MASKED FACE DETECTION AND RECOGNITION USING YOLOV11 AND ARCFACE FOR HEALTHCARE APPLICATIONS

HSIAO-JU LIN [1], SHI-JINN HORNG [2], PIN-SIANG HUANG [3], AND NANDO ALHASANAIN [2]

[1]Department of Orthopedic Surgery, Taoyuan General Hospital, Ministry of Health and Welfare, Taoyuan, Taiwan
[2]Department of Computer Science and Information Engineering Asia University; Department of Medical Research, China Medical University Hospital, China Medical University, Taichung Taiwan Taichung, Taiwan
[3]Agaruda Research, Agaruda, Taipei, Taiwan
E-MAIL: ruru1300837@gmail.com, horngsj@yahoo.com.tw, sean.huang@agaruda.io, nandoalhasanain@gmail.com

Abstract:

Widespread mask usage in healthcare settings has enhanced infection control but poses substantial challenges for conventional face recognition systems due to severe facial occlusion. This occlusion significantly degrades recognition accuracy, limiting the practicality of biometric authentication in real-world clinical environments. To address this issue, we propose a robust masked face authentication framework that integrates real-time mask-state detection with deep face verification to mitigate mask-induced appearance variations. The proposed system employs a YOLO11-based module to simultaneously localize faces and classify mask-wearing status into three categories: no mask, correct mask usage, and incorrect mask usage. The detector is trained on multiple public datasets, including AIZOO, FMD, and MMD, to enhance generalization across diverse scenes and imaging conditions. For identity verification, the system leverages ArcFace-based facial embeddings combined with a multi-metric similarity fusion strategy. In addition, a mask-aware preprocessing scheme is introduced, which expands detected bounding boxes to retain discriminative periocular and upper-face features that remain visible under mask occlusion. Extensive experiments demonstrate that the proposed detection module achieves a face detection rate of 99.0% and a three-class mask classification accuracy of 95.57%. Furthermore, evaluation on masked–unmasked verification scenarios using the RMFD dataset and an internal identity dataset yields an accuracy of 0.9632, a ROC AUC of 0.9854, and an equal error rate (EER) of 0.0368 at an optimized decision threshold of approximately 0.6245. These results indicate that incorporating mask-state awareness with discriminative face embeddings effectively preserves reliable authentication performance in mask-dominated environments.

Keywords:

Masked face recognition; Face mask detection; YOLO; ArcFace; Healthcare authentication; COVID-19

1. Introduction

Biometric technologies are increasingly used in security systems, especially in healthcare, where protecting patient data and controlling access to restricted areas are critical. Facial recognition is preferred due to its contactless and non-intrusive nature, supporting hygiene and reducing contamination risks. It also enhances physical access control (e.g., automated doors) and is more secure and convenient than RFID cards, which can be lost or duplicated [1].

The COVID-19 pandemic significantly impacted facial biometrics, as mandatory mask use causes severe facial occlusion, hiding key features like the nose and mouth. This disrupts the geometric cues used by traditional recognition systems, creating the "masked face recognition problem" and prompting solutions that balance security and public health needs [2].

Existing methods struggle to achieve both security-level accuracy and real-time performance. Early CNN approaches like VGG-16 showed feasibility [3], [4] but are computationally heavy and may exhibit bias [5]. Multi-task models such as DeepMaskNet improve integration [6], while YOLO-based detectors (e.g., YOLOv5, YOLOv7) enhance speed and accuracy [7][8][9], including combinations with InsightFace [9]. Despite advances and the need for liveness detection and XAI [10], the key challenge remains extracting discriminative features from the limited visible upper-face region.

This study proposes a masked face recognition system for healthcare that combines real-time mask detection with robust verification. It uses YOLO11 to localize faces and classify mask status, offering faster inference than earlier YOLO models [7], [8]. For identity verification, ArcFace is adopted to enhance discriminative power under occlusion, outperforming traditional losses and FaceNet-based methods

[11], and proving effective for periocular recognition [12]. By integrating YOLO11 and ArcFace, the system addresses limitations of prior lightweight approaches [13], enabling secure, real-time authentication without removing masks.

The research follows a structured quantitative approach, including dataset aggregation, model training, and evaluation using standard metrics such as Accuracy, Precision, and F1Score.

This study is organized into five sections. Section 2 reviews related work on masked face recognition, object detection, and angular-margin loss functions. Section 3 describes the methodology, including system design and dataset preparation. Section 4 presents experimental results and analysis. Section 5 concludes the work and outlines directions for future research in healthcare-oriented biometric.

2. Related work

Masked face identification is a major challenge for biometric systems, as traditional methods relying on facial features suffer significant accuracy loss under occlusion. This section reviews research on face mask detection, masked face recognition, and general face recognition.

2.1. Face mask detection

Numerous studies have explored mask detection, with YOLO-based methods being widely used. Dewi and Christanto [7] showed that a YOLOv5 model can accurately classify mask usage into three categories: incorrect, with mask, and without mask. By combining the FMD and MMD datasets, their work demonstrates the effectiveness of YOLOv5 for mask classification.

The research by Ibitoye and Osaloni [5] focuses on classifying masked faces using a deep convolutional neural network based on the VGG-16 architecture. This system is aiming to improve accuracy, especially for dark-skinned individuals. Their system addresses biases in existing methods and achieves strong performance with 98% accuracy.

In addition, Kumar et al. [8] compared YOLOv3, YOLOv4, and YOLOv5 for mask detection, emphasizing the trade-off between speed and accuracy. Their results show YOLOv5 performs best on large, complex datasets, and they introduced a more realistic dataset to enhance practical applicability.

2.2. Masked face recognition

On the other hand, the facial recognition system with masks confronts substantial hurdles due to the loss of critical information on the covered regions of the face. To address this, Shahar and Mazalan [14] developed a system that combines deep learning–based mask detection with Procrustes analysis for identity matching. Their approach measures facial similarity through discrepancy values and achieves 97.14% recognition accuracy.

To improve real-world performance, Ullah et al. [6] proposed DeepMaskNet, a multitask CNN that integrates mask detection and face recognition in a single model, independent of mask type. Experiments on a self-developed dataset show that DeepMaskNet outperforms other methods, particularly in accuracy across different datasets.

To support resource-constrained devices, Kocacinar et al. [13] developed a lightweight CNN-based facial recognition system for mobile platforms. Using transfer learning and fine-tuning, the system performs real-time mask detection and face recognition, achieving 90.40% validation accuracy on a dataset of 12 individuals and 1,849 facial samples.

2.3. Face recognition

Putri et al. [1] proposed a YOLO-based smart door lock system for secure access control. The system uses face recognition to unlock a solenoid mechanism upon successful identification, integrating a webcam, ESP32, and electronic lock. Testing showed a face detection accuracy of 94.4%.

On the other hand, Boyko et al. [15] compared Dlib and OpenCV for masked face recognition, finding that Dlib provides more stable performance when facial features are obscured, while OpenCV shows greater variability. Their study offers useful insights for selecting face recognition frameworks in masked environments.

In keeping with the preceding technique, Anjeana and Anusudha [9] developed a real-time face recognition system using YOLOv4 and InsightFace for both face detection and identity recognition. By leveraging ResNet-100 and cosine similarity, the system achieves 98.5% recognition accuracy, demonstrating strong real-world performance

3. Proposed method

This study develops a secure masked face recognition system by combining YOLO11 for face detection and ArcFace for feature extraction. The framework consists of four stages: (a) face detection and preprocessing, (b) embedding generation, (c) datasets and software components, and (d) face recognition and similarity evaluation.

The proposed workflow, shown in Figure 1, describes the complete process from training to verification. First, *Training and Loading YOLO11* trains a face-mask detection model using masked and unmasked images, with the resulting YOLO11 weights (.pt) saved for evaluation. Second, *Face Detection and Pre-processing* applies YOLO11 with multi-scale detection to localize faces, and detected regions are cropped with additional margins to preserve key features before generating the ROI for verification.

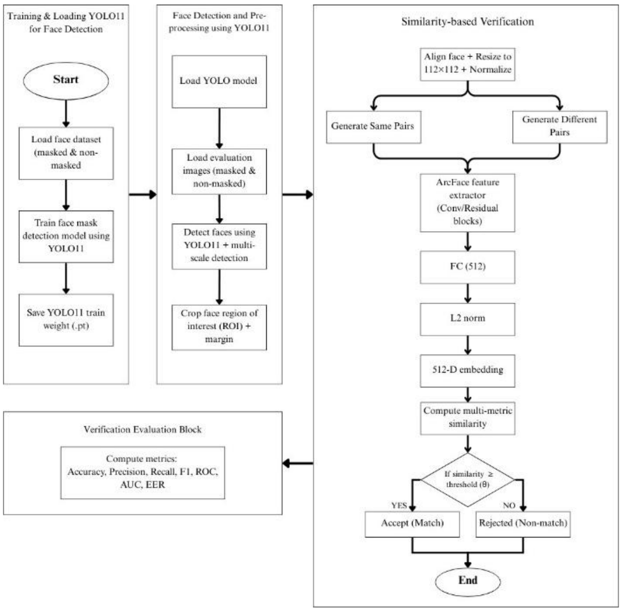

FIGURE 1. Overall system flowchart

Third, *Similarity-based Verification* evaluates genuine and impostor pairs by aligning, resizing (112 × 112), and normalizing faces before generating 512-dimensional ArcFace embeddings. An enhanced similarity score is computed, and pairs are classified as *Accept (Match)* or *Reject (Non-match)* using a threshold θ. Finally, the *Verification Evaluation Block* computes biometric metrics, including Accuracy, Precision, Recall, F1-score, ROC, AUC, and EER, to assess performance under masked conditions.

3.1 Face detection and pre-processing

In the first stage, face detection is performed using the YOLO11 model [16], trained to detect masked and unmasked faces with weights stored in *best.pt*. The YOLO11 architecture (Figure 2) consists of Backbone, Neck, and Head modules that extract features and generate predictions for face detection with and without masks.

The trained model acts as the backbone of the face detection system, identifying both masked and unmasked faces. After detecting the bounding box, the face region is cropped with an added 40% margin to preserve surrounding facial information.

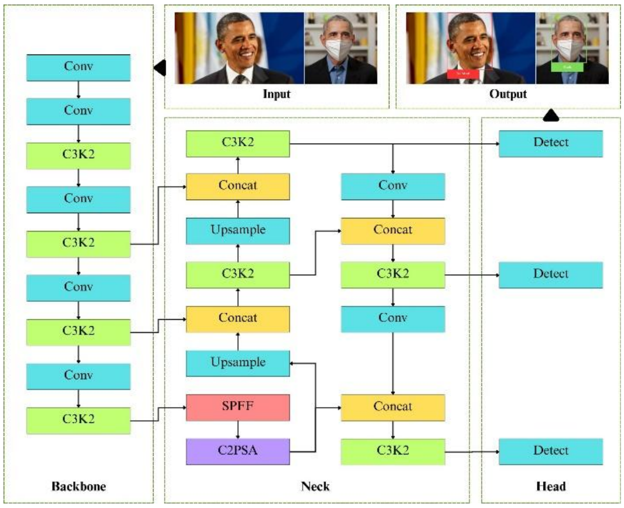

FIGURE 2. YOLO11 system architecture

3.2 Embedding generation

After face detection with the trained YOLO11 model, the system proceeds to facial feature extraction using ArcFace, a lightweight model optimized for masked face recognition.

ArcFace produces a high-dimensional embedding $e \in R^{512}$ that encodes identity-specific information and is commonly trained with an angular-margin loss to increase inter-class separation while reducing intra-class variation.

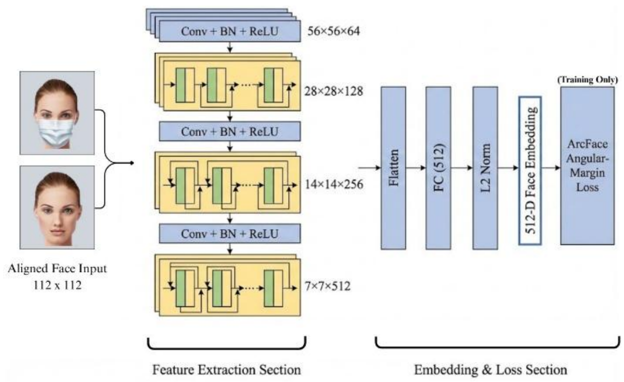

FIGURE 3. ArcFace embedding architecture used for generating 512-dimensional face representations

Although masks cause occlusion and distribution shifts, the visible periocular and upper-face regions still provide sufficient information for stable embedding extraction, especially with complementary features. The embedding architecture (Figure 3) processes a 112×112 aligned face through convolutional layers with batch normalization and ReLU, producing low-level features, followed by residual blocks that progressively reduce spatial size (56×56 → 28×28 → 14×14 → 7×7) while increasing channels from 64, 128, 256, to 512. This enables learning of higher-level periocular and facial structure features that remain visible under mask occlusion.

After the final convolutional stage, features are flattened and passed through a fully connected layer to produce a 512-dimensional vector, which is L_2-normalized for similarity computation. The angular-margin loss is only used during training to improve class separation and is omitted in the workflow since it focuses on inference. During evaluation, ArcFace acts as a fixed feature extractor, and only normalized embeddings are used for similarity scoring and threshold-based decision making.

3.3 Dataset and software component

To improve generalization, the YOLO11 detector is trained on a merged dataset consisting of AIZOO (7,703 images) [17], [18], FMD (763 images with three mask categories) [19], and MMD (670 medical mask images) [17]. After preprocessing, the combined dataset is used to train YOLO11 for face localization and margin expansion before classification and embedding extraction.

In addition to training data, this study uses Moxa3K as a benchmark for CCTV-style mask monitoring [20], and ISL-UFMD/ISL-UFHD datasets for diverse real-world, three-class mask recognition [21]. For masked face recognition, RMFD is used as the main benchmark due to its paired masked and unmasked identity images [22].

All experiments were conducted in Google Colab using a T4 GPU for faster inference. YOLO11 (via the Ultralytics library) was used for face detection and cropping, while a pre-trained ArcFace model from InsightFace handled face alignment and 512-dimensional embedding extraction. The pipeline was implemented in Python using OpenCV for image processing, NumPy/SciPy for computations and cosine similarity, scikit-learn for metrics and PCA for dimensionality reduction, Matplotlib for visualization, and Pickle for embedding storage, enabling comprehensive evaluation of masked and unmasked recognition.

3.4 Similarity computation and evaluation

After the YOLO11-based face detection and cropping, each facial image is processed through the ArcFace model (via the InsightFace library), which outputs a fixed-length 512 dimensional feature vector (embedding) $\mathbf{e} \in R^{512}$.

Relying on a single similarity measure can make verification sensitive to noise and distribution shifts. Therefore, this study fuses multiple metrics into a single similarity score. Besides cosine, Euclidean, Manhattan, and Mahalanobis distances, Pearson correlation and spectral similarity are also computed to produce metric-specific values, including S_{cos}, S_{euclid}, $S_{manhattan}$, S_{mahal}, $S_{pearson}$, $S_{spectral}$, respectively.

Before fusion, each score is normalized to a comparable range to prevent any metric from dominating due to scale differences. The final similarity is a weighted sum:

$$S_{final} = \sum_{i=1}^{m} w_i S_i, \quad (1)$$

where S_i is the normalized score of the i-th metric and w_i is its weight.

During inference, S_{final} is computed for each comparison and used for the final decision or ranking.

The similarity scores are then compared against a threshold τ, where a pair is classified as "same person" if sim $\geq \tau$ and "different person" otherwise. The optimal threshold τ^* is determined by evaluating a range of candidate values and selecting the one that *maximizes the classification accuracy*:

$$\tau^* = arg \max_{\tau}\left(\frac{TP(\tau)+TN(\tau)}{TP(\tau)+TN(\tau)+FP(\tau)+FN(\tau)}\right) \quad (2)$$

Where *True Positives (TP)*: correctly classified same identity pairs; *True Negatives (TN)*: correctly classified different identity pairs; *False Positives (FP)*: different identity pairs incorrectly classified as same; *False Negatives (FN)*: same identity pairs incorrectly classified as different

To further evaluate system performance, standard classification metrics are computed [23]. The metrics are: Precision, Recall, and F1-Score, respectively.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = 2\,\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (5)$$

4. Results

The results show strong performance of the YOLO11 + ArcFace system in masked face recognition, achieving 0.9632 accuracy at an optimal threshold of 0.6245, indicating high discriminative capability.

The final recognition decision is based on a similarity threshold separating genuine and impostor pairs. In this work, the threshold is selected using the Equal Error Rate (EER) criterion on fused similarity scores, aiming to equalize and minimize FAR and FRR. This threshold is used for all verification experiments.

Table 1 reports the global performance at the optimized threshold. The model reaches an accuracy of 0.9632, with macro precision, recall, and F1 score close to 0.9630. The EER is 0.0368 (3.68%). Metrics for the “same” and “different” classes are very similar. This indicates that the system does not strongly favor one decision over the other.

Figure 4 shows FAR and FRR as functions of the decision threshold.

TABLE 1. Performance Metrics at the Optimized Similarity Threshold

Metric	Value
Accuracy	0.9632
Precision (macro)	0.9630
Recall (macro)	0.9630
F1 score (macro)	0.9629
Precision (different)	0.9601
Precision (same)	0.9659
Recall (different)	0.9662
Recall (same)	0.9598
F1 score (different)	0.9631
F1 score (same)	0.9628
ROC AUC	0.9638
EER	0.0368

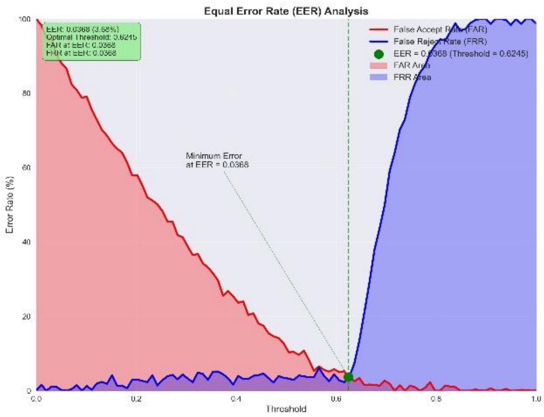

FIGURE 4. EER analysis showing FAR and FRR versus the similarity threshold. The intersection at 0.6245 yields EER = 0.0368

The curves intersect at a threshold of about 0.6245, where both FAR and FRR are 0.0368. This point defines the EER of 3.68% and is used as the operating threshold.

Selecting this threshold balances false acceptance and false rejection errors. A lower threshold reduces FRR but increases FAR, while a higher threshold lowers FAR at the expense of more false rejections. Thus, the EER threshold provides a practical trade-off for this verification task.

Table 2 compares the proposed method with prior masked face recognition studies on RMFD and RMFRD, focusing on accuracy and verification metrics. Direct comparison is limited due to differences in protocols, tasks, thresholds, and evaluation settings. Under a verification protocol on RMFD, the proposed system achieved 0.9632 accuracy using 43,304 balanced pairs, with an EER-optimized threshold producing 0.0368 EER and 0.9854 ROC AUC, indicating strong separation between genuine and impostor pairs.

TABLE 2. Comparison of Recognition Performance on RMFD and RMFRD Datasets

Study / Method	Dataset	Task Type	Performance
YOLO11 + ArcFace (Proposed)	RMFD	Verification	96.32%
CNN+BoF with VGG-16	RMFD	Ident./Verif.	91.30%
Chong et al. (HRNN: HOG + RNN)	RMFD	Verification	95.56%
MFCosface	RMFRD	Not specified	92.15%
MFR-CDC	RMFRD	Not specified	95.22%
MFR-DML + FaceMaskNet-21	RMFRD	Ident./Verif.	82.22%

In related work, Chong et al. [24] reported 95.56% accuracy on RMFD using HRNN, though differences in thresholding and evaluation protocols limit direct comparison. On RMFRD, Hariri’s CNN+BoF method achieved 91.3% using non-masked regions and VGG16 features [25]. Other reported RMFRD results include MFCosface (92.15%), MFR-CDC (95.22%), and MFR-DML with FaceMaskNet-21 (82.22%) [26]. While competitive, the proposed method emphasizes methodological transparency through explicit threshold calibration and multi-metric evaluation for reproducibility and practical validity [26].

5. Conclusions

This study proposes a two-stage masked face authentication system combining (i) YOLO11-based face detection with mask-state classification and (ii) ArcFace-based verification using similarity scoring. The detection stage achieves 99.0% face detection and 95.57% mask classification accuracy, while the verification stage reaches 0.9632 accuracy, and 0.0368 EER at an optimal threshold of ~0.6245. Results show that mask-aware detection and upper-face–focused cropping preserve strong authentication performance under occlusion. Future work will focus on adaptive thresholding, improved domain generalization, occlusion-aware representations, and deployment testing on real-time and edge devices.

Acknowledgements

This work was supported in part by the National Science and Technology Council under contract number 114-2221-E-468 -014 -.

References

[1] H. Putri, S. Hadiyoso, S. B. P. Fatoni, V. Octaviany, A. Wulandari, R. Aprilina, and M. Rosmiati, “Security system for door locks using yolo-based face recognition,” *International Journal on Informatics Visualization*, vol. 9, no. 1, pp. 224–230, 2025.

[2] R. Vidanaarachchi, etc., “Imagery, privacy and ethics: An overview of partially occluded facial biometric analysis in the era of face masks,” *Inte. Conf. on Elect., Comp. and Energy Tech. (ICECET)*, 2021, pp. 1–8.

[3] M. S. Ejaz and M. R. Islam, “Masked face recognition using convolutional neural network,” in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, 2019, pp. 1–6.

[4] M. M. Boulos, “Facial recognition and face mask detection using machine learning techniques,” Master’s Thesis, Montclair State University, 2021. https://digitalcommons.montclair. edu/etd/728

[5] O. T. Ibitoye and O. O. Osaloni, “Masked faces classification using deep convolutional neural network with vgg-16 architecture,” *3rd International Conference on Electrical Engineering and Informatics (ICon EEI)*, 2022, pp. 168–171.

[6] N. Ullah, A. Javed, M. A. Ghazanfar, A. Alsufyani, and S. Bourouis, “A novel deepmasknet model for face mask detection and masked facial recognition,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, pp. 9905–9914, 2022.

[7] C. Dewi and H. J. Christanto, “Automatic medical face mask recognition for covid-19 mitigation: Utilizing yolo v5 object detection.” *Revue d’Intelligence Artificielle*, vol. 37, pp. 627–638, 6 2023.

[8] A. Kumar, A. Kalia, K. Verma, A. Sharma, and M. Kaushal, “Scaling up face masks detection with yolo on a novel dataset,” *Optik*, vol. 239, p. 166744, 2021.

[9] A. N and K. Anusudha, “Real time face recognition system based on yolo and insightface,” *Multimedia Tools and Applications*, pp. 31893–31910, 2024.

[10] S. Khairnar, etc., “Advanced techniques for biometric authentication: Leveraging deep learning and explainable ai,” *IEEE Access*, vol. 12, pp. 153580–153595, 2024.

[11] F. H. Dahri, “Automatic face mask detection and recognition using deep learning,” *ScienceOpen Preprints*, 2022, doi: 10.14293/S21991006.1.SOR-.PP7YAGY.v1.

[12] J. R. R. Manesco and A. N. Marana, “Combining arcface and visual transformer mechanisms for biometric periocular recognition,” *IEEE Latin America Transactions*, vol. 21, pp. 814–820, 2023.

[13] B. Kocacinar, B. Tas, F. P. Akbulut, C. Catal, and D. Mishra, “A realtime cnn-based lightweight mobile masked face recognition system,” *IEEE Access*, vol. 10, pp. 63496–63507, 2022.

[14] M. S. M. Shahar and L. Mazalan, “Face identity for face mask recognition system,” *11th IEEE Symp. on Comp. Appli. & Indust. Electronics (ISCAIE)*, 2021, pp. 42–47.

[15] N. Boyko, O. Basystiuk, and N. Shakhovska, “Performance evaluation and comparison of software for face recognition, based on dlib and opencv library,” *IEEE Second Int. Conf. on Data Stream Mining & Processing (DSMP)*, 2018, pp. 478–482.

[16] G. Jocher and J. Qiu, “Ultralytics yolo11,” 2024, accessed: 2025-05-05. https://github.com/ultralytics/ultralytics

[17] H. Al-Sarrar and H. Al-Baity, “AIZOO and MMD datasets.” figshare, Nov 2022. https://doi.org/10.6084/m9.figshare. 21554835.v1

[18] D. L. de Medeiros, “Aizoo dataset computer vision dataset,” mar 2025, visited on 2025-12-14. https://universe.roboflow. com/diego-lucena-de-medeiros/aizoo-dataset

[19] A. Nabhaan, “Fmd mask dataset computer vision model,” oct 2022, visited on 2025-12-13. https://universe.roboflow.com/ahmad-nabhaan-18ivd/fmd-mask-dataset

[20] B. Roy, etc., “Moxa: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks,” *Trans. of the Indian Nat. Academy of Engi.*, vol. 5, 07 2020.

[21] F. I. Eyiokur, etc., “Unconstrained face mask and face-hand interaction datasets: building a computer vision system to help prevent the transmission of covid-19,” *Signal, Image and Video Processing*, vol. 17(4), p. 1027–1034, 2022.

[22] Z. Wang, etc., “Masked face recognition dataset and application,” *IEEE Trans. on Biometrics, Behavior, and Identity Science*, vol. 5, no. 2, pp. 298–304, 2023.

[23] J. Obi, “A comparative study of several classification metrics and their performances on data,” *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, pp. 308–314, 02 2023.

[24] W.-J. Chong, C. Chin, and T.-S. Ong, “Masked face recognition using histogram-based recurrent neural network,” *Journal of Imaging*, vol. 9, p. 38, 02 2023.

[25] W. Hariri, “Efficient masked face recognition method during the covid-19 pandemic,” 2024. https://arxiv.org/abs/ 2105.03026

[26] M. Mahmoud, M. S. Kasem, and H.-S. Kang, “A comprehensive survey of masked faces: Recognition, detection, and unmasking,” *Applied Sciences*, vol. 14, no. 19, 2024.