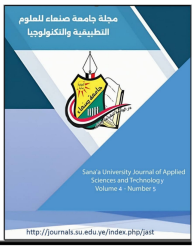
SUN PUBLISHER
مجلة جامعة صنعاء للعلوم التطبيقية والتكنولوجيا
Sana'a University Journal of Applied Sciences and Technology
https://journals.su.edu.ye/index.php/jast/



Deep Learning Approaches for Masked Face Recognition: A Survey of Research During (2020-2024)

Haron M . Ahmed Omer * **and Fadl Ba-Alwi**

Department of Computer Science, Faculty of Computer and Information Technology, Sana'a University, Sana'a, Yemen.

*Corresponding author: haron_omer@yahoo.com

ABSTRACT

The widespread adoption of face masks during the COVID-19 pandemic created new requirements for facial recognition technology, driving focused innovation in Masked Face Recognition (MFR) systems. This survey synthesizes and analyzes key research developments from 2020 to 2024, examining more than twenty pivotal studies in the field. The review provides a structured comparison of contemporary deep learning methodologies, with particular attention to Convolutional Neural Network (CNN) implementations including ResNet, VGG, and MobileNet architectures. It further evaluates the characteristics and performance outcomes associated with principal publicly available datasets employed in MFR model development.

The analysis highlights important considerations for advancing MFR systems, particularly regarding data quality and model generalization. Current research directions emphasize improving the alignment between training data characteristics and real-world application environments. Among available resources, the MaskedFace-Net dataset demonstrates particularly strong performance in benchmark evaluations, with reported recognition rates exceeding 99%, attributed to its specialized design and substantial scale. For practical implementation, the findings support utilizing established CNN-based frameworks, with opportunities for enhancement through complementary approaches such as integration with traditional classifiers.

Looking forward, this survey identifies promising pathways for continued progress in MFR technology, centered on evolutionary advancements within proven architectural paradigms, sophisticated data utilization strategies, and optimizations for deployment efficiency. This comprehensive review offers valuable guidance for researchers and engineers working to create effective and adaptable masked face recognition solutions.

ARTICLE INFO

Keywords:
Machine learning, Deep learning, Face Recognition, Masked Face Recognition, Convolutional Neural Networks.

Article History:
Received: 17-December-2025,
Revised: 14-January-2026,
Accepted: 17-March-2026,
Published: 28 May 2026.

1. INTRODUCTION

The COVID-19 pandemic necessitated widespread mask-wearing, creating unprecedented challenges for conventional facial recognition systems and accelerating research into Masked Face Recognition (MFR). This review comprehensively examines over 20 influential studies published between 2020 and 2024, providing a systematic evaluation of contemporary deep learning approaches. The analysis focuses particularly on Convolutional Neural Network (CNN) architectures, including ResNet, VGG, and MobileNet, while also assessing the characteristics and performance of major publicly available datasets used in MFR development.

The key findings highlight the importance of data quality and model generalization, with current research emphasizing better alignment between training data and real-world conditions. The MaskedFace-Net dataset demonstrated exceptional benchmark performance, with recognition rates exceeding 99%, attributable to its specialized design and substantial scale. For practical implementation, established CNN frameworks remain foundational, with potential enhancements through integration with traditional classifiers. The analysis also identifies critical challenges, including the simulation-reality gap and demographic bias, that must be addressed for equitable deployment.

Looking forward, this survey outlines promising research directions centered on architectural refinements within proven paradigms, advanced data utilization strategies, and deployment optimization. This comprehensive review offers valuable guidance for researchers and practitioners in developing effective and adaptable MFR solutions.

2. METHODOLOGY OF PAPER

2.1. Search Strategy

A systematic search of IEEE Xplore, SpringerLink, Scopus, Google Scholar, ACM, and ScienceDirect was conducted for studies published between 2020-2024 using the keywords ("masked face recognition" OR "MFR") AND ("deep learning" OR "CNN") AND ("COVID-19").

2.2. Inclusion and Exclusion Criteria

Inclusion: peer-reviewed, English, original experiments, recognition focus, deep learning methods. Exclusion: review papers, detection-only studies, insufficient detail, preprints, non-mask face recognition.

2.3. Study Selection Process

Following the PRISMA guidelines, the initial search yielded 350 records. After removing duplicates (280 remaining) and screening titles/abstracts (220 excluded), 60 full-text articles were assessed. 38 were excluded (12 no experiments, 8 insufficient detail, 15 detection focus, 3 duplicates), leaving 22 studies for synthesis.

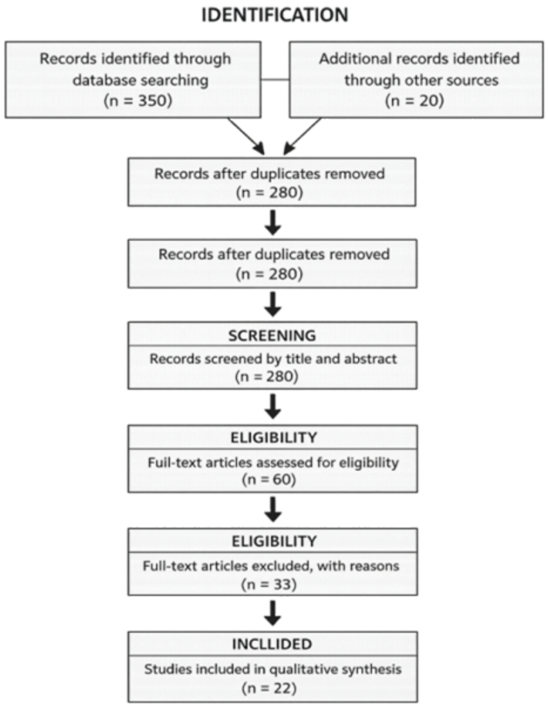


Figure 1. PRISMA flow diagram depicting the study selection process.

2.4. Data Extraction and Analysis

The extracted data included author(s), year, architecture, dataset(s), accuracy metrics, innovations, and limitations. The thematic analysis organized the studies into four categories, which are presented in Section (3).

3. DEEP LEARNING (DL)TECHNIQUES FOR MASKED FACE RECOGNITION

3.1. Introduction to Deep Learning for MFR

Deep learning, a subset of machine learning that uses neural networks, excels at identifying complex patterns in image data.

3.2. Feature Extraction and Reconstruction Methods

Research has focused on extracting robust features from visible facial regions. [1] evaluated VGG16, VGG19, and MobileNet on HMFD (97% accuracy). [2] used VGG-16 with cross-validation on multiple datasets (97.98% accuracy, 97.6 F1-score). [3] proposed a DeepMaskNet (93.33% accuracy). These methods rely heavily on periocular features, making them vulnerable to pose, occlusion, and lighting variations. Additionally, they assume that visible regions contain sufficient identity information—an assumption that may not hold for individuals with distinctive lower-face features.

3.3. HYBRID APPROACHES COMBINING DEEP FEATURES WITH CLASSICAL CLASSIFIERS

Combining CNNs with classical classifiers achieves the highest accuracy. [4] used ResNet50 with SVM, decision trees, and ensemble methods, achieving 99.64% on the RMFD and 99.48% on the SMFD. [5] integrated RPCA with KNN (97% accuracy), demonstrating that multistage preprocessing improves robustness. [6] employed metric learning with FaceMaskNet-21 (88.92% accuracy), offering open-set capabilities. Hybrid approaches decompose recognition into stages, improving interpretability and enabling the optimization of each component independently.

3.4. TEXTURE-BASED AND ENSEMBLE METHODS

[7] combined GLCM texture features with DenseNet-121 and VGG-16 ensemble, achieving 98.56% accuracy on an in-house dataset (38,290 images with proper, partial, and no-mask variations). Texture analysis captures mask-related artifacts (fabric weaves, folds, and boundaries) that can serve as discriminative cues. However, overreliance on mask-specific features risks performance degradation when mask types or placements vary.

3.5. METRIC LEARNING APPROACHES

[6] proposed FaceMaskNet-21, which produces 128-d embeddings (88.92% accuracy on LFW). Unlike classification models, metric learning enables open-set recognition by mapping faces to an embedding space where same-identity faces cluster together. This allows dynamic identity addition without retraining, which is ideal for evolving security applications. However, it requires more training data and sophisticated sampling strategies.

3.6. COMPARATIVE ANALYSIS AND DISCUSSION

Synthesizing the four methodological categories examined, hybrid methods achieved the highest accuracy (99.64%) but required more computational resources. Texture-based methods (98.56%) offer robustness at a higher computational cost. Feature extraction approaches (93-98%) balance accuracy and efficiency. Metric learning (88.92%) enables open-set scenarios with lower absolute accuracy.

For readers unfamiliar with CNNs, foundational explanations are available in [8–10].

As established in reference [8], the combined work of the convolutional and pooling layers constitutes the feature extraction phase of a CNN (see Figure 2).

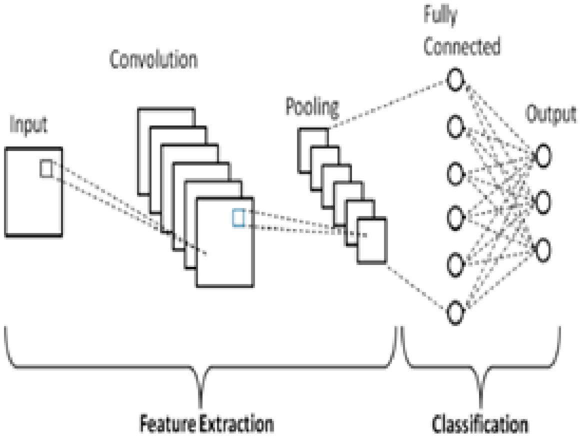


Figure 2. The general structure of Convolutional Neural Network [10]

A summary of the models examined in the preceding sections is provided in Table (2), which classifies them according to the underlying technological approach. This table highlights the predominant techniques employed in the surveyed masked face recognition literature. The analysis confirms that Convolutional Neural Networks (CNNs) are the most frequently utilized architecture for face mask detection. Their prevalence is attributed to key advantages such as spatial and translational invariance, parameter sharing, and scalability. CNNs offer a potent framework for learning and extracting discriminative visual features, rendering them highly effective for tasks including masked face recognition and general computer vision applications. Concurrently, MobileNets have gained significant traction due to their lightweight and efficient design, making them particularly suitable for deployment on mobile and embedded platforms. Their architecture allows for efficient, independent processing of sensor data, capturing distinctive features while modeling complex inter-dependencies, which contributes to enhanced performance and reduced operational latency [11, 12]. Consequently, a combination of deep learning, traditional machine learning, and hybrid sensor-based MobileNet implementations continues to be a favored strategy for developing and deploying mask recognition systems. Furthermore, Table (2) indicates that the highest reported accuracy was attained through a hybrid model that synergizes machine learning classifiers—specifically, Decision Trees and Support Vector Machines (SVMs)—with a deep learning component based on the ResNet50 CNN architecture.

To clearly delineate and identify the distinctions between these various technological approaches, Table (2) provides a description of each method, enumerates their respective advantages, and specifies their primary application domain or classification.

Table (3) also shows the recognition technique technology description. along with the advantages of each technology and the field to which it belongs.

Table[1]: Architectural Trade-offs: **The reviewed methods exhibit distinct trade-offs along several dimensions:**

Method Category	Accuracy	Computational Cost	Flexibility	Training Data Requirements	Open-Set Capability
Feature Extraction (3.2)	High	Moderate	Low	Moderate	Limited
Hybrid (3.3)	Very High	High	Moderate	Moderate	Limited
Texture-Based (3.4)	High	High	Low	High	Limited
Metric Learning (3.5)	Moderate	Moderate	High	Very High	Excellent

Table[2]: Summary of recognition techniques and performance from reviewed works.

Prototype	Techniques	Dataset	Accuracy
[4]	The model was built using ResNet50 for feature extraction, while decision trees, support vector machines (SVMs), and ensemble algorithms were used to classify face masks.	RMFD, SMFD and LFW	99.6%
[7]	The model used a pre-trained CNN DenseNet-121 and Visual Geometry Group (VGG)-16 along with GLCM for texture extraction for face detection and identification.	new dataset known as in-house dataset	98.56%
[1]	This model uses the following CNNs (VGG16, VGG19, MobileNet, and MobileNetV2) to extract texture for face detection and identification.	HSTU Masked Face Dataset (HMFD)	reaching up to 97%
[6]	This model uses the FaceMaskNet-21 deep learning network, which was designed to produce 128-d encodings using the OpenCV library, Python, and deep learning libraries. To obtain accurate and fast results on a masked face, deep metric learning was then used. This technique differs from traditional deep learning in that instead of taking a single input and classifying it, real decimal numbers of 128 d are output (128-d encodings).	Labeled Faces in the Wild (LFW) dataset	overall accuracy of 88.92%
[2]	A VGG-16 CNN model was used for image recognition, which contains 16 convolutional layers, max-pooling, activation, and fully connected (FC) layers. the model is designed to extract a region of interest (ROI) from each image in the dataset.	RMFD Masked-FaceNet Celebrities Face Recognition (CFR)	97%
[5]	robust principal component analysis (RPCA) for extract features combines deep-learning-based mask detection, landmark and oval face detection, and robust principal component analysis (RPCA)	Labeled Face in the Wild and Stimulated Masked Face Data Set (LFW-SMFD)	97%
[3]	This model uses advanced DeepMasknet for features extraction for face detection and identification.	(MDMFR) dataset	93.33%

3.7. COMPARATIVE PERFORMANCE ANALYSIS

Multidimensional analysis reveals critical insights: (1) accuracy-complexity trade-off: highest accuracy (99.64%) requires substantial resources; (2) lightweight architectures (MobileNet) offer 97% accuracy for edge deployment; (3) no studies evaluated demographic fairness—a critical gap; (4) specialized models lack cross-dataset validation; and (5) metric learning enables open-set applications despite lower accuracy.

3.8. PERSISTENT CHALLENGES ACROSS METHODOLOGICAL APPROACHES

Despite architectural diversity, fundamental challenges persist: (1) information loss from occluded lower-face features forces reliance on the periocular region; (2) simulation-reality gap: synthetic-trained models fail on real-world data; (3) demographic bias amplification from biased source datasets; (4) inadequate evaluation protocols (single-dataset focus, no standardization); and (5) computational efficiency vs. accuracy trade-offs limiting edge deployment.

3.9. SUMMARY OF KEY FINDINGS

Key findings: Hybrid approaches achieved the highest accuracy (99.64%) for closed-set applications. Lightweight architectures (97%) enable edge deployments. Metric learning (88.92%) supports open-set scenarios. Critical gaps: no demographic fairness evaluation, limited cross-dataset validation, and unaddressed simulation-reality gap. MaskedFace-Net offers benchmark leadership (>99% accuracy) but requires real-world validation.

4. DATASETS FOR DEVELOPING FACE MASK RECOGNITION ALGORITHMS

Table[3]: summarizes the Recognition Techniques Technology Description.

Field	Technique	Description	Advantages
Face Detection	Oval	Detect oval or square shaped masks for face detection.	Helps in detecting faces or objects that are partially occluded.
	Landmark	Detects major facial landmarks such as eyes, nose, and mouth.	Helps align the face to improve model accuracy.
Machine Learning	Decision Trees	A supervised learning method that partitions data into branches based on feature values.	Can handle both categorical and continuous data.
	SVM (Support Vector Machine)	A supervised classifier that finds the hyperplane that best separates data into distinct classes.	Works well with smaller datasets with clear separation margins.
Deep Learning (CNN)	VGG-16 or VGG-19	Deep convolutional neural networks with 16 layers for image classification (VGG-16) or 19 layers for image classification (VGG-19)	It performs well on image classification tasks and is widely used.
	DenseNet-121	A deep learning model with dense connections between layers, where each layer gets input from all previous ones.	Efficient use of parameters, better gradient flow and improved performance Reduces overfitting and improves feature reuse.
	Inception	A deep learning model that uses multiple filter sizes at each layer.	Captures features at various scales.
	ResNet V1	The first version of the Residual Network using skip connections to solve gradient issues.	Reduces the vanishing gradient problem.
	ResNet50	A residual network that introduces skip connections, allowing for deeper networks without vanishing gradients	Highly accurate and effective in deep networks and scalable.
	MobileNetV2	A lightweight CNN designed for mobile and embedded devices, optimized for efficiency and speed.	Reduces computational load while maintaining high performance.
Deep Learning (Mask Detection)	FaceMaskNet-21	A specialized model for detecting whether people are wearing face masks in images or video streams.	High accuracy in identifying people wearing masks in various environments.
	Deep Learning-based Mask Detection	Uses deep learning models to detect whether a person is wearing a mask or not.	High accuracy and adaptable to new mask types and facial variations.

Table[4]: Multi-dimensional comparison of reviewed approaches

Approach	Ref	Architecture	Accuracy (%)	Inference Speed	Model Size	Mask Type Robustness	Demographic Fairness	Open-Set Capability
Feature Extraction	[1]	VGG16/VGG19/ MobileNet	97.0	High	Moderate-High	Moderate	Not evaluated	Limited
Feature Extraction	[2]	VGG-16 + custom CNN	97.98	Moderate	High	Good	Not evaluated	Limited
Feature Extraction	[3]	DeepMaskNet	93.33	Moderate	High	Moderate	Not evaluated	Limited
Hybrid	[4]	ResNet50 + SVM/DT / Ensemble	99.64	Moderate	High	Excellent	Not evaluated	Limited
Hybrid	[5]	RPCA + DL + KNN	97.0	Moderate	Moderate	Good	Not evaluated	Limited
Texture-Based	[7]	DenseNet-121 + VGG-16 + GLCM	98.56	Low-Moderate	Very High	Good	Not evaluated	Limited
Metric Learning	[6]	FaceMaskNet-21	88.92	High	Low-Moderate	Moderate	Not evaluated	Excellent

4.1. The Critical Role of Data Quality Datasets

serve as the foundational pillar upon which the face mask Dataset quality, diversity, and scale fundamentally constrain model performance. This section critically examines the available datasets, categorizing them as synthetic or real-world, and analyzes their strengths, limitations, and biases.

4.2. Categorization of Masked Face Datasets

Based on their method of creation, masked face datasets can be broadly classified into two categories: synthetically generated (where masks are digitally superimposed on existing face images) and real-world captured (where images are collected from individuals physically wearing masks). Each category has distinct advantages and inherent limitations.

4.3. Synthetic Datasets: Advantages and Limitations

Synthetic datasets (MaskedFace-Net, adapted CASIA/LFW) offer scalability and control at a minimal cost. MaskedFace-Net's 67,049 masked images enabled large-scale training. However, they suffer from a simulation-reality gap: digitally superimposed masks lack physical realism (texture, fit variations, and lighting interactions). Models that excel on synthetic benchmarks (>99% accuracy) often underperform on real-world footage.

4.4. Real-World Datasets: Strengths and Challenges

Real-world datasets (RMFD, MAFA, HMFD) capture authentic mask variations (fabric types, fitting styles, and environmental interactions).

They provide ecological validity but face scale limitations (HMFD: 5,280 vs. MaskedFace-Net: 67,049), demographic imbalances, and annotation complexities.

4.5. Performance Patterns Across Dataset Types

The synthetic dataset performance was remarkably high (99.92%), but failed to transfer. The real-world accuracy was lower but more indicative (RMFD: 97.98%, MAFA: 87.30%). Cross-dataset evaluation remains rare, limiting the generalization assessment.

4.6. The Critical Issue of Demographic Bias

Foundational datasets (LFW, CASIA) exhibit demographic imbalances that are inherited and potentially amplified in masked versions. No reviewed study evaluated demographic fairness, which is a critical ethical concern for deployment.

4.7. Dataset Recommendations and Future Directions

For the benchmark performance: MaskedFace-Net. For deployment: RMFD and MAFA. For fairness, curate balanced subsets and collect targeted real-world data. Future priorities include real-world datasets with demographic annotation, generative models for high-fidelity synthetic data, and cross-dataset evaluation standards.

4.8. Summary of Dataset Analysis

MaskedFace-Net achieved >99% accuracy, which is optimal for initial development. The CASIA enables diverse masked variants. Researchers must address the simulation-reality gap and demographic biases.

In summary, dataset selection must balance benchmark performance against real-world applicability, with careful attention to inherent biases and the simulation-reality gap.

5. CONCLUSION

This survey provides a systematic and critical examination of deep learning approaches for masked face recognition, synthesizing findings from 22 key studies published between 2020 and 2024. Based on this comprehensive analysis, we draw the following evidence-based conclusions and recommendations:

5.1. Summary of Key Findings

Datasets: MaskedFace-Net achieved >99% benchmark accuracy but exhibited a simulation-reality gap. The RMFD and MAFA datasets provide real-world evaluations. Techniques: Hybrid models (ResNet50+SVM) achieved the highest accuracy (99.64%). Lightweight architectures (MobileNet) offer 97% accuracy for edge deployment.

Metric learning enables open-set scenarios (88.92%). Gaps: No demographic fairness evaluation, limited cross-dataset validation, and unaddressed simulation-reality gap.

5.2. Recommendations for Practice

Based on these findings, we offer the following recommendations for researchers and practitioners:

1- **For maximum benchmark performance**, MaskedFace-Net [17] was utilized for training and evaluation, and hybrid architectures combining deep feature extraction (ResNet50) with classical

Table[5]: Summary of benchmark datasets for masked face recognition

Dataset	No of Images	No of Masked Images	No of Un-Masked Images	Type	Advantage	Prototype	Accuracy
CASIA [13]	494,414	4,007	490,407	Synthetic (adaptable)	The CASIA dataset contains a large number of images and a number of different individuals. This large amount of data contributes to the training of a robust model that can better generalize to new data.	[13]	96%
RMFD [2]	95,000	5,000	90,000	Real-world	–	[2]	97.98%
LFW [14]	26,350	13,117	13,233	Synthetic (adaptable)	–	[14]	97%
HMFD [1]	5,280	5,280	–	Real-world	HMFD was developed, which incorporates photos of individuals who are appropriately wearing masks, ensuring that the mouth, nose, and cheeks are totally covered. The dataset includes a wide variety of perspectives, ages, and genders.	[1]	97%
COMASK20 [15]	2,754	2,754	–	Real-world	–	[15]	87%
MAFA [16]	35,806	911	30,811	Real-world	–	[16]	87.30%
MaskedFace-Net [17]	133,783	67,049	66,734	Synthetic	The MaskedFace-Net dataset is specifically designed to identify faces covered by masks. The dataset includes thousands of images covering thousands of individuals. It contains images of masked faces collected from real-life situations of people wearing masks in different locations, lighting conditions, and shooting angles. It also contains images of faces of different ages, genders, and demographic characteristics, which helps reduce bias.	[18]	99.92%
CFR [2]	2,549	–	2,549	Real-world	–	[2]	97.98%
MDMFR [3]	6,006	3,174	2,832	Synthetic	–	[3]	93.33%
In-house [7]	38,290	28,859	9,431	Real-world	an in-house dataset was developed of images of 2,500 people: 38,290 facial images, of which 17,596 have proper masks, 11,263 have partial masks (improperly worn), and 9,431 have no masks.	[7]	98.56%
ISL-UFMD [19]	21,816	10,689	10,618	Real-world	–	[19]	98.2%

classifiers (SVM, ensemble methods) were employed, as demonstrated in [4]

2- **For real-world deployment**, training and evaluation should be prioritized on real-world datasets (RMFD [2], MAFA [16] that capture authentic mask variations. Data augmentation simulating diverse mask materials and fitting conditions was incorporated to bridge the simulation-reality gap.

3- **For edge and mobile applications**, lightweight architectures (MobileNet variants [1] that balance accuracy (97%) with computational efficiency are leveraged, enabling real-time performance on resource-constrained devices.

4- **For open-set applications** requiring dynamic identity addition without retraining, consider metric learning frameworks [6] that, despite their lower accuracy, provide inherent scalability and flexibility.
5- **For robustness-critical applications**, texture-based methods [7] and RPCA-enhanced pipelines [5] demonstrate superior resilience to mask variations and occlusions.
6- **Critically, incorporate fairness evaluation** into development pipelines is critical. Supplement existing datasets with balanced demographic subsets, and report performance disaggregated by race, gender, and age group—a practice conspicuously absent from the current literature.

5.3. STUDY LIMITATIONS

Limitations: temporal scope (2020-2024), metric reliance (varying protocols), fairness data absence, potential publication bias, English-only publications.

5.4. CONTRIBUTIONS OF THIS SURVEY

Despite these limitations, this survey makes several original contributions to the field:

1- **Systematic methodology**: Application of PRISMA guidelines ensures comprehensive, reproducible literature coverage.
2- **Thematic organization**: Categorizing approaches into four methodological themes enables a structured comparison and reveals patterns obscured in chronological reviews.
3- **Multi-dimensional evaluation framework**: Extending beyond accuracy to consider computational efficiency, model size, robustness, and fairness provides a more complete picture of model capabilities and trade-offs.
4- **Critical dataset analysis**: The categorization of datasets as synthetic versus real-world, with an explicit discussion of their respective biases and limitations, informs appropriate dataset selection for different development goals.
5- **Evidence-based recommendations**: Practical guidance for researchers and practitioners based on systematic analysis rather than anecdotal evidence.
6- **Identification of critical gaps**: The documentation of the simulation-reality gap, absence of demographic fairness evaluation, and scarcity of cross-dataset validation provides a clear agenda for future research.

5.5. FINAL REMARKS

The field of masked face recognition has advanced remarkably in response to the COVID-19 pandemic, with deep learning approaches achieving impressive benchmark performances. Hybrid architectures combining deep feature extraction with classical classifiers currently represent the state of the art, whereas lightweight models enable edge deployment, and metric learning frameworks offer unique advantages for open-set applications. The MaskedFace-Net dataset provides an unparalleled resource for initial model development; however, its synthetic nature necessitates validation on real-world data.

However, the transition from laboratory demonstrations to reliable, equitable, real-world systems requires addressing the persistent challenges documented throughout this review: bridging the simulation-reality gap, systematically evaluating demographic fairness, developing standardized cross-dataset benchmarking protocols, and moving beyond accuracy-centric optimization toward more comprehensive evaluation frameworks that consider computational efficiency, robustness, and fairness as first-class citizens.

This survey provides researchers and practitioners with a structured roadmap for navigating these challenges, supporting the development of MFR systems that are accurate, robust, fair, and deployable in the diverse, uncontrolled conditions of real-world applications.

6. FUTURE RESEARCH DIRECTIONS

The rapid evolution of masked face recognition technologies opens several promising avenues for enhancing the performance and applicability of deep learning models. Future research should build upon the strong foundation of convolutional neural networks (CNNs) while addressing the persistent challenges identified in Section (3.8): information loss, the simulation-reality gap, demographic bias, and evaluation protocol inadequacies through the following directions:

Architectural innovations: adaptive receptive fields, attention modules, hierarchical fusion

- **Training paradigms**: progressive learning, meta-learning, generative augmentation, domain adaptation
- **Deployment optimization**: hardware-aware design, multi-scale processing, adaptive inference
- **Evaluation frameworks**: unified benchmarks, fairness metrics, robustness suites, and cross-dataset standards
- **Fairness and bias**: bias-aware datasets, mitigation techniques, explainability, and ethical guidelines
- **Simulation-reality bridge**: photorealistic generation, domain-invariant learning, hybrid training

Concluding Remarks

The continued advancement of masked face recognition requires a multi-faceted approach that builds upon established CNN methodologies while innovating architectural design, training strategies, deployment optimization, and evaluation frameworks. Progress demands moving

Summary of Future Research Priorities

Priority Area	Key Directions
Architectural Innovation	Adaptive receptive fields, attention modules, hierarchical fusion
Training Paradigms	Progressive learning, meta-learning, generative augmentation, domain adaptation
Deployment Optimization	Hardware-aware design, multi-scale processing, adaptive inference
Evaluation Frameworks	Unified benchmarks, fairness metrics, robustness suites, cross-dataset standards
Fairness and Bias	Bias-aware datasets, mitigation techniques, explainability, ethical guidelines
Simulation-Reality Bridge	Photorealistic generation, domain-invariant learning, hybrid training

beyond narrow accuracy optimization toward a holistic conception of system capability encompassing computational efficiency, demographic fairness, robustness to real-world variations, and adaptability to evolving conditions. These interconnected directions will enable more capable and practical systems to address the unique challenges of facial occlusion scenarios.

REFERENCES

[1] M. O. Faruque, M. R. Islam, and M. T. Islam, “Advanced masked face recognition using robust and light weight deep learning model,” *Int. J. Comput. Appl.*, vol. 186, no. 33, pp. 1–8, 2024.

[2] H. Al-Dmour, A. Tareef, A. M. Alkalbani, A. Hammouri, and B. Alrahmani, “Masked face detection and recognition system based on deep learning algorithms,” *J. Adv. Inf. Technol.*, vol. 14, no. 2, pp. 286–294, 2023.

[3] N. Ullah, A. Javed, M. A. Ghazanfar, A. Alsufyani, and S. Bourouis, “A novel deepmasknet model for face mask detection and masked facial recognition,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 10, pp. 9905–9914, 2022.

[4] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, “A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic,” *Measurement*, vol. 167, p. 108 288, 2021.

[5] M. Eman, T. M. Mahmoud, M. M. Ibrahim, and T. Abd El-Hafeez, “Innovative hybrid approach for masked face recognition using pretrained mask detection and segmentation, robust pca, and knn classifier,” *Sensors*, vol. 23, no. 15, p. 6727, 2023.

[6] R. Golwalkar and N. Mehendale, “Masked-face recognition using deep metric learning and facemasknet-21,” *Appl. Intell.*, vol. 52, pp. 13 268–13 279, 2022.

[7] O. P. Gupta, A. P. Agrawal, and O. Pal, “Masked facial recognition using ensemble convolutional neural network and grey-level co-occurrence matrix,” *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 34, no. 1, pp. 524–533, 2024.

[8] H. Ji, O. Alfarraj, and A. Tolba, “Artificial intelligence-empowered edge of vehicles: Architecture, enabling technologies, and applications,” *IEEE Access*, vol. 8, pp. 61 020–61 034, 2020.

[9] E. Mbunge, S. G. Fashoto, and S. Bimha, “A review of trends and machine learning computational models,” *Int. J. Bus. Intell. Data Min.*, vol. 20, no. 2, pp. 192–207, 2022.

[10] V. H. Phung and E. J. Rhee, “A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets,” *Appl. Sci.*, vol. 9, no. 21, p. 4500, 2019.

[11] D. Sinha and M. El-Sharkawy, “Thin mobilenet: An enhanced mobilenet architecture,” in *IEEE UEMCON*, 2019, pp. 280–285.

[12] Y. Liu, C. Miao, J. Ji, and X. Li, “Mmf: A multi-scale mobilenet based fusion method for infrared and visible image,” *Infrared Phys. & Technol.*, vol. 119, p. 103 894, 2021.

[13] S. Mishra, “A face recognition method using deep learning to identify mask and unmask objects,” M.S. thesis, University of North Dakota, 2022.

[14] M. Eman, T. M. Mahmoud, M. M. Ibrahim, and T. Abd El-Hafeez, “Innovative hybrid approach for masked face recognition using pretrained mask detection and segmentation, robust pca, and knn classifier,” *Sensors*, vol. 23, no. 15, p. 6727, 2023.

[15] H. N. Vu, M. H. Nguyen, and C. Pham, “Masked face recognition with convolutional neural networks and local binary patterns,” *Appl. Intell.*, vol. 52, pp. 5497–5512, 2022.

[16] A. Kantarcı, F. Ofli, M. Imran, and H. K. Ekenel, “Bafmd (bias-aware face mask detection dataset),” *arXiv preprint arXiv:2211.01207*, 2022.

[17] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, “Maskedface-net—a dataset of correctly/incorrectly masked face images in the context of covid-19,” *Smart Health*, vol. 19, p. 100 144, 2021.

[18] K. Hammoudi, A. Cabani, H. Benhabiles, and M. Melkemi, “Validating the correct wearing of protection mask by taking a selfie: Design of a mobile application ’checkyourmask’ to limit the spread of covid-19,” *Comput. Model. Eng. & Sci.*, vol. 124, no. 3, pp. 1049–1059, 2020.

[19] F. I. Eyiokur, H. K. Ekenel, and A. Waibel, “Unconstrained face mask and face-hand interaction datasets: Building a computer vision system to help prevent the transmission of covid-19,” *Signal, Image Video Process.*, vol. 17, pp. 1027–1034, 2023.