International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A RELIABLE MASKED FACE RECOGNITION SYSTEM FOR PERSON'S SAFETY AT PUBLIC PLACES

Rajesh Kumar Rai[1*], Shambhu Shankar Bharti[2*] and Neelam Dwivedi[3*]

[1*] Computer Science & Engineering Department, Aryabhatta Knowledge University, Patna, Bihar, India.

Email: rajeshkrrai16@gmail.com

[2*]Assistant Professor, Computer Science & Engineering Department, LNJPIT Chapra, Chhapra, Bihar, India.

Email: shambhu4u08@gmail.com

[3*] Assistant Professor , Computer Science & Engineering Department, United College of Engineering & Research, Prayagraj, Uttar Pradesh, India.

Email: neelamdw@gmail.com

Abstract

Wearing face masks in public places has become common since the COVID-19 pandemic. It provides the safety of a person from various viral diseases and dust in polluted spaces. Wearing face masks in public places is a significant concern for public safety as it has become a pervasive pattern that a person comes wearing a mask, performs a crime, and escapes. This crime pattern is a significant concern for security agencies for surveillance work, identity verification, and many other tasks. Identifying that person has become a tedious task for the security agency. Therefore, to identify a person even while they are wearing a face mask, an automated and efficient masked facial recognition system is required. This research proposed a reliable model to identify a person's identity even when wearing a facemask using the FaceNet model. The proposed model in this work enhances security by incorporating a robust pipeline that includes masked face detection, alignment, and embedding extraction using a pre-trained FaceNet network. These embeddings have been compared using distance-based metrics to verify or identify individuals, even when facial regions are partially occluded. The effectiveness of the suggested work is confirmed by testing using masked images, which yields an accuracy of over 97 % when the model is trained using either masked images or a combination of masked and unmasked images. Results from experiments demonstrate that even when a person is wearing a facemask, the suggested model can correctly identify them.

Keywords: Masked Face Recognition, FaceNet, MaskTheFace, SVM

1 Introduction

A face recognition system automatically recognizes or verifies a person by applying facial characteristics or features. Traditional face recognition systems identify or verify a person by extracting facial features like the nose, mouth, jawline, and overall facial structure. Face recognition systems are vital in most computer vision and machine learning applications like public safety, surveillance, user authentication, and many more.

However, due to the COVID-19 pandemic, using face masks in public places is the new normal, introducing a significant challenge to existing face recognition technologies. Recently, we have seen many instances like terror attacks in a Hotel in Bengaluru [1] and many more [2, 3] in which attacker wears face masks to hide their identity. This is happening worldwide, as per the data revealed by the Los Angeles Police Department. There were 1,164 criminal incidents in which a suspect was wearing a mask from Jan 1 to Oct 31, 2020, while during the same period the year before, there were just 200 such reports [4]. Therefore, it can be said that most people wear face masks to avoid contagious diseases, while criminals use them to hide their identity during crime. According to previous studies, the recognition accuracy rate of current facial recognition technology for unmasked faces is 97.7% [5].

On the other hand, the accuracy decreases to 50% for masked faces, and the algorithm occasionally malfunctions, rendering the present technology extremely ineffective. Wearing face masks reduces accuracy because the face is over 50% obscured, particularly the nose and mouth, which makes it challenging for traditional face recognition algorithms to correctly identify or validate the face. Additionally, the accuracy of facial recognition systems is also impacted by the forms and colors of the mask.

As an alternative to face mask detection, there are other biometric technologies, such as iris and finger point recognition, among many others. However, avoiding social contact and wearing a face mask are the best ways to stop the virus from spreading, according to the US Centers for Disease Control and Prevention [6]. Therefore, due to the contagious behavior of diseases like COVID-19, using a biometric verification system like finger point recognition that involves the person in contact is unsafe. However, an iris recognition system does not require a person in contact, but for such a system, the person's iris should be scanned very closely. Apart from this, the security personnel keep surveillance at public places like roads, airports, Railway stations, and malls using CCTV Cameras. Therefore, utilizing the Iris detection mechanism for surveillance purposes in public places is not a feasible solution.

In the current scenario, when social distancing and wearing a face mask is nec-essary, there is a need for an improved face recognition system that can identify or verify a person by looking at the masked face. This will improve the person's safety, property safety, and surveillance in public places, especially in high-security or public-facing applications. Thus, the Masked Face Recognition system identifies or verifies a person even if their face is partially occluded by wearing masks. However, the main issue in this task is extracting the relevant features using only the upper parts of the face, like eyes, eyebrows, and forehead. This issue is minimized by using deep learn-ing because, in deep learning, the user only passes the raw image, and the machine extracts the relevant features. Apart from deep learning, the transfer learning concept provides high accuracy in such systems. Moreover, researchers have proposed various techniques using pre-trained models like FaceNet [7], VGGFace [8], and ArcFace [9]. Some techniques detect the mask in phase 1 while extracting the relevant features from the unmasked face areas. Thus, robust and reliable masked face recognition systems are necessary for public safety and

International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

surveillance in public places like airports, malls, and Railway stations [10–12].

In this work, we have proposed a Reliable Masked Face Recognition System for person's Safety in Public Places that overcomes the challenge of partial facial occlusion due to face masks. The proposed work utilizes the pre-trained model FaceNet for mapping the facial images to a compact embedding space while using Support Vector Machines (SVM) to classify the person. The proposed work's strength lies in its ability to recognize faces using features that remain visible even when a mask is worn, primarily the eyes, forehead, and upper face structure.

Next section discusses the literature review of Masked Face Recognition System.

2 Literature Review

This section critically reviews the Face recognition System. The facial recognition sys-tem can be broadly divided into two categories based on feature extraction. The first categories of the system recognize the face by using explicitly identified features while later extracting the feature by itself only by providing raw images. Additionally, face features may include geometry features such as the length and width of the nose, the distance between the mouth and the nose, and texture features such as SURF (Speeded Up Robust Features), SIFT (Scale-Invariant Feature Transform), and LBP (Local Binary Patterns) [13]. Additionally, a lot of researchers employed appearance-based features such facial landmarks like the corners of the eyes and the points of the nose, as well as PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) [14]. In contrast, later one utilizes deep learning models like Facenet, Deep-Face [15], VGGFace, and many more to extract the features from the raw images. Explicit extracted feature-based approaches are simple and computationally cheap but sensitive to pose, occlusion, variations in illumination, and noise. In contrast, later on, it is accurate and robust to pose, lighting, and occlusion while needing a large dataset for proper training. Therefore, a deep learning-based approach is preferable for the Masked Face Recognition System. Liao et al. [16] proposed a deep learning method using CNNs to handle occlusion in general face recognition tasks. They presented Multi-Keypoint Descriptors using the alignment-free face representation method. His work showed that their proposed research is superior in recognizing partial faces without requiring alignment, but it is still insufficient to detect the face correctly while wearing a mask. Taigman et al. [15] introduced DeepFace, which represents a face from a nine layer deep neural network that can handle alignment and occlusion to some extent but is not robust enough to mask.

Moreover, In recent times, researchers have used classical network structures or pre-trained networks like AlexNet [17], VGGNet [18], GoogleNet [19], ResNet [20], and many more, which are widely used as the backbone networks of face recognition models [21]. These models are fine-tuned by applying the transfer learning approaches. Wang et al. (2020) [22] fine-tuned the existing ResNet-50 [20] model by training with the masked face datasets, showing improved accuracy over standard models. Anwar and Raychowdhury [23] proposed Masked Face Recognition for Secure Authentication using transfer learning on pre-trained CNNs like VGG16 and MobileNet to adapt to masked datasets. However, a masked dataset is required to fine-tune the existing pre-trained model. This requires a large amount of masked images. In recent research, various masked image datasets, like Real-world Masked Face Dataset

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(RMFRD) [22] and Synthetic Masked Face Dataset (SMFRD) [24], still need many images with several types of masks. To solve this problem, many researchers have presented techniques like MaskTheFace [23], which can convert an unmasked image into masked images with various masks. This reduces the scarcity of images.

Still, there are many challenges in the masked face recognition system, like Occlusion-Invariant Learning, reliability, Loss of discriminating features, increased intra-class variation (same person looks different with a mask), and decreased inter-class variation (different people look similar when masked). Therefore, in this paper, we have presented a reliable masked face recognition system that can identify whether a person wears a mask or not in an accurate way.

The following Section discusses the proposed approach in detail.

3 Proposed Approach for Masked Face recognition

The suggested method's flow chart is displayed in Fig. 1. Training and testing are the two components of the suggested methodology. Raw images are used as input during the training phase. Raw images are then preprocessed. In preprocessing, two operations are performed: resizing and normalizing. The proposed work converts all input images into equal sizes of 160 * 160. Once all the images are converted into equal sizes then, the normalizations of the images are done using the following three operations:

- **Pixel Normalization**- It scales pixel values in the range of 0-1.
- **Pixel Centering**- It scales pixel values so that the pixel's mean becomes 0.
- **Pixel Standardization**- It scales pixel values so that the mean and variance of the pixel become 0 and 1, respectively.

Following image normalization, the face and facial landmarks are identified using Multi-task Cascaded Convolutional Networks (MTCNN) [25], a 2016 algorithm that uses a cascading series of neural networks to quickly and accurately detect, align, and extract facial features from digital images. The MTCNN algorithm performs the

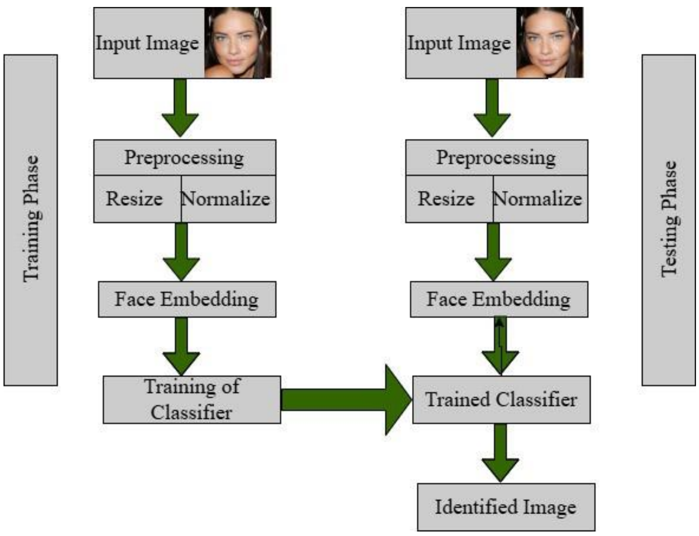


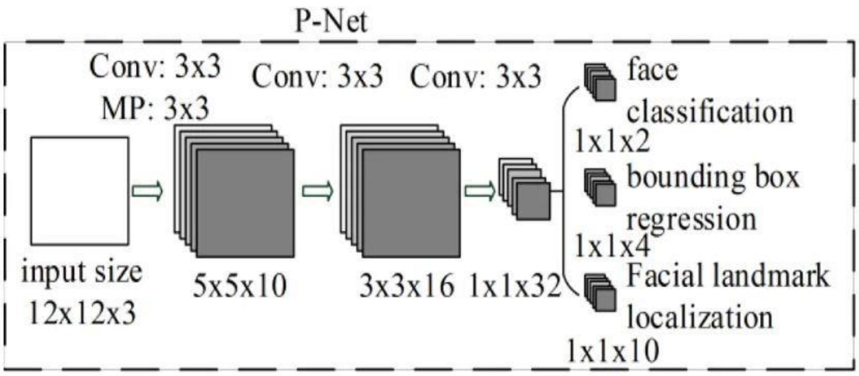
Fig. 1: Flow Chart of the Proposed Approach

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

above tasks in three phases with the help of three networks: Proposal network (P-Net), refinement network (R-Net), and output network (O-Net).

- **Proposal Network (P-Net):** P-Net is used in the initial step to build a set of potential bounding boxes that might contain a face. Convolutional filters are used to generate feature maps from raw images, and then fully connected layers are applied to forecast the likelihood of a face. Figure 2 presents the block diagram of the Proposal Network.

Fig. 2: Block Diagram of the Proposal Network [25]



- **Refinement Network (R-Net):** In the second phase, candidate bounding boxes are taken as input and crop the corresponding regions. These trimmed areas are also scaled, and to identify the area as a face, a number of convolutional and fully connected layers are used. Figure 3 presents the block diagram of the Refinement Network.

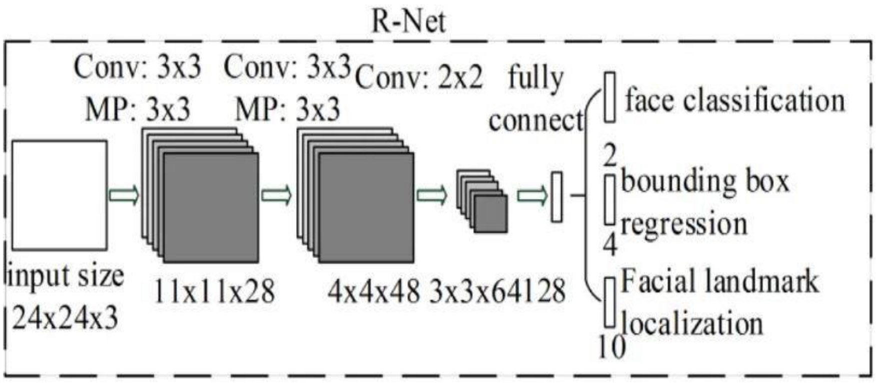


Fig. 3: Block Diagram of the Refinement Network [25]

- **Output Network (O-Net):** In the third phase, the matching areas of the input image are cropped using the R-Net's revised bounding boxes as input. Furthermore, again by applying a series of convolutional and fully connected layers, it refine the location of the detected face further followed by extraction of facial landmarks. Figure 4 presents the block diagram of the Output Network.

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

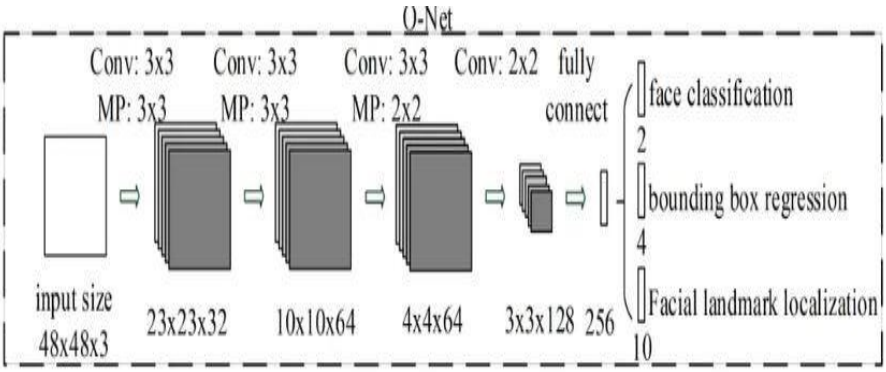


Fig. 4: Block Diagram of the Output Network [25]

Once the facial landmarks are identified, face embedding is done using the pretrained FaceNet model. FaceNet model was proposed by Google Researchers in 2015. The figure 5 shows the architecture of the FaceNet model. Inception network has been used

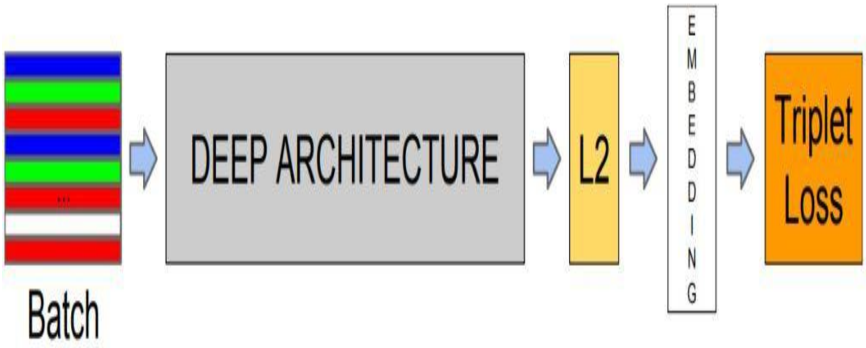


Fig. 5: FaceNet Model [7]

as Deep Architecture in the proposed work. The Figure 6 shows the Inception Layer used in the FaceNet Model that is used to embed the faces of an input image.

type	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj (p)	params	FLOPS
conv1 (7×7×3, 2)	112×112×64	1							9K	119M
max pool + norm	56×56×64	0						m 3×3, 2		
inception (2)	56×56×192	2		64	192				115K	360M
norm + max pool	28×28×192	0						m 3×3, 2		
inception (3a)	28×28×256	2	64	96	128	16	32	m, 32p	164K	128M
inception (3b)	28×28×320	2	64	96	128	32	64	L_2, 64p	228K	179M
inception (3c)	14×14×640	2	0	128	256,2	32	64,2	m 3×3,2	398K	108M
inception (4a)	14×14×640	2	256	96	192	32	64	L_2, 128p	545K	107M
inception (4b)	14×14×640	2	224	112	224	32	64	L_2, 128p	595K	117M
inception (4c)	14×14×640	2	192	128	256	32	64	L_2, 128p	654K	128M
inception (4d)	14×14×640	2	160	144	288	32	64	L_2, 128p	722K	142M
inception (4e)	7×7×1024	2	0	160	256,2	64	128,2	m 3×3,2	717K	56M
inception (5a)	7×7×1024	2	384	192	384	48	128	L_2, 128p	1.6M	78M
inception (5b)	7×7×1024	2	384	192	384	48	128	m, 128p	1.6M	78M
avg pool	1×1×1024	0								
fully conn	1×1×128	1							131K	0.1M
L2 normalization	1×1×128	0								
total									7.5M	1.6B

Fig. 6: Inception Layer used in FaceNet Model [7]

Received: August 04, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

FaceNet model embeds each image into a vector X = [e_1, e_2, ..., e_n], where each $e_i \in R^{128}$.

After the image embedding is done, the Support Vector Machine Classifier is used for classification purposes. A Support Vector Machine has been used as a classifier because it works well with small to medium sized data and performs excellently in high dimensional space. Moreover, SVM also avoids overfitting with kernel trick and margin maximization. In this proposed work, the SVM learns a decision boundary in the embedding space that separates identities.

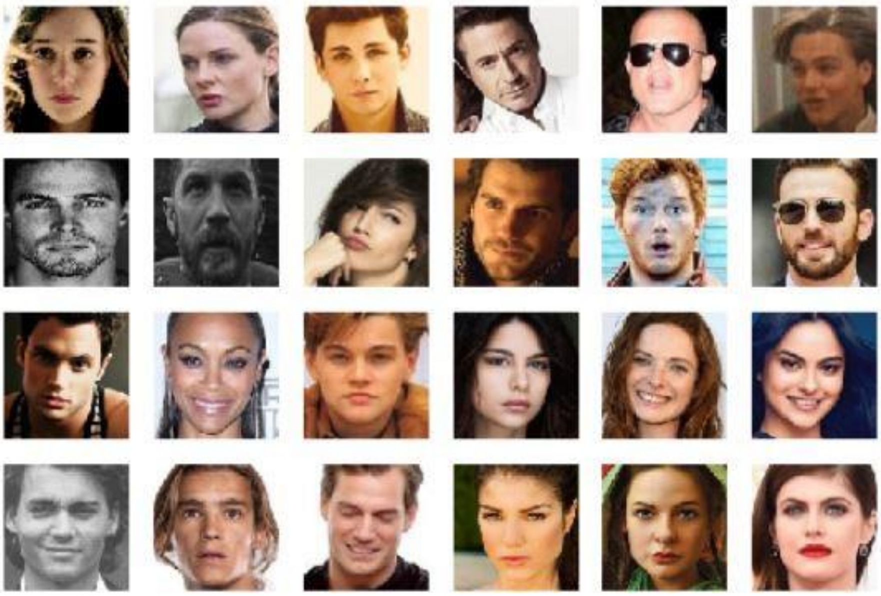
Once the training phase is completed, then in the testing phase, the embedding of an image is done in the same way as the training phase, and then the embedded vector is passed through the trained SVM classifier for the classification.

The next section, presents the details of experimentation.

4 Experimentation

The Pins Face Recognition dataset [26] is used to validate the proposed approach for reliable masked face detection. The images of this dataset have been collected from Pinterest and cropped. This dataset contains the images of 105 celebrities containing 17534 faces. All these images are unmasked images. Figure 7 shows some sample images of this dataset.

Fig. 7: Sample Image of PINS Datatset



To confirm the reliability of the proposed approach, three experiments are conducted.

- Experiment #1: Training on unmasked images and testing on masked images.
- Experiment #2: Training on a combined dataset (masked + unmasked), and
- Experiment #3: Training and testing on only masked images.

These experiments allowed us to assess the generalization ability of the model under different conditions.

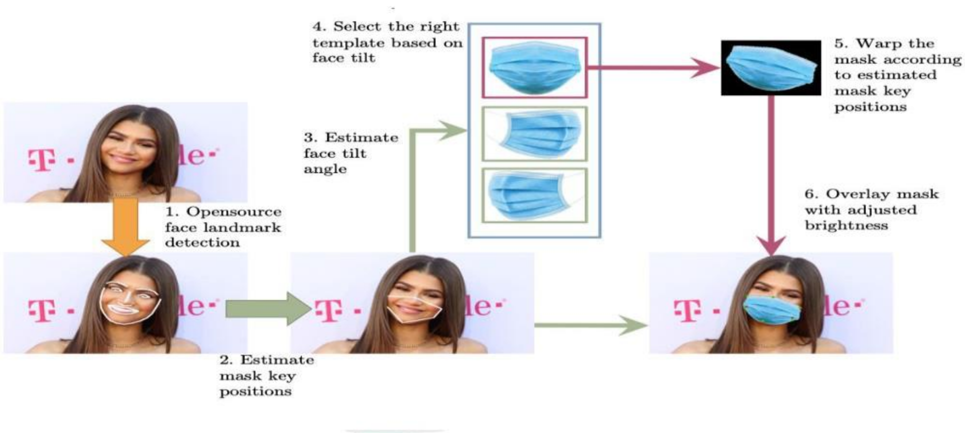
International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Since the PINS Dataset contains only unmasked images, the next task is to pre-pare the masked images for the conduction of the second and third experiments. The following subsection details the process of obtaining masked images from unmasked images.

4.1 Collection of Masked Images

The MaskTheFace tool has been used in the proposed approach to convert masked images from unmasked ones. MaskTheFace tool detects the face in an image using a dlib-based face landmark detector. The Figure 8 shows the block diagram of the MaskTheFace tool working.

Fig. 8: Block Diagram of MaskTheFace [26]



MaskTheFace provides several types of masks from which to select. The MaskTheFace tool reduces limitations like face angle, mask fit, lighting conditions, and many more and applies a mask on the face.

Currently, the MaskTheFace applies five types of masks on the faces of unmasked images: Surgical, N95, KN95, Cloth, and Gas. The Figure 9 shows sample images of masked faces obtained by applying the MaskTheFace tool.

4.2 Experiment #1

In Experiment #1, all the unmasked images are used for training. Before starting the

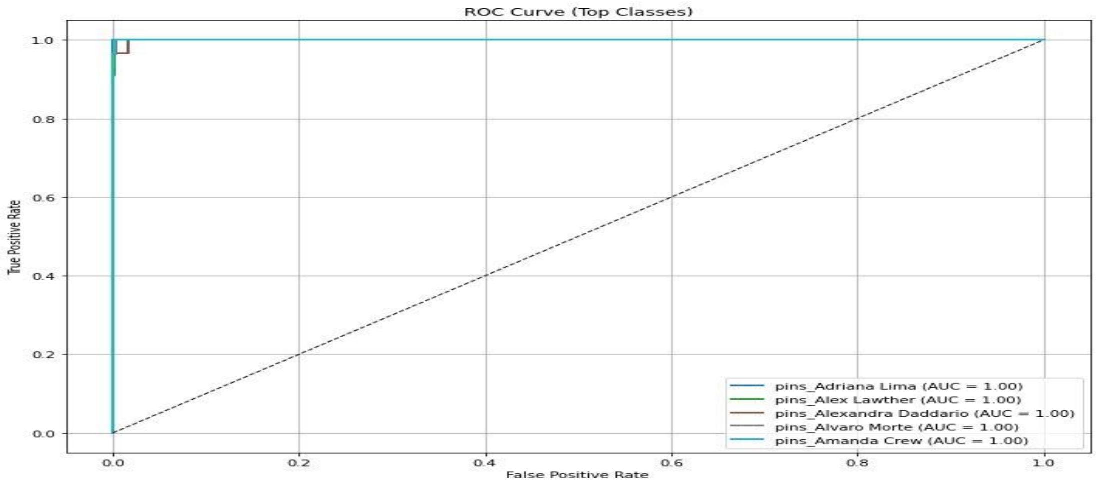
International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

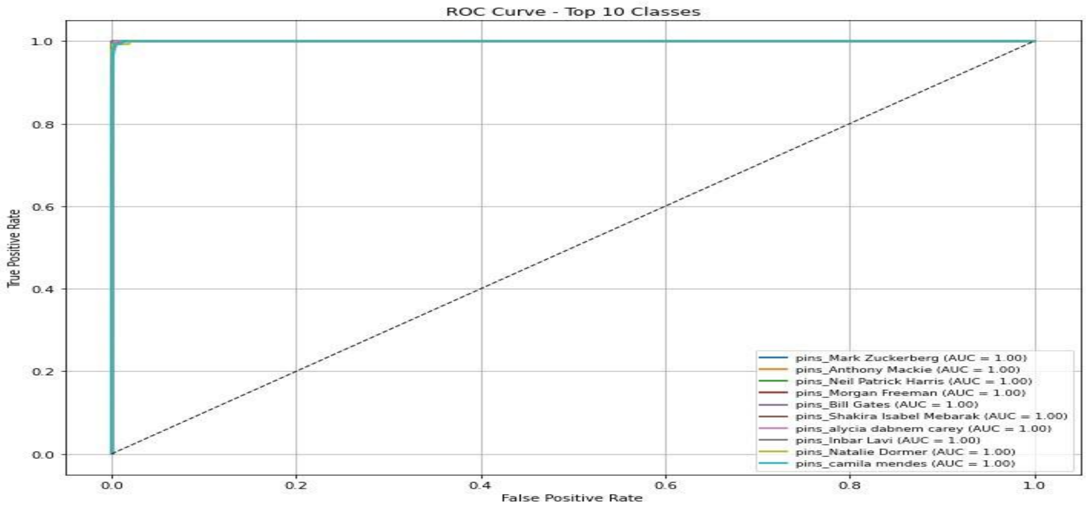
training, 150 unmasked images of each person are stored in the training directory. In addition, each user has five masked images saved in a testing folder. Thereafter, the embedding of each image in the training folder and testing folder correspondingly is obtained by applying the FaceNet model. It gives a 128-dimensional vector of each image. The embeddings of images in the training folder are then passed to the support vector machine classifier for training purposes. Once the model is trained, the 128-dimensional vector corresponding to each image is passed to the trained support vector machine classifier. Accuracy obtained for the experiment #1 is 86.67%. Since, in this experiment only unmasked images are used for training and only masked images are used for testing purposes, it is par from the other existing approaches.

4.3 Experiment #2



In Experiment #2, masked and unn ages are used for training. Before starting the

10.(a)



training, 150 unmasked and 150 mask obtained using the MaskTheFace tool for each person were stored in the training di n addition, each user has five masked images saved in a testing folder. Thereafter, the embedding of each image in the training folder and testing 28- dimen ing the image embeddings from the training folder. The trained support vector machine classifier

10.(b)

Fig.10: ROC curve for (a). Five person's image (b). Ten person's image

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

receives the 128-dimensional vector that corresponds to each image after the model has been trained for identification. Accuracy obtained for the experiment #2 is 97.29%. The Figure 10 a and b shows the ROC curve of five different person's image and ten different person's image.

From this ROC curve, it can be seen that the Area Under Curve (AUC) =1.0 confirms the effectiveness of the proposed approach. Figure 11 shows the confusion matrix of the proposed approach for Experiment #2. Since most predictions of this confusion matrix are diagonal, it confirms the accurate classification of the proposed approach.

As can be seen, more than 97% accuracy is obtained for the proposed approach; thus, it can be used for real-life applications. Apart from this, since masked images are used only by applying the FaceTheMask tool for training purposes, the high accuracy obtained provides its applicability for public safety. Since unmasked images of a person

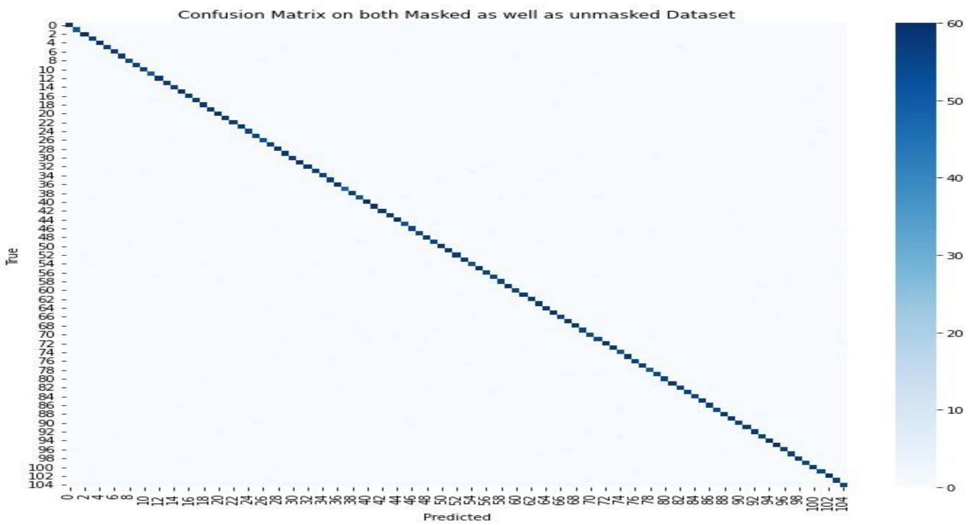


Fig. 11: Confusion Matrix for Experiment #2

are readily available to any security agency, masked images of the same person can be obtained using the MaskTheFace tool. Once the masked and unmasked types of images are collected, the model can be easily trained and help identify the criminals quickly. It will help reduce crime and increase people's safety in public places.

4.4 Experiment #3

In Experiment #3, masked images obtained using the MaskTheFace tool for each person are used for training. Five masked images of each person are also used for the test. The embedding of each masked image for training and corresponding testing is obtained by applying the FaceNet model. It gives a 128-dimensional vector of each image. The embeddings of images used for the training are then used to train the support vector machine classifier. Once the model is trained, the 128-dimensional vector corresponding to each image is passed to the trained support vector machine classifier. Accuracy obtained for the experiment #3 is 97.66%.

The Figure 12 a and b shows the ROC curve of Ten different person's image and confusion matrix for the proposed approach of Experiment #3. From this Figure, it can be seen that the Area Under Curve (AUC) =1.0 confirms the effectiveness of the proposed approach. Apart

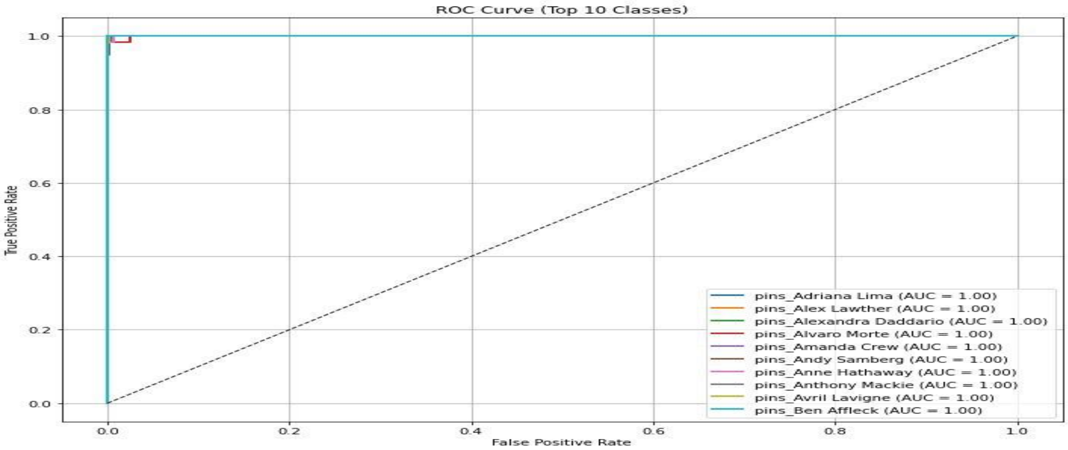
Received: August 04, 2025

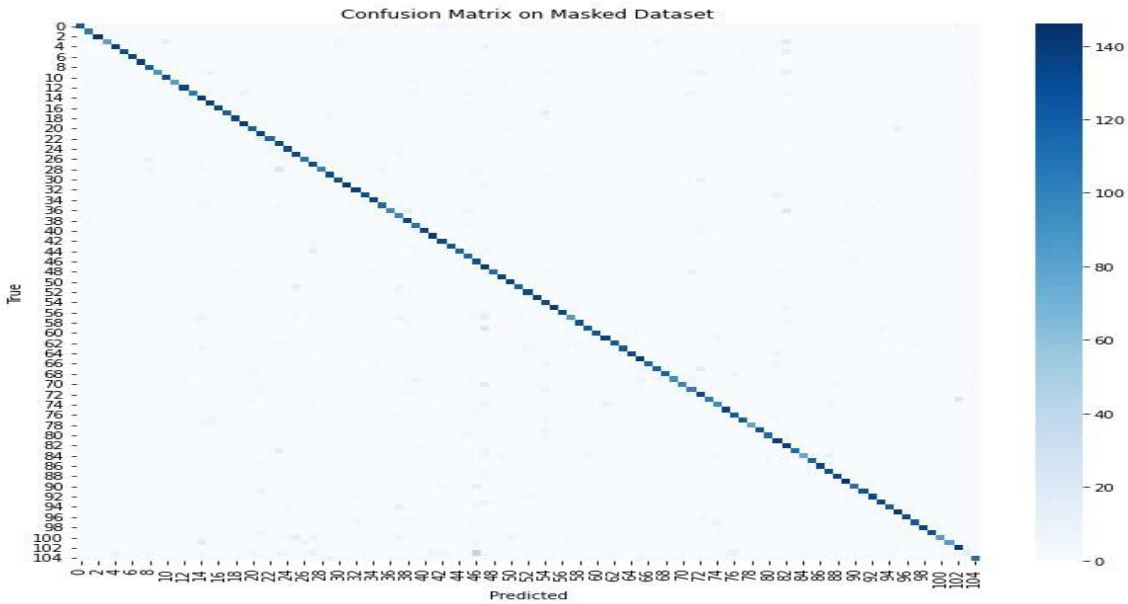
International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

from this, it is seen that most predictions of this confusion matrix are diagonal, it confirms the accurate classification of the proposed approach. As can be seen, more than 97% accuracy is obtained for the proposed approach; thus, it can be used for real-life applications. Moreover, by comparing the results of Experiment #2 and Experiment #3, it can be concluded that in both cases, accuracy is approximately the same and higher than 97%, confirming the reliability of the





12.(b)

Fig. 12: ROC curve for a. Ten person's image b. Confusion Matrix for Experiment #3

proposed approach. Therefore, by seeing the results of the experiments, it can be concluded that the proposed approach can be used to identify whether the criminals are wearing face masks or not. Also, by utilizing the proposed approach, surveillance can be done more effectively in public places and, hence, will increase the person's safety in public places like malls, roads, Restaurants, Railway Stations, and many more.

5. Conclusions

The primary goal of the proposed work is to develop a new reliable model that can identify a person even wearing face masks. A reliable model using FaceNet and SVM is presented in this work. The PINS image dataset is used to validate the proposed work. Further, to confirm the reliability of the proposed approach, three experiments are conducted in various scenarios. In the first experiment, the model is trained with unmasked images; in the second experiment, the model is trained with the combi-nation of masked and unmasked images, while in the third experiment, the model is trained with masked images. In all three experiments, the model is tested with masked images only. Experimental results reveal that more than 97% accuracy is obtained for experiment #2, and experiment #3 confirms the effectiveness of the proposed approach. Moreover, there is only a 0.4% difference in accuracy for experiment #2 and experiment #3, which confirms the reliability of the proposed approach.

Statements and Declarations: No funding was received to assist with the preparation of this manuscript. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Acknowledgements: I want to express my heartfelt gratitude to the supervisor, Dr. Shambhu Shankar Bharti, and DRC members, who have critically reviewed our work and provided relevant feedback and all support related to this research work. I also express my sincere gratitude to all the Aryabhatta Knowledge University, Patna, Bihar officials, for providing us with all the institutional support during our research work.

Data availability statement

All data that support the findings of this study are included within the article.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. E.W. Desk. Bengaluru rameshwaram cafe blast: What we know so far (2024). URL: https://indianexpress.com/article/cities/bangalore/bengalururameshwaram-cafe-blast-9190110/

2. E.W. Desk. One year later, the ‘masks’ still help jnu attacker escape law (2021). URL https://sabrangindia.in/ one-year-later-masks-still-help-jnu-attacker-escape-law/

3. D.W. Team. Karnataka: Man stabbed to death by masked attackers, third attack in 8 days (2022). URL: https://www.dnaindia.com/india/ report-man-stabbed-to-death-by-masked-attackers-in-karnataka-2972382

4. E. Ward. A spate of mask-wearing criminals (2020).URL:

https://xtown.la/2020/11/23/criminals-wear-masks-covid/#:~:text=From%20Jan.

%201ct.%2031%2C%20there%20were%201%2C164%20criminal,Don%E2%80%99t%20expect%20the%20trend%20to%20reverse%20itself%20soon.

5. S. Singh, S. Prasad, Techniques and challenges of face recognition: A critical review. Procedia computer science **143**, 536–543 (2018)

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

6.H.R. Güner, I˙. Hasanoğlu, F. Akta¸s, Covid-19: Prevention and control measures in community. Turkish Journal of medical sciences **50**(9), 571–577 (2020)

7.F. Schroff, D. Kalenichenko, J. Philbin, *Facenet: A unified embedding for face recognition and clustering*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 815–823

8. K. Balamurali, S. Chandru, M.S. Razvi, V.S. Kumar, *Face spoof detection using vggface architecture*, in *Journal of Physics: Conference Series*, vol. 1917 (IOP Publishing, 2021), p. 012010

9.J. Deng, J. Guo, N. Xue, S. Zafeiriou, *Arcface: Additive angular margin loss for deep face recognition*, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 4690–4699.

10.S.S. Bharti, R.K. Rai, N. Dwivedi, Random forest based robust human activity detection system using hybrid feature. Engineering Research Express **6**(3), 035203 (2024).

11.N. Dwivedi, D.K. Singh, D.S. Kushwaha, Orientation invariant skeleton feature (oisf): a new feature for human activity recognition. Multimedia Tools and Applications **79**(29), 21037–21072 (2020).

12.N. Dwivedi, D.K. Singh, D.S. Kushwaha, A novel approach for suspicious activity detection with deep learning. Multimedia Tools and Applications **82**(21), 32397–32420 (2023).

13.A. Gupta, D. Raj, *Comparative Analysis of Texture-Based Algorithms LBP, LPQ, SIFT, and SURF Using Touchless Footprints*, in *International Conference on Artificial Intelligence and Sustainable Engineering: Select Proceedings of AISE 2020, Volume 2* (Springer, 2022), pp. 423–439.

14. Pattnaik, A. Dev, A. Mohapatra, A face recognition taxonomy and review framework towards dimensionality, modality and feature quality. Engineering Applications of Artificial Intelligence **126**, 107056 (2023).

15. Y. Taigman, M. Yang, M. Ranzato, L. Wolf, *Deepface: Closing the gap to human-level performance in face verification*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2014), pp. 1701–1708.

16. R. Liao, A. Schwing, R. Zemel, R. Urtasun, Learning deep parsimonious representations. Advances in neural information processing systems **29** (2016).

17. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (2017).

18. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).

19. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Van-houcke, A. Rabinovich, *Going deeper with convolutions*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 1–9.

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
20. K. He, X. Zhang, S. Ren, J. Sun, *Deep residual learning for image recognition*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

21. Y. Ge, H. Liu, J. Du, Z. Li, Y. Wei, Masked face recognition with convolutional visual self-attention network. Neurocomputing **518**, 496–506 (2023).

22. Z. Wang, B. Huang, G. Wang, P. Yi, K. Jiang, Masked face recognition dataset and application. IEEE Transactions on Biometrics, Behavior, and Identity Science **5**(2), 298–304 (2023).

23. A. Anwar, A. Raychowdhury, Masked face recognition for secure authentication. arXiv preprint arXiv:2008.11104 (2020).

24. B. Huang, Z. Wang, G. Wang, K. Jiang, Z. He, H. Zou, Q. Zou, *Masked face recognition datasets and validation*, in *Proceedings of the IEEE/CVF international conference on computer vision* (2021), pp. 1487–1491.

25. K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks. IEEE signal processing letters **23**(10), 1499–1503 (2016).

26. StephENCT. Pins dataset (2023). URL https://www.kaggle.com/datasets/ troyct/105-classes-pins-dataset