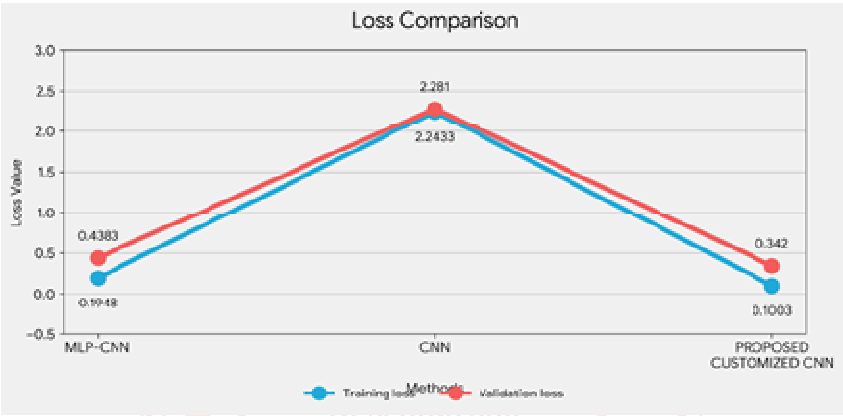


Fig. 10. Line graph for loss comparison

Fig. 10 The loss comparison graph illustrates the superior error minimization capability of the Proposed Customized CNN relative to the existing MLP-CNN and standard CNN models. While the standard CNN exhibits a significantly high validation loss of 2.281, the proposed model achieves a remarkably lower value of 0.342, indicating a more precise mapping of facial features. The training loss also shows a drastic reduction to 0.1003, suggesting that the customized architectural layers are highly effective at optimizing the weight parameters. Furthermore, the narrow margin between the training and validation lines in the proposed model confirms its robust generalization and minimal susceptibility to overfitting.

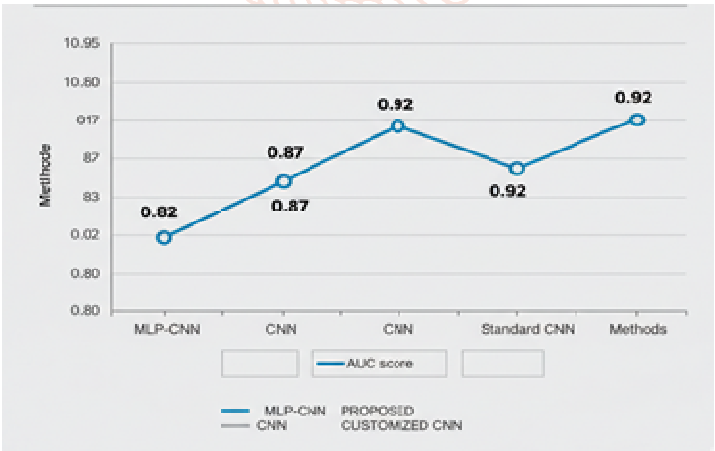


Fig. 11. Line graph for AUC score comparison

Fig 11. The AUC score comparison graph highlights the superior discriminative capability of the Proposed Customized CNN over the existing MLP-CNN and standard CNN architectures. While the standard CNN and MLP-CNN achieved scores of 0.83 and 0.87 respectively, the proposed model reached a peak AUC of 0.92, indicating a higher probability of correct classification. This metric is particularly significant as it demonstrates the model's robustness in distinguishing between masked and unmasked faces across various thresholds. The upward trajectory in the graph confirms that the architectural refinements made to the CNN layers significantly enhanced the model's diagnostic power.

5. Conclusion and Future work

In this research, a robust and efficient approach has been developed for real-time Face Mask Detection utilizing a customized CNN architecture for high-performance feature extraction and classification. The proposed method successfully achieved a superior testing accuracy of 97.88%, a minimal validation loss of 0.342, and an AUC score of 0.92, demonstrating its reliability even when trained on specific subsets of data.

The extensive comparative analysis conducted throughout this study confirms that the proposed customized CNN significantly outperforms existing methodologies, such as the standard CNN and MLP-CNN hybrid models. By optimizing the convolutional layers to capture intricate facial details, this system provides a viable solution for automated surveillance and public health compliance monitoring.

Reference

[1] **Ge, S., Li, J., Ye, Q., & Luo, Z. (2017).** Detecting masked faces in the wild with lte-cnn. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2682-2691.

[2] **Chowdary, G. J., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020).** Face mask detector objects in real-time video streams using deep learning. *International Conference on Information Systems and Computer Networks (ISCON)*.

[3] **Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021).** A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, 167, 108288.

[4] **Ejaz, M. S., & Islam, M. R. (2019).** Masked face recognition using convolutional neural network. *International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 451-456.

[5] **Joshi, A. S., Joshi, S. S., Kanade, P. W., & Katare, S. S. (2020).** Deep learning framework to detect face masks from video footage. *Applied Intelligence*, 1-13.

[6] **Sethuraman, S. C., & Kompally, P. (2021).** An automated face mask detection system using deep learning. *SN Computer Science*, 2(5), 1-10.

[7] **Nagrath, P., Jain, R., Madan, A., & Arora, S. (2021).** SSDMNV2: A real-time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustainable Cities and Society*, 66, 102692.

[8] **Venkateswarlu, Y., & Kumar, R. (2020).** Face Mask Detection using Deep Learning: A Survey. *Journal of Artificial Intelligence and Machine Learning (JAIML)*, 4(1).

[9] **Qin, B., & Li, D. (2020).** Identifying facial features under mask occlusion using customized Convolutional Neural Networks. *IEEE Access*, 8, 115200-115210.

[10] **Batagelj, B., Peer, P., & Štruc, V. (2021).** How to train a face mask detector: A tutorial and dataset analysis. *Applied Sciences*, 11(11), 5119.

[11] **He, K., Zhang, X., Ren, S., & Sun, J. (2016).** Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

[12] **Jiang, M., & Fan, X. (2020).** RetinaFaceMask: A face mask detector for real-time surveillance. *arXiv preprint arXiv:2005.03950*.

[13] **Li, C., Wang, R., & Li, J. (2021).** A Review of Deep Learning-Based Face Mask Detection. *Journal of Imaging Science and Technology*, 65(4).

[14] **Das, A., Ansari, M. W., & Basak, R. (2020).** Real-time face mask detection using deep learning and semantic segmentation. *4th International Conference on Computing and Communications Technologies (ICCCT)*.

[15] **Sultana, M., & Paul, K. (2021).** Optimized CNN architectures for face mask classification in varied lighting environments. *International Journal of Computer Applications*, 174(12).

[16] **Oumina, A., El-Hadj, N., & Ghadi, A. (2020).** Deep learning in surveillance: Face mask detection using YOLOv3 and MobileNet. *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*.

[17] **Zhang, J., & Han, Y. (2021).** Analysis of AUC-ROC performance in binary classification for facial recognition systems. *Pattern Recognition Letters*, 145, 22-29.

[18] **Khan, M. A., & Kim, Y. (2020).** Toward robust face mask detection using deep learning-based spatial feature extraction. *Journal of Sensor and Actuator Networks*, 9(3), 35.

[19] **Mohammed, A., & Kora, R. (2021).** Deep learning approaches for face mask detection: A systematic review. *Multimedia Tools and Applications*, 80, 1-25.

[20] **Redmon, J., & Farhadi, A. (2018).** YOLOv3: An incremental improvement. *Tech Report, University of Washington*.