Proposed Activation Function Based Deep Learning Approach for Real-Time Face Mask Detection System

NAY KYI TUN[1], AYE MIN MYAT[2]
[1]Faculty of Computer System and Technology, Myanmar Information Institute of Technology
Mandalay, MYANMAR

[2]University of Technology
(Yadanapon Cyber City) Pyin Oo Lwin, MYANMAR

Abstract: — The ongoing global pandemic has underscored the importance of effective preventive measures such as wearing face masks in public spaces. In this paper, we propose a deep learning-based approach for real-time face mask detection to aid in enforcing mask-wearing protocols. Our system utilizes convolutional neural networks (CNNs) to automatically detect whether individuals in images or video streams are wearing masks or not. The proposed system consists of three main stages: face detection, face mask classification, and real-time monitoring. Firstly, faces are localized in the input image or video frame using a proposed face detection model. Then, the detected faces are fed into a proposed CNN model for mask classification, which determines whether each face is covered with a mask or not. Finally, the system will provide real-time monitoring and alerts authorities or stakeholders about non-compliance with mask-wearing guidelines. We evaluate the performance of our system on publicly available datasets and demonstrate its effectiveness in accurately detecting face masks in various scenarios. Additionally, we discuss the challenges and limitations of deploying such a system in real-world settings, including issues related to privacy, bias, and scalability. Overall, our proposed face mask detection system offers a promising solution for automated monitoring and enforcement of face mask policies, contributing to public health efforts in mitigating the spread of contagious diseases.

Key-words: — CNN, Face Mask, Detection, Classification, YOLO.
Received: March 27, 2024. Revised: September 5, 2024. Accepted: September 24, 2024. Published: October 17, 2024.

1. Introduction

The emergence of the COVID-19 pandemic has necessitated the adoption of stringent public health measures, with widespread face mask usage being one of the most effective means of preventing the spread of infectious diseases in communal settings. However, ensuring compliance with mask-wearing protocols in public spaces presents a formidable challenge for authorities and organizations worldwide. In response to this challenge, the development of automated face mask detection systems has garnered significant attention as a promising solution to enforce mask-wearing guidelines efficiently.

This article provides an in-depth exploration of recent advancements in face mask detection technology, examining the underlying methodologies, applications, challenges, and future prospects of such systems. As the pandemic continues to evolve, the need for reliable and scalable solutions for monitoring mask compliance remains paramount. By leveraging cutting-edge technologies such as computer vision, deep learning, and edge computing, researchers and practitioners have made substantial progress in the development of robust and real-time face mask detection systems.

Convolutional neural networks (CNNs), in particular, have demonstrated remarkable capabilities in accurately detecting faces and distinguishing between masked and unmasked individuals. Moreover, the integration of complementary technologies such as thermal imaging, edge computing, and privacy-preserving techniques has further enhanced the utility and efficacy of face mask detection systems in diverse operational settings.

Beyond the immediate imperative of pandemic management, face mask detection systems hold immense potential for addressing broader societal challenges, including security surveillance, access control, and public safety monitoring. By leveraging the insights gained from research and real-world deployments, policymakers, businesses, and public health authorities can develop evidence-based strategies to promote mask compliance and mitigate the risk of disease transmission.

However, the widespread adoption of face mask detection systems also raises important ethical, legal, and societal considerations. Issues related to privacy, bias, algorithmic fairness, and consent must be carefully addressed to ensure that these systems are deployed responsibly and ethically. Furthermore, the interoperability and standardization of face mask detection technologies are essential to facilitate seamless integration with existing infrastructure and interoperability across different platforms.

In light of these considerations, this article aims to provide a comprehensive overview of the YOLO [5] based face mask detection systems, offering insights into the technological advancements, practical applications, and ethical implications of this rapidly evolving field. By synthesizing existing research and identifying key challenges and opportunities, we seek to inform future research directions and contribute to the development of effective and equitable solutions for promoting mask compliance and safeguarding public health.

2. Aim

The aim of this article is to provide a comprehensive analysis of face mask detection systems, encompassing the technological advancements, practical applications, challenges, and ethical considerations in this emerging field.

3. Related Work

A. Velip and A. Dessai [1] proposed a multi-task learning framework combining face detection and mask classification, achieving high accuracy in diverse environmental conditions. This study proposes a real-time face mask detection system using deep learning techniques. The authors employ a CNN architecture for face detection and mask classification.

S. Sakshi et al. [2] present a face mask detection system based on a CNN model trained on a large dataset of masked and unmasked faces. Their system exhibits robust performance across various environmental conditions and lighting scenarios. Additionally, they explore the impact of different CNN architectures on detection accuracy.

N. Kowalczyk et al. [3] propose a comprehensive face mask detection and recognition system integrated with thermal imaging technology. Their system combines deep learning-based mask detection with facial recognition to identify individuals and enforce mask-wearing protocols in public spaces. The study emphasizes the importance of multimodal approaches for enhanced accuracy and reliability.

D. Singh and S. K. Joshi [4] introduce a lightweight CNN model tailored for real-time face mask detection on edge devices with limited computational resources. Their system achieves competitive performance while maintaining low computational overhead, making it suitable for deployment in resource-constrained environments such as embedded systems and IoT devices.

These studies represent a diverse range of approaches and methodologies for face mask detection, spanning from real-time systems using deep learning to privacy-preserving solutions leveraging federated learning. By building upon and extending the findings of these works, our proposed face mask detection system aims to contribute to the advancement of this critical area of research.

4. Methodology

In this paper, we proposed our new YOLO-based detection algorithm for face mask detection system. The block diagram of our proposed system consists of the training section and the detection section that were shown in fig.1.

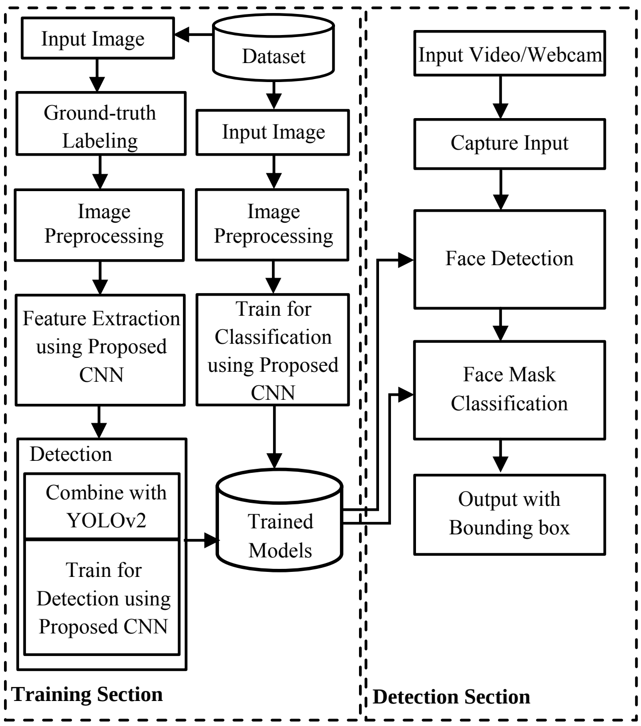


Fig. 1. Block diagram of proposed face mask detection system.

In the training section, firstly we make a ground truth dataset with the original image dataset for detection. And we apply preprocessing to the image. We extract features of the image by using our proposed Convolutional Neural Network without consisting of the fully connected layer. The architecture is concatenated with the YOLOv2 detection network model for the part of detection. And then, we train the image dataset. As for the classification model training, the input dataset images are applied preprocessing such as resizing and scaling. The input image size is changed to 224x224 pixels. The smaller the size of image, the fewer the features of image. After resizing the face mask images, all these images are trained by using the proposed Convolutional Neural Network for face mask classification.

In the detection section, the input is obtained from the video or webcam. The proposed system is captured these input frames and detected face by using our proposed detection model. After detecting the face, the system applied face mask classification which is based on the result of face detection.

In the process flow diagram of the proposed system, the system is started by capturing the input frame from the webcam. Face are detected by using the proposed object detection method which is based on the YOLOv2 algorithm[6]. If the input frame detects face, the proposed CNN architecture is applied for the classification of face mask. And then, the value of classification results is shown as an output. If the input frame does not contain a face or does not detects face, the process of detection will return to the input stage, which is shown in fig. 2.

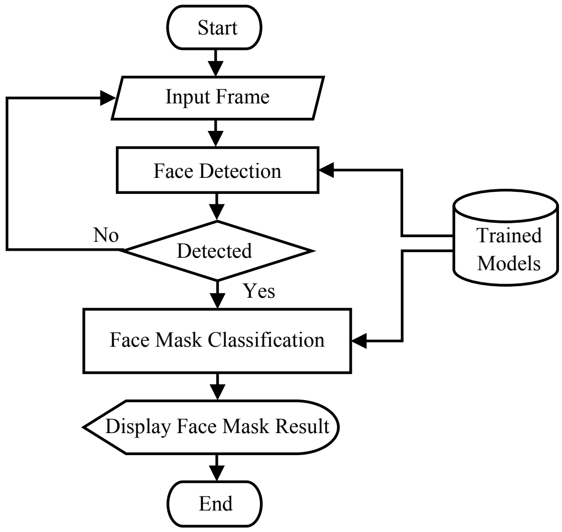


Fig. 2. Flow chart of proposed face mask detection system.

4.1 Preprocessing

In the stage of the preprocessing for detection, data augmentation is applied to the ground truth label image. We used data augmentation to improve the network accuracy and we also made a random transformation to original data to increase the labeled training data. We apply random horizontal flipping and random X/Y scaling. We also use color space transformation to convert RGB to HSV color space. And then, we jitter image color to make randomly augment the color of each pixel.

As for the stage of training for face mask classification, the data augmentation techniques are also used that includes scale augmentation and position augmentation. Scale augmentation technique and position augmentation is used to augment the scale and position of training images. We crop and resize the detected face image into 224x224x3 pixel image. After the stages of preprocessing, we trained the detected model and classification model by using our proposed CNN architectures. The preprocessing result of detection and classification are shown in Fig.3 and Fig.4.

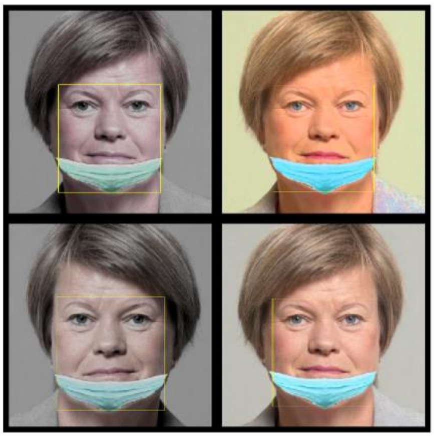


Fig. 3. Preprocessing of input dataset image for detection

Fig. 4. Preprocessing of input dataset image for classification.

4.2 Proposed Detection Network Based on Yolov2 Algorithm

In our proposed system, we make detection of face by using the proposed detection model which is transferred from the YOLOv2 object detection model. Before the development of the YOLOv2 algorithm, we augmented the dataset which is not only to enlarge our utilized dataset artificially but also to lessen the likelihood of overfitting throughout the training stage [7].

In the architecture of CNN, the input layer receives the image with a size of 416x416 with RGB. The architecture and parameters of the proposed face detection model is shown in fig. 5 and Table I. The architecture consists of the feature extraction part and detection part. In the feature extraction, we used our tiny CNN architecture which is shown in Table I. In a typical CNN, full-connected layers are usually placed toward the end of the architecture [8]. In our proposed architecture, we operate a series of convolution processes without consisting of a fully connected layer because we want to replace the YOLO detection sub-network. We apply the YOLO convolution layers, YOLO transform layer, and YOLO output layer. Finally, the result of face detection displays as an output. There are 8 convolution layers and each convolution layer consists of convolution, batch normalization and proposed activation function layer.

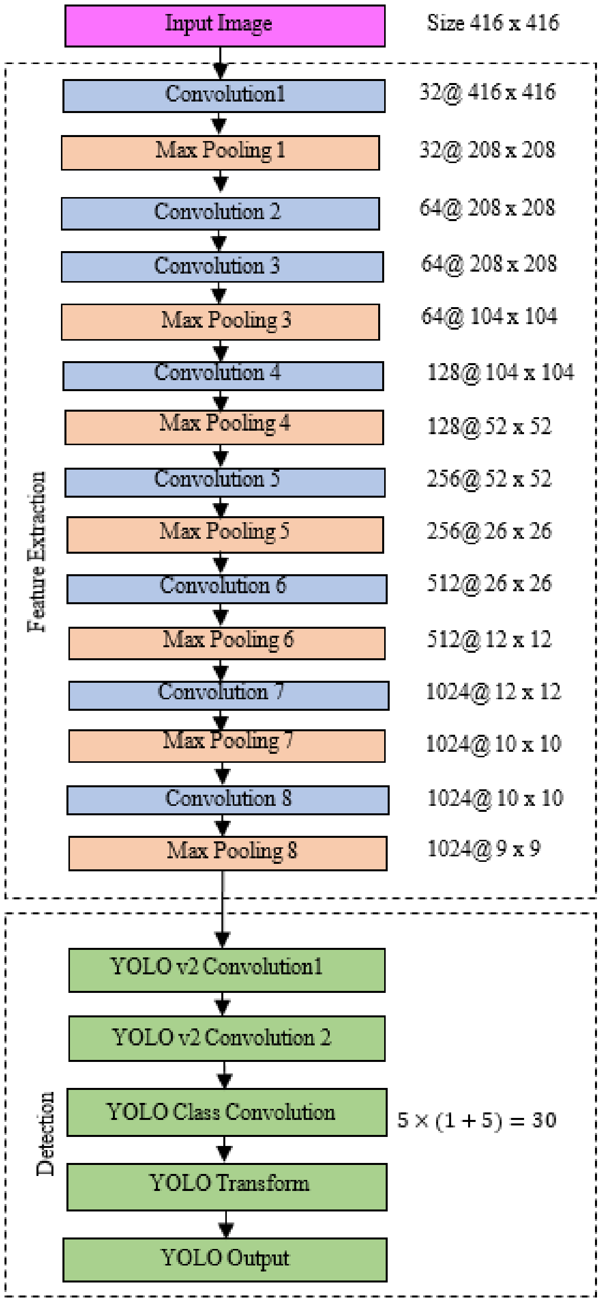


Fig. 5. Proposed detection network architecture.

TABLE I. THE LAYER VALUES OF PROPOSED CNN ARCHITECTURE FOR FACE DETECTION

Type	Filters	Size/Stride	Output
Conv1	32	3 × 3	416 × 416
MaxPool1		2 × 2/2	208 × 208
Conv2	64	3 × 3	208 × 208
Conv3	64	3 × 3	208 × 208
MaxPool3		2 × 2/2	104 × 104
Conv4	128	3 × 3	104 × 104
MaxPool4		2 × 2/2	52 × 52
Conv5	256	3 × 3	52 × 52
MaxPool5		2 × 2/2	26 × 26
Conv6	512	3 × 3	26 × 26
MaxPool6		3 × 3/2	12 × 12
Conv7	1024	3 × 3	12 × 12
MaxPool7		3 × 3/2	10 × 10
Conv8	1024	3 × 3	10 × 10
MaxPool8		2 × 2/2	9 × 9

4.3 Proposed Classification Network Architecture

After the detection of the face, the process of classification is applied to the detected face by using the proposed CNN model. There are several architectures in the field of Convolution Networks that have a name. The most common are AlexNet, VGGNet, and GoogleNet. The architecture of the proposed classification method using Convolutional Neural Network is shown in fig. 6.

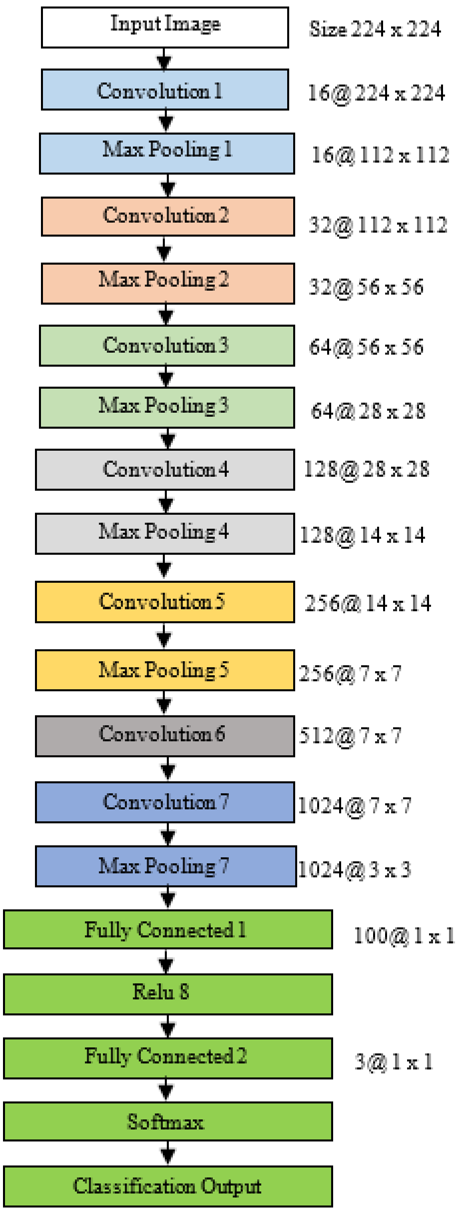


Fig. 6. The architecture of the proposed CNN for facemask classification.

In the architecture of CNN for face mask classification, we resize the input image into 224x224 with RGB. There are seven convolution layers and six max-pooling layers. After passing the two Fully-Connected layers and Softmax layer, the final size will be reduced to 1x1x2. The classification output layer produces the output corresponding to each group. The output is related to the 3 classes for the classification of without face mask, with face mask and wrong position of face mask. The parameters of the proposed CNN for face mask classification are shown in Table II. The number of filters and the size of output feature maps are shown in it.

TABLE II. THE PARAMETERS OF PROPOSED CNN RCHITECTURE FOR FACE MASK CLASSIFICATION

Layer	Filter/Weight	Number of Parameters
Convolutional Layer (C1)	((3×3×3)+1)×16	448
Convolutional Layer (C2)	((3×3×16)+1)×32	4,640
Convolutional Layer (C3)	((3×3×32)+1)×64	18,496
Convolutional Layer (C4)	((3×3×64)+1)×128	73,856
Convolutional Layer (C5)	((3×3×128)+1)×256	295,168
Convolutional Layer (C6)	((3×3×256)+1)×512	1,180,160
Convolutional Layer (C7)	((3×3×512)+1)×1024	4,719,616
Fully Connected (F1)	((1024)+1)×100	102,500
Fully Connected (F2)	((100)+1)×3	303
Total parameter		6,395,187

4.4 Proposed Activation Function

For the feature extraction of CNN, the proposed activation function was used to follow the convolution layer. The proposed activation function which is shown in fig.7. The proposed activation function extracts the feature without reducing the negative value of the convoluted feature map.

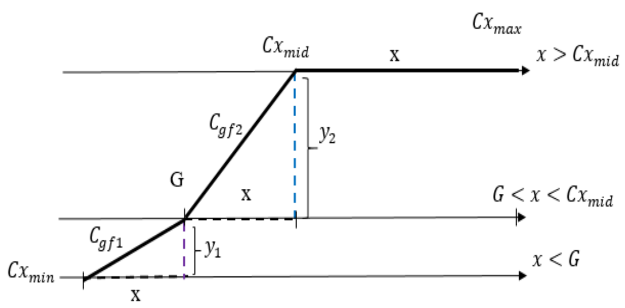


Fig. 7. Proposed Activation Function.

The mathematical form of the proposed activation function is shown in equation (3.1).

$$\text{PAct} = \begin{cases} x & , \quad x > Cx_{mid} \\ x \times C_{gf2} & , \quad G < x < Cx_{mid} \\ x \times C_{gf1} & , \quad x < G \end{cases} \qquad (3.1\text{ a})$$

$$Cx_{mid} = \frac{Cx_{max} + Cx_{min}}{2} \qquad (3.1\text{ b})$$

$$G = \frac{Cx_{max} - Cx_{min}}{4} + Cx_{min} \qquad (3.1\text{ c})$$

$$C_{gf2} = \frac{x - G}{Cx_{mid} - G} \times y_2 \qquad (3.1\text{ d})$$

$$C_{gf1} = \frac{x - Cx_{min}}{G - Cx_{min}} \times y_1 \qquad (3.1\text{ d})$$

Where x is a convolution layer value, Cx_{max} is maximum value of convolution layer, Cx_{min} is minimum value of convolution layer, Cx_{mid} is midpoint value of convolution layer, G is gradient start point of feature value, C_{gf1}, C_{gf2} are gradient of feature values and y_1 and y_2 are the constant leak factors . In the proposed system, the value of y_1 is 1 and the value of y_2 is 0.01.

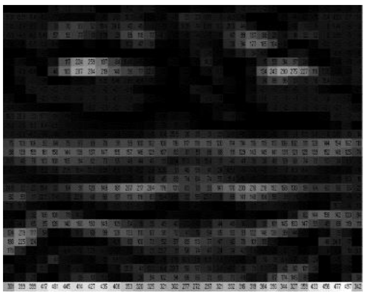
5. Result and Discussion

The training of detection network used an Stochastic Gradient Descent with Momentum (SGDM) optimizer optimizer, with an initial learning rate of 0.001, mini-batch size of 5, and 32 maximum number of epochs. The training of detection model is done to 7200 images of 1 class. The detector model is then saved and used for the detection of 2400 testing images. This is done by randomly using 60% of all images for training, 20% for validation and 20 % for testing. The average precision of detection is 0.97%. Our proposed system is work well in different positions and light conditions. It also works in blur conditions.

As for the face mask classification, the training used the SGDM optimizer, with an initial learning rate of 0.001, mini-batch size of 30 and 7 maximum number of epochs. Then training is done to 15000 images of 3 classes. The classifier model is then saved and used for the classification of 3,000 validation images, and 30,00 testing images. This is done by randomly using 60% of all images for training, 20% for validation and 20% for testing.

(a)

(b)



(c)

Fig. 8. Visualization results of feature extraction (a) Original image, (b) Feature extraction with ReLU activation and (c) Feature extraction with proposed activation

The process of proposed feature extraction is compared with ReLU activation function which is shown in fig.8. The proposed activation function based feature extraction method extract the feature of image more accuratly than ReLU activation function based method.

The avearge precision of proposed detection model is compared with YOLOv1, YOLOv2 that is shown in fig.9. The process is based on training with SGDM optimizer. The average precision of YOLOv1 based model is 0.919, YOLOv2 based model is 0.941 and proposed model is 0.979. According to the result, the SGDM optimizer based proposed detection model has a higher accuracy than YOLOv1 and YOLOv2.

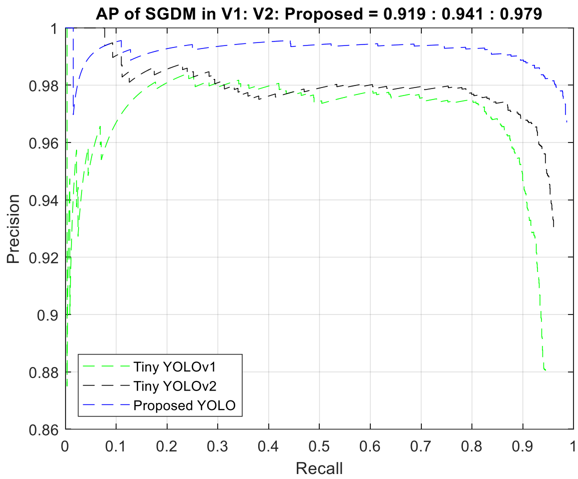
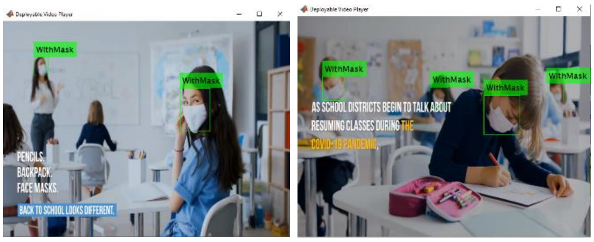


Fig. 9. Comparative analysis of the detection model.

Proposed detection model is tested face mask detection in daylight, roomlight, different background color, horizontal poistion and vertical position. We expressed our experimental result of detection and classification that includes working in the real-time conditions in fig.10.

(a)

(b)



(c)

(d)

Fig. 10. Result of face mask detection (a) Face mask detection on image, (b) Face detection on image (c) Face mask detection on video and (d) Face mask detection on real-time.

The confusion matrix of the face mask classification is shown in fig.11 which is based on the 3 classes of 3000 test dataset images.Table III shows the accuracy result of the proposed classification model that is obtained from the confusion matrix results. According to the average result, the precision is more than recall. The average result of the negative samples of all sentences is over 99.9% which shows the specificity of the classification model. The accuracy of each class is over 96% and the average accuracy of all classes is 96%.

	FM	Without FM	Wrong FM
FM	942	12	46
Without FM	12	947	41
Wrong FM	50	15	935

Fig. 11. Confusion Matrix of Proposed System.

TABLE III. ACCURACY RESULTS OF FACE MASK CLASSIFICATION

Class	Precision	Recall	Specificity	F1 Score	Accuracy
1	0.938	0.942	0.968	0.940	0.9592
2	0.972	0.947	0.985	0.959	0.9724
3	0.914	0.935	0.955	0.924	0.9489

6. Conclusion

In conclusion, face mask detection systems represent a crucial technological advancement in the ongoing efforts to combat infectious diseases and promote public health. Throughout this article, we have explored the evolution of face mask detection technology, from traditional image processing methods to sophisticated deep learning algorithms. Our analysis has highlighted the significant strides made in improving the accuracy, reliability, and scalability of face mask detection systems. Real-world deployments and case studies have demonstrated the practical utility of these systems in enforcing mask-wearing protocols and mitigating the spread of contagious diseases in various settings. Looking ahead, future research and development efforts should focus on addressing the remaining challenges and limitations of face mask detection systems, including dataset bias, robustness to environmental conditions, and interoperability across different platforms.

References

[1] A. Velip and A. Dessai, "Face Mask Detection Using Machine Learning Techniques," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-5, doi: 10.1109/ASIANCON55314.2022.9908873.

[2] S. Sakshi, A. K. Gupta, S. Singh Yadav and U. Kumar, "Face Mask Detection System using CNN," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 212-216, doi: 10.1109/ICACITE51222.2021.9404731.

[3] N. Kowalczyk, M. Sobotka and J. Rumiński, "Mask Detection and Classification in Thermal Face Images," in IEEE Access, vol. 11, pp. 43349-43359, 2023, doi: 10.1109/ACCESS.2023.3272214.

[4] D. Singh and S. K. Joshi, "Masked Face Detection using Lightweight Deep Learning based Model for IoT based Healthcare applications," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-7, doi: 10.1109/MysuruCon55714.2022.9972485.

[5] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.

[6] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 66518-66520.

[7] C. R. Alimboyong, A. A. Hernandez and R. P. Medina, "Classification of Plant Seedling Images Using Deep Learning," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 1839-1844.

[8] Salman Khan, Hossein Rahmani, Syed Afaq Ali Shah, Mohammed Bennamoun, "A Guide to Convolutional Neural Networks for Computer Vision, " 2018, pp. 56.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
No funding was received for conducting this study.

Conflict of Interest
The authors have no conflicts of interest to declare that are relevant to the content of this article.