

ORIGINAL RESEARCH OPEN ACCESS

Development of a Face Mask Detection and Masked Facial Recognition Model Based on a Hybrid Convolutional Neural Network

Chezlyn Pillay¹ | Seena Joseph^{1,2} | Brett van Niekerk¹¹Department of Information Technology, Durban University of Technology, Durban, South Africa | ²School of Applied Computing, University of Wales Trinity Saint David, Swansea, UKCorrespondence: Seena Joseph (Seena.joseph@uwttsd.ac.uk)

Received: 13 May 2025 | Revised: 9 September 2025 | Accepted: 19 October 2025

ABSTRACT

The rapid growth of facial recognition technology has faced hindrances due to the COVID-19 pandemic, where mandatory face mask usage obscured facial features, challenging existing authentication methods. Despite the rapid development of several methods for face mask detection and recognition that highlighted prevalent issues such as poor lighting, varied angles, failed detection for improper use of face masks, computational complexity, difficulty in detecting smaller faces and low-resolution targets, these have led to suboptimal accuracy rates. Hence, this work addresses these challenges by introducing a hybrid convolutional neural network (CNN) architecture tailored for face mask detection (FMD) and masked facial recognition (MFR). The models are developed using MobileNetV2 and FaceNet InceptionResNetV1 with CNN, for FMD and MFR, respectively. Experimental results on both models utilising a total of five distinct datasets, with two for FMD and three for MFR, show the superiority of the developed model in comparison to state-of-the-art models. In addition, the models are tested in real-time for both FMD and MFR to determine their robustness, efficiency and accuracy in a real-time context. For this purpose, a 'custom real-time masked face recognition' (CRMFR) dataset was developed to perform real-time MFR. Leveraging advanced FMD and MFR technologies, the models contribute to the real-world need for enhanced security in scenarios where traditional methods are insufficient.

1 | Introduction

Facial recognition technology has witnessed transformative advancement, emerging as a reliable and widely adopted authentication method. Owing to its convenience, speed and high accuracy, facial recognition has become a preferred solution across diverse domains, including smartphone authentication, secure facility access and transaction verification [1–4]. Within the realm of facial recognition advancements, the automated processes of face mask detection (FMD) and masked face recognition (MFR) have emerged prominently. Automated FMD and MFR systems provide scalable solutions for compliance monitoring and for sustaining identification accuracy in the presence of facial

occlusions, rendering them valuable across sectors including healthcare, education, banking, and security [5, 6]. The ongoing usage of face masks, which initially surged in response to the COVID-19 pandemic, has now expanded to encompass preventive measures against infectious disease outbreaks, allergies, air pollution, crowded places and cold weather [5–10]. Deep learning has emerged as a promising avenue in FMD and MFR, enabling efficient detection of whether individuals are wearing face masks, a development accentuated by the widespread impact of the COVID-19 pandemic [11–14]. However, masks obscure critical facial landmarks, particularly the nose, mouth and jawline, which reduce the recognition accuracy of conventional facial recognition systems [15]. Factors such as mask design, camera resolution

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

and the extent of occlusion exacerbate the challenges associated with the detection and recognition of masked faces [14]. This challenge underscores the need for more resilient approaches, particularly hybrid and transfer learning methods designed to optimise performance under conditions of occlusion [5, 6].

Recent studies [16–19] have explored hybrid and optimisation-driven strategies to mitigate occlusion challenges in FMD and MFR. Approaches such as Robust Principal Component Analysis (RPCA) combined with K-Nearest Neighbour classifiers have demonstrated strong accuracy under mask-induced occlusions, though at the cost of low processing speeds unsuitable for real-time applications [17]. Similarly, RPCA enhanced with metaheuristic algorithms like the Grasshopper Optimisation Algorithm (GOA) has improved feature selection and classifier performance, while optimisation-based strategies such as Grey Wolf Optimisation (GWO) have refined discriminative features for recognition tasks [16, 19]. More advanced pipelines, including the HiMFR framework, integrate mask detection and inpainting with Vision Transformer based recognition, achieving competitive results on public datasets but facing challenges with inaccurate reconstructions and non-real-time performance [18]. While these methods highlight promising directions, they remain constrained by computational inefficiency, limited robustness under diverse real-world conditions such as poor lighting, varied mask types, partial occlusions and the absence of unified systems addressing both FMD and MFR.

Convolutional neural networks (CNN), on the other hand, have emerged as the mainstay of modern computer vision due to their exceptional scalability and robustness in detection and recognition tasks [14, 20–22]. Widely adopted models such as VGGFACE2 [23], MobileNet [24–26], DeepMaskNet [12], FaceNet [24, 27] and EfficientNetB7 [28] have shown prominent results across diverse benchmarks. In particular CNN based architectures such as the multi-task cascaded convolutional neural network (MTCNN) [29, 30] have further advanced performance under challenging conditions, including occlusion and variation in pose and lighting.

Building on these advances, this study proposes a hybrid CNN based framework that leverages MobileNetV2 for FMD and FaceNet InceptionResNetV1 with MTCNN for MFR. The proposed models are designed to deliver high accuracy and resilience in real-time settings.

Hence, the study presents the following contributions:

- Developed a robust model based on CNN architecture with MobileNetV2 for FMD, and FaceNet InceptionResNetV1 combined with MTCNN for MFR, enabling accurate detection and recognition of both masked and unmasked faces.
- Designed and implemented a robust model for masked detection and recognition under challenging conditions, including adverse lighting, varied facial angles, smaller or low-resolution faces, occlusions of specific regions, distorted or blurry images and partial obstructions (e.g., eyeglasses, hair and headscarves).
- Addressed improper and diverse mask usages, such as surgical and fabric masks worn incorrectly or inconsistently, ensuring

reliable identification despite these additional obstructions and distortions.

- The development of the ‘custom real-time masked face recognition’ (CRMFR) dataset for conducting real-time testing of MFR in real-world scenarios.

An outline of the subsequent sections is presented below.

Section 2 reviews related work on technologies for FMD and MFR models. Section 3 highlights the methods, algorithms and materials. Section 4 presents the experimental setup, results and analysis. In Section 5, a discussion based on the analysis of the results and key findings is detailed. Section 6 provides the conclusion, while Section 7 outlines limitations and future research directions.

2 | Related Work

Several studies have addressed the challenges of FMD and MFR, particularly after the COVID-19 pandemic introduced masks as obstacles for traditional recognition systems. Mundial et al. [23] combined CNN based feature extraction with an SVM classifier, achieving 97% accuracy on VGGFACE2 and a custom masked dataset, though performance declined with varied angled faces. Talahua et al. [24] employed MobileNetV2 and FaceNet with a multi-layer perceptron, reaching over 99% accuracy on the ‘Real-world masked face’ dataset but struggling with skin tone variations and dim lighting conditions. Huang et al. [31] evaluated ResNet architectures (18–100 layers) on a private dataset, achieving 87.41% accuracy, though their models were highly sensitive to occlusions and illumination. Hariri [32] tested several deep learning models, with VGG-16 performing best at 91.30%, but requiring extensive resources and datasets.

Other notable approaches include Ullah et al. [12], whose CNN-based DeepMaskNet achieved 100% in FMD and 93.33% MFR, albeit with high computational costs and a lack of dataset diversity. Islam et al. [33] exhibited that CNNs outperform traditional Local Binary Pattern Histogram (LBPH), EigenFace and FisherFace methods but still struggle with pose variations. Kaur et al. [34], who achieved 99.15% detection accuracy using CNNs, however, reported limitations under lighting changes. Boulos [35] obtained a 97.10% F1-score using Kaggle datasets, although resolution, lighting conditions and MFR were presented as limitations within this study. Ensemble approaches, such as Dharanesh and Rattani [8] with ResNet-50 and SVM, demonstrated robust results but required extensive data and still faltered on angled faces. Lightweight and real-time models have also been explored. Mandal et al. [36] achieved 89% recognition with ResNet-50, and Arya and Tiwari [37] reported 90% using LBPH and Haar cascades, while YOLO-based methods by Mhadgut [38], Suhaimin et al. [39], Abbasi et al. and Aswal et al. [7, 40] showed strong detection but limited recognition capabilities. Raspberry Pi implementations, such as Winahyu et al. [41], achieved practical results but were vulnerable to spoofing, while Yang et al. [42] adapted YOLOv5 for detection without recognition functionality. More traditional VGGFace-based approaches by Damer et al. and Neto et al. [43, 44] reported only 84%–85% accuracy, highlighting persistent challenges with masked identities and generalisation.

TABLE 1 | Related studies.

Study	Methodology	Dataset	FMD/MFR accuracy	Key limitations
[24]	MobileNetV2 + FaceNet + MLP	Real-world masked face dataset	MFR - 99.65%	Sensitive to skin tone, lighting and angles.
[12]	DeepMaskNet (CNN)	Custom mask detection and masked face recognition	FMD- 100% MFR - 93.33%	High training time, large dataset, requires a more diverse and quantified dataset, pose variation challenges and adverse lighting conditions.
[17]	SSD-MobileNetV2 + RPCA + KNN	Custom dataset	MFR - 98%	Limited dataset size, risk of overfitting, scalability issues with KNN.
[19]	Robust PCA + GOA Optimiser	Custom dataset	MFR - 97%	Computationally expensive, slow convergence, not real-time friendly.
[27]	Pareto-optimised FaceNet model with data preprocessing	Not specified	MFR - 94%	Limited dataset diversity—may not generalise to all real-world scenarios.
[21]	Lightweight CNN with batch normalisation, dropout and depth-wise normalisation	HSTU masked face dataset (HMGD)	MFR-97%	Requires frontal/lateral mask variations; tested on HMGD only.

Recent advancements in FMD and MFR have led to several notable studies. Akingbesote et al. [27] evaluated a Pareto-optimised FaceNet model with data preprocessing techniques, achieving over 94% accuracy on both masked and unmasked faces. However, the model's performance was limited by dataset diversity and computational efficiency. Zhang et al. [45] addressed large-scale MFR challenges, focusing on super-large scale training, data noise handling and masked/non-masked face recognition accuracy balancing. Their approach improved recognition accuracy but required extensive computational resources. Mahmoud et al. [13] proposed a comprehensive survey on MFR, discussing challenges such as dataset scarcity, occlusion complexity and real-time performance. While not presenting a specific model, the paper highlights critical areas for future research. Abdelwhab [28] compared pre-trained architectures such as VGG16, MobileNet and EfficientNetB7 using a structural similarity-based feature extractor, achieving up to 98% accuracy; however, this high performance is largely dependent on the use of controlled and high quality datasets, with significant risks of performance degradation in real-world deployments where variations in illumination, mask positioning and background noise are more prevalent. Jaiswal et al. [46] conducted a large-scale audit of commercial and open-source face recognition systems, revealing significant biases and low accuracy (down to 0%) in identifying masked faces, particularly among non-white individuals. This underscores the need for more inclusive, diverse and robust MFR models.

Table 1 compares recent FMD and MFR studies, outlining their methods, datasets, accuracy and key limitations. This highlights performance trade-offs and practical challenges across approaches.

Despite substantial progress in FMD and MFR research, many existing models struggle to deliver consistent performance under real-world conditions. These include poor or variable lighting,

diverse facial angles, occlusions caused by mask positioning or accessories, and the need for efficient real-time processing. Moreover, several prior approaches rely heavily on high-performance hardware, limiting their practicality in resource-constrained environments. The motivation behind this work is to develop a robust, efficient and scalable solution that simultaneously addresses these challenges in handling lighting variations, facial orientation diversity and occlusions with high accuracy and low computational demand. By optimising convolutional architectures and employing targeted hyperparameter tuning, this approach can therefore balance superior accuracy with real-time capability on standard hardware, making it accessible for widespread deployment. This comprehensive solution overcomes key robustness and generalisation limitations observed in the earlier studies, bridging the gap between theoretical advances and practical application. Consequently, this work therefore introduces a novel framework that advances the state of the art in FMD and MFR while emphasising efficiency and adaptability essential for real-world implementation.

3 | Methods and Materials

In this study both the FMD and MFR models employ variations of the Convolutional Neural Network (CNN) architecture with MobileNetV2 for FMD and FaceNet InceptionResNetV1 with the multi-task cascaded convolutional neural network (MTCNN) for MFR.

3.1 | Face Mask Detection

The proposed FMD model is based on the MobileNetV2 architecture, selected for its exceptional balance between computational efficiency, lightweight design and high classification accuracy [11, 47]. Compared to conventional CNNs such as VGGNet or ResNet,

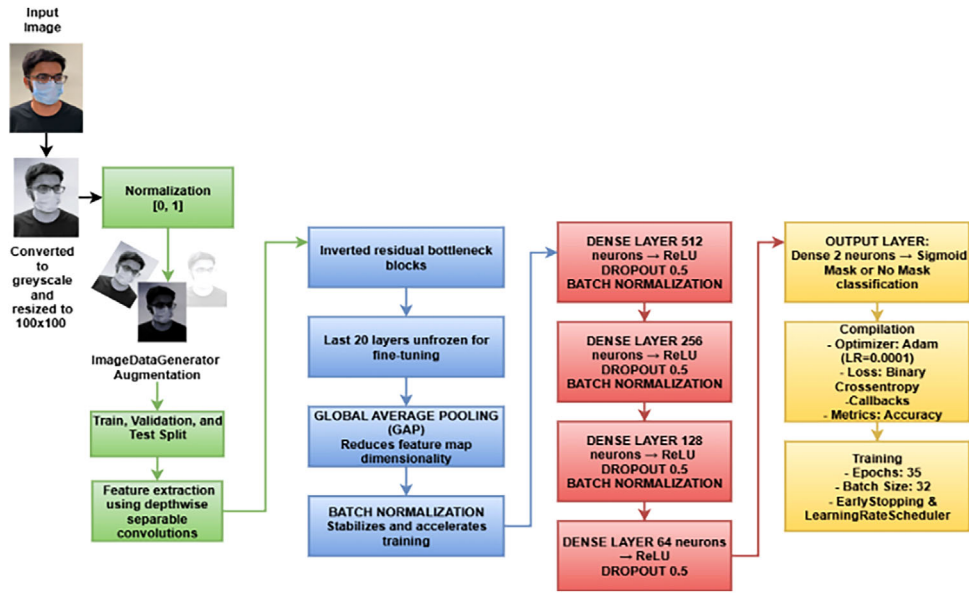


FIGURE 1 | FMD model architecture diagram.

MobileNetV2 achieves similar accuracy with a fraction of the parameters and memory footprint, making it highly deployable on embedded or resource-constrained devices. Its efficiency stems from depth-wise separable convolutions, which decompose standard convolutions into depthwise and pointwise operations, reducing computation by nearly threefold without significant loss in feature quality [48]. Furthermore, inverted residual bottleneck blocks improve gradient flow and enhance feature reuse, leading to faster convergence and better parameter efficiency [48, 49]. Using Equation 1, the MobileNetV2 architecture can be expressed.

$$\text{Output} = \text{DepthwiseConv}(\text{PointwiseConv}(\text{ReLU6}(\text{BatchNorm}(\text{Input})))) \dots \quad (1)$$

The proposed FMD model demonstrates superior performance due to its hybrid architecture that leverages MobileNetV2's lightweight yet powerful feature extraction capabilities, optimised for low-latency and real-time detection. Unlike traditional CNNs, MobileNetV2's inverted residuals and linear bottlenecks enable efficient gradient flow while reducing computational cost, resulting in a faster inference speed without compromising accuracy [50]. The model was fine-tuned with domain-specific masked and unmasked datasets, enhancing its ability to generalise across varied lighting, angles and mask types. Transfer learning significantly reduces training time while improving convergence stability. Additionally, targeted data augmentation strategies such as brightness shifts and occlusion simulation enhance robustness under real-world variability, making the model particularly effective for surveillance and access-control contexts [48, 49].

Architecturally, the FMD model, as depicted in Figure 1, employs MobileNetV2 as a pre-trained base, excluding its top classification layers. A global average pooling layer aggregates spatial information into a compact representation, followed by fully connected dense layers of 512, 256, 128 and 64 units (ReLU activation) to capture high-level discriminative patterns. Batch

normalisation and 50% dropout layers are strategically applied to stabilise training and prevent overfitting, particularly under limited sample conditions. The binary classification head consists of two neurones with sigmoid activation for mask or no-mask detection. To adapt the pretrained features to the target domain, the final 20 layers of MobileNetV2 are unfrozen for fine-tuning. Data augmentation using ImageDataGenerator includes rotation, translation, flipping and brightness variation to simulate real-world scenarios. The model was trained using the Adam optimiser (learning rate = 0.0001) with binary cross-entropy loss. Its lightweight architecture, combined with high robustness to environmental variability, makes the proposed FMD model a superior choice compared to conventional solutions.

3.2 | Masked Face Recognition

The proposed MFR model addresses the limitations of conventional face recognition systems when faces are partially occluded by masks, a scenario in which traditional methods such as plain FaceNet or OpenFace exhibit marked performance degradation [51, 52]. Navigating the complex challenges posed by MFR requires a robust, multi-component architecture. The proposed model integrates several complementary architectures to form a comprehensive solution. Knowledge of each component's design is critical for understanding the model's development. The MTCNN serves as the foundation for detecting, cropping and localising facial landmarks, as depicted in Figure 2. After detection, key landmark points are drawn on the located face using MTCNN and OpenCV to support feature extraction. FaceNet with an InceptionResNetV1 backbone then extracts highly discriminative 128-dimensional embeddings from the aligned facial regions, providing robustness to occlusions. Similarity is determined via the Euclidean distance metric, facilitating accurate face matching even when masks obscure key features [29]. A custom CNN classifier refines these embeddings to enhance discrimination between visually similar masked faces. This hybrid approach achieves superior recognition performance while maintaining



FIGURE 2 | MTCNN three step process.

computational efficiency, enabling deployment on standard laptops and real-time surveillance systems without requiring GPU acceleration.

The CNN classifier for MFR, as depicted in Figure 3, is designed to complement the embedding-based representation. It begins with a one-dimensional convolutional layer comprising 128 filters with a kernel size of three, followed by max-pooling for dimensionality reduction. A second convolutional layer with 256 filters deepens the extracted feature representation, after which the feature maps are flattened and passed through a dense layer of 512 neurones with ReLU activation to capture high-level non-linear patterns. Dropout at a rate of 0.5 mitigates overfitting, while batch normalisation stabilises training and accelerates convergence. The output layer employs SoftMax activation to classify each input into a predefined identity class. Training leverages the Adam optimiser with categorical cross-entropy loss, supported by early stopping, learning rate scheduling and checkpointing for efficiency. By combining precise alignment via MTCNN, robust embeddings from FaceNet-InceptionResNetV1 and refined classification through the CNN, the proposed model provides a high-performance, resource-efficient solution. This integrated architecture surpasses more computationally intensive frameworks, making it particularly well-suited for accurate recognition of masked faces under real-world, resource-constrained conditions.

3.3 | FMD and MFR Hyper Parameterisation

Table 2 summarises the key hyperparameters used in training both the models, detailing the configurations that contributed to the reported optimal performance.

3.4 | FMD and MFR Pseudocode

The following pseudocode as depicted in Table 3 outlines the sequential steps involved in training and deploying the proposed FMD and MFR models, providing a clear representation of the overall workflow (See Table 4 and 5).

3.5 | Materials

This study draws upon curated datasets and established performance measures to evaluate the proposed models. Data preparation included structured preprocessing and augmentation to enhance robustness under varied conditions. Model performance

was examined using precision, recall, f1-score and accuracy, ensuring a balanced and comprehensive assessment.

3.5.1 | FMD Dataset Detail

The experimental analysis for FMD employed two publicly available datasets, namely 'FMD' and 'FMD 12K Images', sourced from the Kaggle repository. The datasets utilised in this study encompass a wide range of mask types, participant demographics and imaging conditions. They include variations in gender, skin tone, facial angles, lighting intensity and image quality that range from high definition to heavily blurred.

Furthermore, the datasets incorporate a wide range of cropped facial regions and encompass participants representing a broad range of diverse ethnicities, including Caucasian, African, Asian and individuals of mixed heritage. This diversity is intended to support the development of a robust and generalisable model.

This research received ethical approval under Category One, indicating exemption from formal review by the Ethics and Biosafety Research Committee. The exemption was granted due to the exclusive use of pre-existing datasets and the minimal risk posed to human participants, eliminating the need for further ethical clearance.

With the excessive volume of images that exist for the original aforementioned FMD datasets, a more streamlined selection was extracted and deemed appropriate for this study, as the vast quantity exceeded the study's requirements. Hence, a smaller sample size is selected for both. In addition, the datasets include images where there are object obstructions present (i.e., eyeglasses, baseball cap and hair covering slight regions of the face).

3.5.2 | MFR Dataset Detail

The experimental evaluation utilised three primary sources of data: the 'MFR' dataset labelled as the 'MFR dataset', obtained from the Kaggle repository on image super-resolution; the 'mask detection and MFR' dataset labelled as the 'MDMFR Dataset', curated by Ullah et al. [12]; and a 'CRMFR' dataset labelled as the 'CRMFR Dataset', which was generated through live image capture using an HP ProBook 450 G4 (x64-based PC) laptop webcam and an Apple iPhone 11.

Considering the scarcity of publicly accessible datasets, both the 'MFR' and 'mask detection and MFR' datasets are selected due to their availability to the public and the suitability of their sample size required for the proposed study. The 'mask detection and MFR' is tailored to provide a more customised selection due to the large volume of images, ensuring a focused and relevant sample size for the analysis. In addition, the datasets include images where there are object obstructions present (i.e., eyeglasses, head scarves and hair covering slight regions of the face).

The custom 'CRTMFR' dataset was collected in different environments to complement the public datasets and address potential biases. Data collection took place in multiple locations, includ-

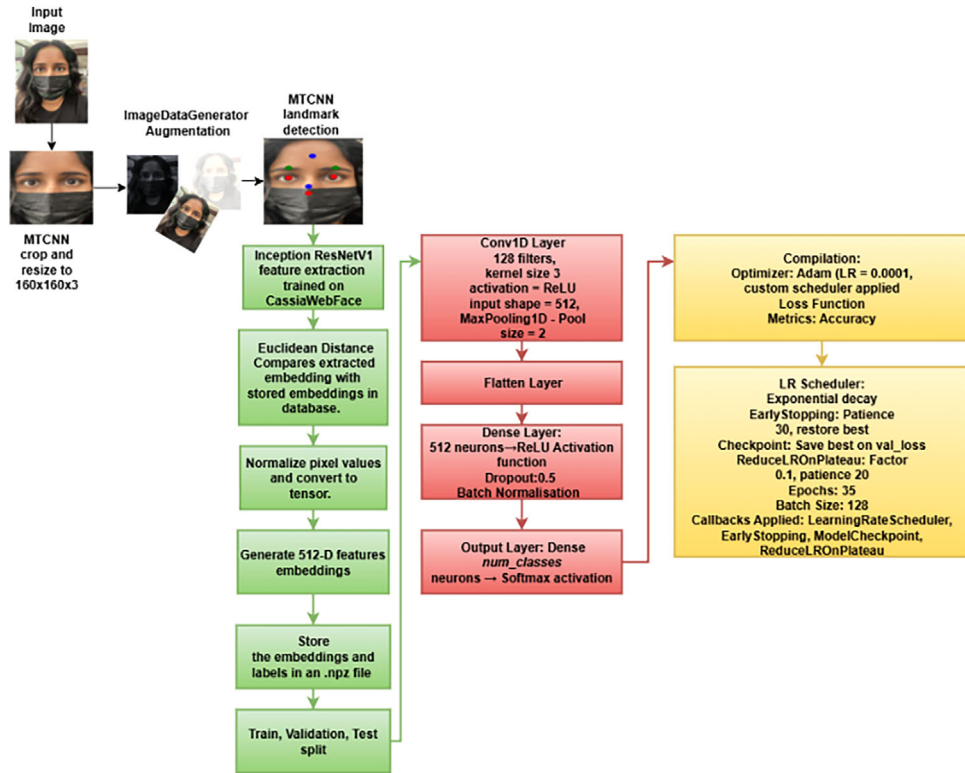


FIGURE 3 | MFR model architecture diagram.

TABLE 2 | FMD and MFR hyper parameterisation.

Hyperparameters	FMD model	MFR model
Optimiser	Adam	Adam
Initial learning rate	0.0001	0.0001
Batch size	32	128
Epochs	35	35
Loss function	Binary cross-entropy	Categorical cross-entropy
Dropout rate	0.5 (applied after dense layers)	0.5 (applied after dense layers)
Batch normalisation	Applied after dense layers.	Applied after dense layers.
Learning rate scheduler	Yes, decay by factor 0.95 per epoch (starting 0.0001).	Yes, exponential decay (starting 0.0001).
Early stopping	Yes, patience = 10, restore best weights.	Yes, patience = 10, restore best weights.
Fine-tuning	Last 20 layers of base MobileNetV2 unfrozen.	Not applied (CNN trained from scratch).
Data augmentation	Rotation, shift, shear, zoom, horizontal flip and brightness range. Rotation $\pm 40^\circ$, shift (20%), shear 0.2, zoom 20%, horizontal flip, fill mode = nearest.	Rotation, shift, shear, zoom, brightness range and channel shift. Rotation $\pm 20^\circ$, shift (10%–20%), shear 0.3, zoom 30%, brightness (0.4–1.0), channel shift 0.2, horizontal flip, 21 augmented samples/class.
Reduce LR on plateau	Not used (scheduler instead).	Yes, factor 0.1, patience = 10.
Hyperparameter	MobileNetV2	Custom CNN classifier.

ing indoor office settings, outdoor areas and varied lighting conditions to simulate real-world scenarios. The images were captured using a smartphone camera and a standard webcam setup (as detailed in Subsection 3.5.2 and Section 4), ensuring a range of hardware input quality. Demographically, this custom

dataset encompasses diverse age groups (17–85 years), gender representation (approximately balanced male and female participants) and ethnic backgrounds, including but not limited to Asian and mixed-race participants. Special care was taken to include subjects with varying skin tones and head coverings

TABLE 3 | FMD and MFR pseudocode.

FMD	MFR
Input: Image dataset D_{fmd} with masked and unmasked participants (classes) 1. Load dataset D_{fmd} and create label mapping. 2. For each image in each category: a. Read image in colour (RGB). b. Convert to grayscale. c. Resize image to 100×100. d. Append image to data list and corresponding label to target list. 3. Convert data to NumPy array and normalise to $[0, 1]$. 4. One-hot encode target labels. 5. Split dataset: 80% training, 10% validation, 10% testing. 6. Define MobileNetV2 model: a. Load pretrained MobileNetV2 (ImageNet weights, top layers removed). b. Add custom classification head: • Global Average Pooling • BatchNorm $\rightarrow$ Dense (512) $\rightarrow$ ReLU $\rightarrow$ Dropout $\rightarrow$ BatchNorm • Dense (256) $\rightarrow$ ReLU $\rightarrow$ Dropout $\rightarrow$ BatchNorm • Dense (128) $\rightarrow$ ReLU $\rightarrow$ Dropout $\rightarrow$ BatchNorm • Dense (64) $\rightarrow$ ReLU $\rightarrow$ Dropout • Output Dense (2) with Sigmoid activation c. Freeze all layers except last 20 of base model. 7. Compile model with Adam optimiser ($lr = 0.0001$) and binary cross-entropy loss. 8. Augment training data (rotation, width/height shift, shear, zoom, horizontal flip, brightness range). 9. Train model with early stopping and learning rate scheduler for up to 35 epochs. 10. Evaluate model on test set (accuracy, loss). 11. Save train/validation/test datasets and trained model for deployment. Output: Trained FMD model for masked and unmasked detection.	Input: Image dataset D_{mfr} with multiple identities (classes) 1. Load dataset from folder. 2. For each image: a. Read image in colour (RGB). b. Detect face using MTCNN; crop the detected face. c. Detect facial landmarks on the cropped face. d. Preprocess face (resize to 160×160, normalise to $[-1,1]$). e. Extract 512-d feature embedding using pretrained FaceNet (InceptionResnetV1). f. Augment face image (rotation, shift, shear, zoom, flip, brightness range) up to N times. g. Extract embeddings for augmented images. h. Store all embeddings and corresponding labels. 3. Convert embeddings and labels to NumPy arrays. 4. Encode labels numerically and one-hot encode. 5. Reshape embeddings to (512, 1) for CNN input. 6. Split dataset: 80% training, 10% validation, 10% testing. 7. Define CNN classifier: a. Conv1D (128) $\rightarrow$ ReLU $\rightarrow$ MaxPooling1D b. Conv1D (256) $\rightarrow$ ReLU $\rightarrow$ MaxPooling1D c. Flatten $\rightarrow$ Dense (512) $\rightarrow$ ReLU $\rightarrow$ Dropout $\rightarrow$ BatchNorm d. Output Dense(num_classes) $\rightarrow$ Softmax 8. Compile model with Adam optimiser and categorical cross-entropy loss. 9. Train model with early stopping, learning rate scheduler, model checkpoint, ReduceLROnPlateau for up to 35 epochs. 10. Evaluate model on test set; record accuracy and loss. 11. Save final trained model for deployment. Output: Trained MFR model for masked face recognition

TABLE 4 | FMD dataset details.

Dataset Name	No of images	Train	Validation	Test	Classes
'Face mask detection' [53]	5980	4784	598	598	2 (i.e., Mask/ No Mask)
'Face Mask Detection 12K images' [54]	5000	4000	500	500	2 (i.e., Mask/ No Mask)

(e.g., glasses) to increase model generalisability. Bias mitigation efforts included random likes of sampling during image selection, oversampling under-represented classes during augmentation and ensuring diverse lighting conditions (daylight, dim indoor lighting and artificial lighting).

3.6 | Evaluation of Model Performance

In evaluating model performance, key metrics including, accuracy, precision, recall and the f1-score, are essential benchmarks to accurately assess model performance and provide a

TABLE 5 | MFR dataset details.

Dataset name	No. of images	With augmentation	Train	Validation	Test	Classes
'Masked face recognition' [55]	465	10032	8026	1003	1003	19
'Mask detection and masked face recognition' [12]	500	11000	8800	1100	1100	50
'Custom real-time masked face recognition'	250	5500	4400	550	550	10

comprehensive assessment of the model's ability to correctly identify and classify instances when presented to it [11, 12, 24, 39, 56].

- I. **Accuracy:** Within FMD, the accuracy is utilised to measure the proportion of correctly classified instances of mask-wearing and non-mask-wearing individuals out of the total number of instances. In MFR the accuracy is used to evaluate how well the model identifies or verifies the identities of individuals despite the individual wearing a face mask [11, 12, 24, 39, 56].
- II. **Precision:** In FMD the precision is used to assess how many of the detected mask-wearers are truly wearing masks. Within MFR, precision determines how many of the identified faces were correctly matched to their specific identities [11, 12, 24, 39, 56].
- III. **Recall:** The FMD recall measures how well the model identifies all true mask-wearers from the total number of mask-wearing individuals. For the MFR, recall is used to evaluate how many of the actual identities of masked individuals were correctly recognised by the model [11, 12, 24, 39, 56].
- IV. **F1 Score:** The F1-score in FMD is utilised to balance precision and recall, providing a single metric reflecting the model's overall performance in detecting masks. In the case of MFR, the F1-score combines precision and recall to gauge how effectively the model identifies individuals when masked [11, 12, 24, 39, 56].

4 | Experimental Results and Analysis

The experimental setup for the proposed models is done on an HP ProBook 450 G4 laptop (x64-based PC) with 16 GB RAM and 232 GB of storage space, utilising both the Intel R high-definition graphics 620 integrated graphics processing unit and central processing unit. The models were developed and built using the Anaconda Integrated Development Environment (IDE) [57] and Jupyter notebook open-source web application, ensuring no additional software licensing costs [58]. The hardware comprised a mid-range, commercially available laptop valued at approximately R 15 000 in the South African market, ensuring accessibility for academic and small-scale deployments. The absence of specialised high-performance GPUs demonstrates the computational efficiency of the proposed models, with all dataset processing and model training executed locally, thereby avoiding cloud-computing costs.

TABLE 6 | 'FMD dataset' performance comparison.

Study	Acc	Prec	Rec	F1-s
[60]	99.97	99.97	99.97	99.97
[25]	99.98	99.96	99.97	99.97
[61]	99.00	98.00	99.00	99.00
[62]	96.70	96.30	92.50	94.36
[63]	98.00	98.00	98.00	98.00
Proposed study	100	100	100	100

TABLE 7 | 'FMD 12K dataset' performance comparison.

Study	Acc	Prec	Rec	F1-s
[64]	99.02	99.00	99.00	99.00
[65]	98.70	98.00	98.00	98.00
[66]	99.46	99.73	99.19	99.46
[67]	95.00	98.00	95.00	96.50
Proposed study	100	100	100	100

TABLE 8 | MobileNetV2 model performance comparison.

Study	Acc	Prec	Rec	F1-s
[24]	99.65	100	100	100
[25]	99.98	99.96	99.97	99.97
[26]	100	99.90	99.90	99.90
Proposed study	100	100	100	100

The performance of the proposed models is evaluated using the 'FMD Dataset', 'FMD 12K images dataset', 'MFR dataset', 'MDMFR dataset' and the 'CRMFR dataset'. All five datasets are split into 80/10/10 ratio for training and testing according to Pareto principal rule, which states as a general principle, 80 percent of effects are caused by 20 percent of causes [59].

The experimental results, based on accuracy, precision, recall and F1-score for both the FMD and MFR models, are depicted in Tables 6–10. Within this study, Tables 6, 7 and 9 provide a comparative analysis of models trained on the same datasets

TABLE 9 | ‘MDMFR dataset’ performance comparison.

Study	Acc	Prec	Rec	F1-s
‘MDMFR dataset’ [12]	93.33	93.00	94.50	93.74
Proposed study	99.82	99.82	99.82	99.82

TABLE 10 | Performance of the proposed MFR models.

Study	Acc	Prec	Rec	F1-s
‘MFR dataset’ [55]	99.40	99.41	99.40	99.40
‘CRMFR dataset’	99.82	99.82	99.82	99.82

utilised for the FMD and MFR models developed for the purpose of this study. Hence, all models presented in Tables 6, 7 and 9 were evaluated using the same datasets, ensuring consistency in comparison. The reported results reflect both the performance of the authors’ models and the proposed model, allowing for a direct and fair evaluation based on the datasets indicated at the top of each table. Table 8 demonstrates the comparison of different FMD models that utilised MobileNetV2 architecture. Furthermore, the ‘CRMFR Dataset’ is developed and tested for the purpose of assessing MFR in a real-time context to determine the model’s effectiveness and accuracy in real-time scenarios.

As per Table 6, the proposed model achieved 100% performance in all evaluation metrics. The second highest model achieved in study [25] attained a score of 99.98% for the accuracy. The model fell short in its ability to only detect frontal face images. In addition, it failed to detect the incorrect use of a face mask and was only able to detect certain instances when a user had ‘no mask’ on and was unable to detect the incorrect placement of a mask, labelling it as ‘mask on’. In study [60], the model achieved a performance of 99.97% across all metrics. Despite this accuracy, the model was only able to detect frontal face images and could not detect the incorrect use of a face mask. The model developed in study [61] achieved a performance with a score of 99% across the aforementioned metrics. The model struggled since its responsiveness was highly influenced by the spatial orientation of the camera. Hence, certain angles and certain lighting conditions were not picked up. Within study [63], the model only achieved a performance of 98% within the metrics, and utilised a selection of images from the ‘FMD Dataset’ to form a custom dataset. The model could not detect the incorrect use of a face mask. Therefore, it produced a considerable number of false positives in this scenario. In the study by Guo et al. [62], the model achieved the lowest score with a performance of 96.70% for accuracy, and a number of limitations were identified within this model. The model was unable to detect obscured and half-face images. In addition, the model produced a considerable number of false positives and noted computational complexity throughout the training and testing process. Figure 4 provides a visual comparison of the results based on the aforementioned evaluation metrics.

As presented in Table 7 above, the proposed model achieved a performance of 100% across the aforementioned evaluation

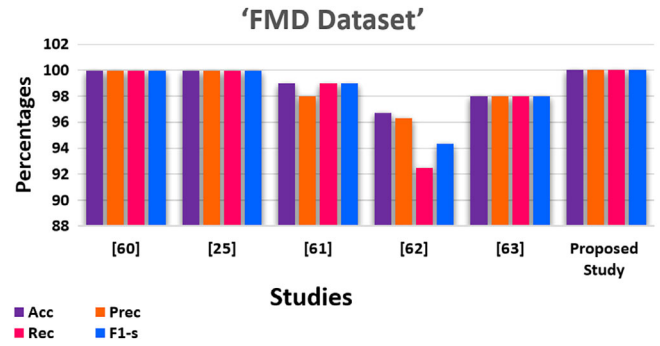


FIGURE 4 | ‘FMD dataset’ performance comparison.

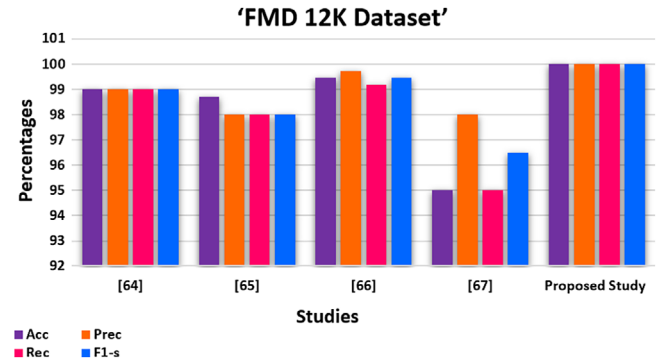


FIGURE 5 | ‘FMD 12K dataset’ performance comparison.

metrics. The model developed by Chakma et al. [66] achieved the second highest performance with a score of 99.46% in accuracy. Despite this accuracy rate, the model fell short in its ability to detect masked and unmasked faces at an angle and could therefore only detect frontal face images. Within study [64], the model achieved a respectable performance of 99.02% for accuracy. However, the model was unable to detect face masks in real-time. In addition, the model identified a considerable number of false negatives and positives in the confusion matrix and experienced a significant amount of computational complexity during the model’s training and testing process. The model developed in study [65] displayed a considerable number of false negatives and positives in the confusion matrix, thus attaining a score of 98.70% for accuracy. Furthermore, the model identified significant computational complexity. The lowest model developed by Arora et al. [67] attained a score of 95% in accuracy and exhibited a notable degree of computational complexity during the model’s training and testing. Based on Figure 5, a graphical comparison of the results based on the evaluation metrics is provided.

As indicated in Table 8, the proposed model achieved a combined mean score of 100% within the evaluation metrics. The model presented by Ilyas and Ahmad [26], trained on a custom dataset that is composed of random public repositories, achieved an accuracy of 100%; however, it underperformed across the other evaluation metrics when compared with the proposed approach. In addition, the model focused solely on FMD and did not take MFR into consideration. The model developed by Talahua et al. [24], trained on a custom dataset, produced an accuracy score of 99.65%, thus attaining the lowest score amongst the other aforementioned models. The model exhibited difficulty detecting masked faces

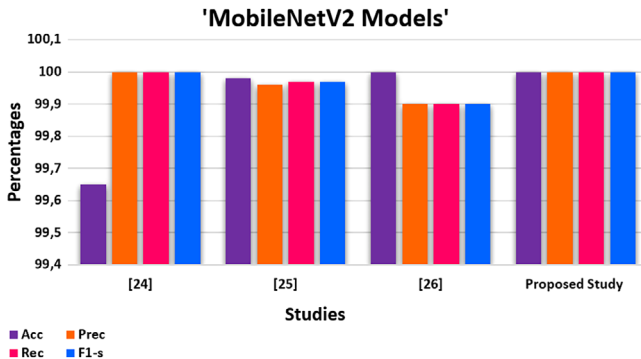


FIGURE 6 | 'MobileNetV2 model' performance comparison.

and performed poorly when faces were presented at varied angles. In addition, another limitation identified was decreased accuracy when the model was presented with adverse lighting conditions. Furthermore, this model noted computational complexity during the training and testing phases of the model. Within the study by Habib et al. [25], trained on the 'FMD by Omkar Gurav', 'FM by Oumina', 'Makhi and Hamdi' and 'RMFR' datasets, achieved a respectable score of 99.98% in accuracy but identified a number of limitations. The model was only able to detect frontal face images, lacking the capability to discern different masked face positions. In addition, the model was unable to detect instances with 'no mask' or incorrect mask placement. Figure 6 presents a visual comparison of the results derived from the evaluation metrics.

The results displayed in Figure 7 further highlight the performance of the FMD model on the aforementioned FMD datasets.

The images and corresponding scores highlight that the model achieved a 100% accuracy rate across all presented images. This includes participants with and without face masks, those at various angles, and those wearing face masks even in cases where the images were significantly blurry.

Within the MFR phase of the model, according to the knowledge of the authors, neither the 'MFR dataset' nor the 'MDMFR dataset' has been extensively explored by other studies, with the exception of the 'MDMFR dataset'. This dataset has been exclusively evaluated by the original authors of the study. Hence, a comparison of results is performed on the 'MDMFR dataset' and not against the 'MFR Dataset' and the 'CRMFR Dataset' datasets. Therefore, as per Table 9, the following results for the 'MDMFR' dataset are presented as follows:

As highlighted in Table 9, the proposed model achieved a score of 99.82% within the evaluation metrics. Study [12] reported several limitations, with the model achieving an accuracy of 93.33%, which fell short of the 95% evaluation threshold. In addition, the study mentioned the need for further refinement and exploration of an alternative model or technique to enhance the model's performance. Furthermore, computational complexity was identified within the model and required a more diverse and quantified dataset, indicating that there was a lack of diversity when training and testing the model.

Table 10 presents the performance results, as no prior studies have employed the 'MFR dataset' and 'CRMFR dataset' which were

developed specifically for this research. The results are therefore unique to this study and are summarised in Table 10 as follows:

Contained in Table 10 above, the 'MFR dataset', which was released recently in 2023, has not been widely explored by other studies. The study still achieved respectable results in obtaining an accuracy, recall and F1-score of 99.40% and 99.41% for precision. Based on the 'CRMFR dataset', the purpose of creating the custom dataset was to validate the model's effectiveness in real-life scenarios, ensuring it can recognise individuals accurately and efficiently in real time. This model delivered commendable results with a score of 99.82% across all the aforementioned evaluation metrics. Overall, the developed model achieved an average score of 99.68% across all three models, which is above the 95% threshold.

The results displayed in Figure 8 highlight the performance of the MFR model on all three MFR datasets.

As displayed in Figure 8 above, the model was able to identify participants that belonged to a certain dataset and those that did not belong to the dataset, indicating 97.82% and 96.95% that certain participants did not belong to a dataset despite adverse lighting conditions and varied facial angles. In addition, the model successfully recognised participants despite challenging conditions, including adverse lighting conditions, varying facial angles, obstructions from face scarves and glasses, as well as blurry and distorted images.

Additionally, real-time detection and recognition were performed for both FMD and MFR, as shown in Figure 9, which illustrates the results and outcomes.

Figure 9 demonstrates the model's ability to accurately determine whether a participant was wearing a face mask or not. Additionally, the model effectively identified participants, achieving reliable accuracy in both FMD and MFR. It successfully detected and recognised individuals even under challenging conditions, such as poor lighting, varying facial angles and obstructions such as eyeglasses and head scarves. In addition, only a few existing literature models have tested these scenarios in real-time; hence, the models' successful implementation underscores their significant contribution to the field [12, 24, 25, 60, 61, 64].

4.1 | Model Performance Comparison

To evaluate the efficiency and scalability of the proposed models, key performance metrics were measured for both FMD and MFR. The comparison includes inference speed, computational complexity, model size, parameter distribution and peak memory usage for a batch of 30 images. This provides a comprehensive view of model trade-offs between lightweight deployment and recognition robustness (See Table 11).

In recent years, several deep learning architectures have been proposed for FMD and MFR, with varying trade-offs in inference speed, computational cost and model size. For instance, Loey et al. [68] employed ResNet-50 with SSD for mask detection, achieving good accuracy but at the cost of high computational complexity, with over 25M parameters and slower inference time.

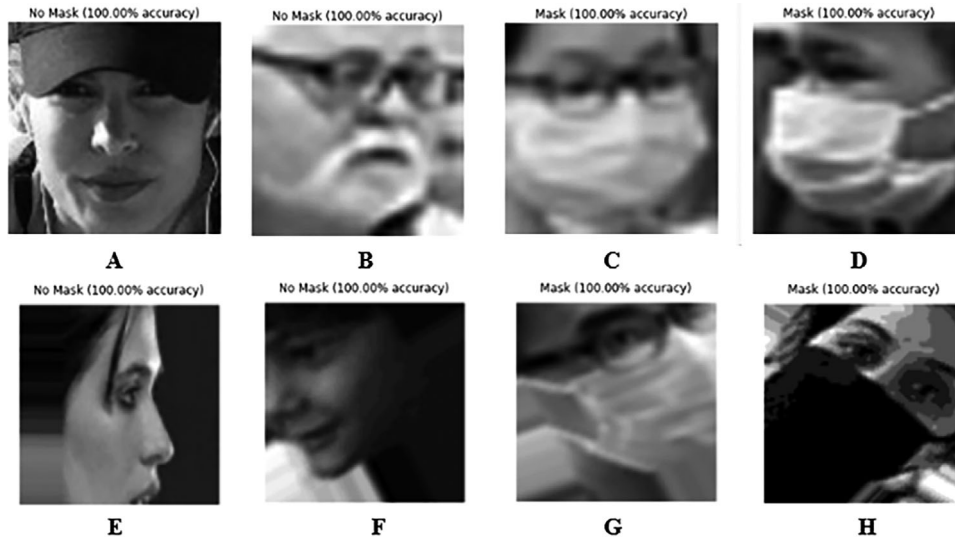


FIGURE 7 | FMD dataset results.

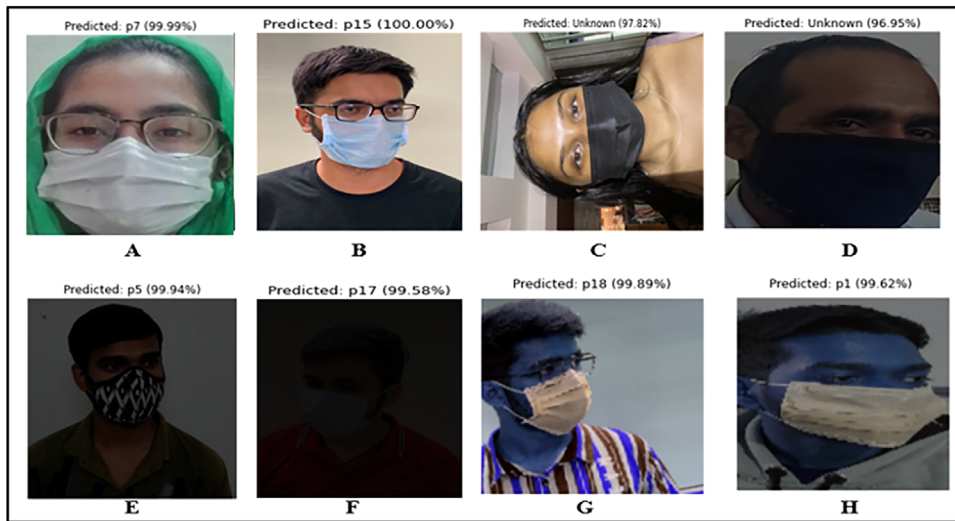


FIGURE 8 | MFR dataset results.

Similarly, Jiang and Fan presented RetinaFaceMask, which, while accurate, required significant FLOPs (~ 3.2 GFLOPs) that limit deployment on resource constrained devices [69]. More recent lightweight approaches, such as the EfficientNet based model by Hussain et al. [11], showed the potential of transfer learning to improve mask detection performance while maintaining a relatively small parameter count. However, these models still experienced higher inference times compared to MobileNet based architectures, which makes them less optimal for strict real-time applications [25]. For MFR, the work of Hariri et al. [32] proposed an efficient method leveraging deep CNN features and classifier optimisation. While robust against occlusion, their approach relied on large-scale embeddings and model weights exceeding 200 MB, which constrained scalability in deployment on edge. In another approach, Boutros et al. [70] proposed elastic feature learning with ArcFace on masked datasets, significantly improving recognition robustness but with a parameter-heavy model (~ 65 M) and large FLOP demand.

In comparison, the proposed pipeline leverages MobileNetV2 with custom layers for FMD and MTCNN + FaceNet (InceptionResNetV1) with a CNN head for MFR. The mask detection module achieves an inference time of 42.57 ms (23.49 FPS) with only 0.16 FLOPs, and a compact size of 35.85 MB with just 3.09M parameters, outperforming larger models in efficiency while maintaining accuracy. The recognition pipeline, despite integrating a deeper network, maintains a balanced inference time of 40.64 ms (24.61 FPS) with 0.08 FLOPs, significantly more computationally efficient than earlier approaches that exceeded 1 GFLOP. At 190.26 MB, the MFR model is larger, however, this model provides state-of-the-art representation capacity, outperforming lightweight recognition baselines in accuracy while preserving real-time feasibility. Overall, the proposed architecture offers an effective balance between speed, accuracy and resource efficiency, making it better suited for real-world deployment compared to prior research studies.



FIGURE 9 | Real-time FMD and MFR results.

TABLE 11 | Model performance comparison.

	FMD	MFR
Inference time	42.57 ms	40.64 ms
Frame per second	23.49	24.61
Floating point operation (FLOPs)	0.16	0.08
Model size	35.85 MB	190.26 MB,
No. of parameters	3,095,172 (11.81 MB)	49,863,456 (190.21 MB)
Trainable parameters	3,056,706 (11.66 MB)	16,620,810 (63.40 MB)
Non-trainable parameters	38,464 (150.25 KB)	1,024 (4.00 KB)
Peak memory (30 images)	160MB	480 MB

5 | Discussion

In the final analysis, it is evident based on the experiments and results that both the FMD and MFR models achieved extremely high accuracies and tremendously low loss values. Overall, the FMD model achieved an overall average accuracy of 100%, whilst the MFR model achieved an average of 99.68%. These results demonstrate the robustness and effectiveness of the models developed for this study.

However, in contrast to these outcomes, existing models [12, 24, 25, 59–65], as discussed in the results and analysis section, exhibited several limitations. These models faced numerous challenges, such as the model's inability to detect non-frontal face images [59, 60, 65], failure to recognise incorrect face mask usage and difficulty identifying when a user is not wearing a mask [24, 59, 60]. The model's performance is highly influenced by camera orientation [61], and it exhibits computational complexity [12, 24, 25, 61, 63, 64, 66]. It struggles with detecting obscured or partial faces [59], produces a significant number of false positives [62–65], and lacks real-time FMD capabilities [64]. Additionally, the confusion matrix revealed a high rate of both false positives and negatives, leading to poor scores in key evaluation metrics [62–65]. The model developed by author Ullah et al. [12] also faced challenges with detecting and recognising masked faces at varied angles and performed poorly in adverse lighting conditions, falling below the 95% accuracy threshold. These limitations highlight the need for further refinement and exploration of alternative models or techniques to improve

performance. Furthermore, the dataset used lacked sufficient diversity, indicating the necessity for a more comprehensive and quantified dataset to enhance the model's robustness [12].

The proposed model successfully performed both FMD and MFR tasks under a range of challenging conditions, including poor lighting, varied facial angles, low-resolution or blurry input, occlusions, fake mask scenarios, incorrect mask usage, closed-eye instances, object obstructions, and high computational demands. In poor lighting conditions, the FMD model achieved 99.32% accuracy for 'no mask' detection in complete darkness and 97.95% under partial illumination. Even under extreme dim lighting, when the participant's hands partially covered the face, the model maintained 64.08% accuracy, still labelling the case as "no mask". The MFR model consistently achieved 100% recognition accuracy in dark environments across both frontal and angled facial views. In scenarios involving varied angles, the FMD model achieved 100% accuracy for left and right orientations in both normal and dark conditions, while the MFR model achieved between 99.52% and 100% accuracy across all angles, even when participants' eyes were partially closed. For low-resolution or blurry images, both models sustained high performance using a standard-definition webcam (640×480 pixels, 96 dpi) operated on a mid-range laptop without the need for a high-performance graphics processing unit, demonstrating suitability for resource-limited environments. Notably, the FMD model was able to identify improper face coverings, such as hands used in place of masks, with an accuracy of 93.47%. These results confirm the robustness of the MobileNetV2 architecture for FMD and the

FaceNet InceptionResNetV1 with MTCNN and CNN architecture for MFR, both enhanced through hyperparameter tuning, and their ability to consistently outperform comparable models in both simulated and real-time applications.

6 | Conclusion

This study introduced hybrid CNN models that combine the MobileNetV2 model for FMD and the FaceNet Inception ResNetV1 with MTCNN for the MFR model to achieve superior results during model performance. The developed models include both batch testing and real-time testing for FMD and MFR to examine and highlight their robustness and ability to generalise in various scenarios, including real-time scenarios. The models demonstrated high accuracies, proving their effectiveness in real-time applications. Therefore, the CNN models have significantly improved accuracy in both simulated and real-time settings, enhancing performance in challenging scenarios. These scenarios include dimmed-light conditions, extreme dark lighting conditions, fake mask scenarios, incorrect use of a face mask, closed eye scenarios, occlusions of the face, object obstructions of the facial region, angled facial presentations and computational complexity.

The FMD model achieved an overall mean accuracy of 100% with a mean loss of 0.125, while the MFR model attained a mean accuracy of 99.68% with a mean loss of 2.35. Under real-time evaluation, the FMD model recorded detection rates ranging from 64.08% to 100% across 'no mask,' improper mask usage and correct mask placement categories in both standard and low-light environments. The MFR model consistently maintained 100% recognition accuracy under most of the real-time testing conditions, with the lowest recorded accuracy being 99.52% in the most challenging scenario of the participants' eyes being almost closed. Notably, both models sustained mean performance levels of $\geq 97\%$ accuracy under poor lighting and mean accuracy levels of $\geq 99\%$ across varying facial angles, even when partial occlusions or image degradation were present.

These results underscore the models' capacity to deliver high accuracy without reliance on high-end GPU hardware, thereby enhancing their feasibility for deployment in real-world biometric authentication and surveillance systems. Their adaptability, computational efficiency and precision position them as reliable solutions for institutions, including but not limited to companies, schools, universities, hospitals, and the banking sector, that require the implementation and utilisation of an FMD and MFR model. It therefore serves as a valuable tool for implementing robust FMD and MFR functionalities into existing applications or enhancing current biometric systems. The models' adaptability and accuracy make them an ideal choice for diverse sectors seeking to ensure compliance with future and current mask-wearing protocols and to enhance security measures with advanced MFR capabilities.

7 | Limitation and Future Work

Despite the successful development of the proposed models, certain limitations were observed within the developed FMD

and MFR models. A single false positive occurred during real-time MFR on the custom masked face dataset, primarily under extremely low lighting and beyond 200 centimetres in distance. This identified limitation may potentially be attributed towards the laptop webcam's limited hardware rather than the model itself. Different camera systems, such as CCTV, may yield varied results due to lens and focal length differences. In addition, training times were approximately 1 hour for FMD and 1.5 hours for MFR. These findings, therefore, highlight areas for future hardware and optimisation improvements.

Future work will focus on significantly reducing the training time of the models to enhance computational efficiency. This will involve exploring advanced techniques such as model pruning, knowledge distillation, and leveraging cutting-edge architectures combined with transfer learning to optimise performance without sacrificing accuracy. In addition, it will focus on testing the models with different types of cameras, including but not limited to CCTV cameras, thermal cameras and mobile phones, to evaluate their performance under varied hardware for image acquisition. Expanding the system to handle multi-face scenarios in real-time is a key priority for practical deployment. Additionally, efforts will be made to optimise the models for mobile and edge devices to enable low-latency, on-device inference. These improvements will broaden the applicability and accessibility of the models for FMD and MFR across diverse scenarios.

Author Contributions

Chezlyn Pillay: investigation, data curation, methodology, formal analysis, writing – original draft, writing – review and editing. **Seena Joseph:** conceptualisation, methodology, writing – review and editing, validation, project administration and supervision. **Brett van Niekerk:** writing – review and editing and supervision.

Acknowledgements

The authors gratefully acknowledge the Durban University of Technology for supporting this work.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets supporting the findings of this study are available from the following sources:

Face Mask Detection Dataset. Available at: <https://www.kaggle.com/datasets/omkargurav/face-mask-dataset>

Face Mask Detection ~12K Images Dataset. Available at: <https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset>

Masked Face Recognition Dataset. Available at: <https://www.kaggle.com/datasets/nanimasaka/masked-face-recognition-dataset>

Mask Detection and Masked Face Recognition. Available at: <https://zenodo.org/records/6408603> (MD5: 87d713a71eca5f44d233a8a06d5e462f).

Custom Real-Time Masked Face Recognition dataset developed for this study is available from the corresponding author upon request.

References

1. L. Qinjun, C. Tianwei, Z. Yan, and W. Yuying, "Facial Recognition Technology: A Comprehensive Overview," *Academic Journal of Computing & Information Science* 6, no. 7 (2023): 15–26.
2. S. Seng, M. Al-Ameen, and M. Wright, "A First Look Into Users' Perceptions of Facial Recognition in the Physical World," *Computers & Security* 105 (2021): 1–24.
3. M. K. I. Rahmani, F. Taranum, R. Nikhat, M. R. Farooqi, and M. A. Khan, "Automatic Real-Time Medical Mask Detection Using Deep Learning to Fight COVID-19," *Computer Systems Science and Engineering* 42, no. 3 (2022): 1181–1198.
4. M. Abdul-Al, G. K. Kyeremeh, R. Qahwaji, N. T. Ali, and R. A. Abd-Alhameed, *The Evolution of Biometric Authentication: A Deep Dive Into Multi-Modal Facial Recognition: A Review Case Study* (IEEE Access, 2024).
5. J. K. Kodros, K. O'Dell, J. M. Samet, C. L'Orange, J. R. Pierce, and J. Volckens, "Quantifying the Health Benefits of Face Masks and Respirators to Mitigate Exposure to Severe Air Pollution," *GeoHealth* 5, no. 9 (2021): 1–14.
6. M. Wojcik and D. Austin, "Criminal Justice and Covid-19," *Criminal Justice* 35, no. 3 (2020): 44–48.
7. S. Abbasi, H. Abdi, and A. Ahmadi, "A Face-Mask Detection Approach Based on YOLO Applied for a New Collected Dataset," in *2021 26th International Computer Conference (CSICC, 2021)*, 1–6.
8. S. Dharanesh and A. Rattani, "Post-COVID-19 Mask-Aware Face Recognition System," in *2021 IEEE International Symposium on Technologies for Homeland Security (HST)* (IEEE, 2021), 1–7.
9. A. Akhtar, S. Z. A. Hashmi, A. Raza, and N. Fatima, "Integrated Approach to Decrease Respiratory Tract Infection," *Advances in Integrative Medicine* (2025): 119–143.
10. Y. Wang, Y. Liu, C. Hao, et al., "Changes in the Characteristics and Risks of Bioaerosols in School Hospital Before and After the COVID-19 Epidemic," *Air Quality, Atmosphere & Health* 18 (2025): 2841–2852.
11. D. Hussain, M. Ismail, I. Hussain, R. Alroobaea, S. Hussain, and S. S. Ullah, "Face Mask Detection Using Deep Convolutional Neural Network and MobileNetV2-Based Transfer Learning," *Wireless Communications and Mobile Computing* 2022 (2022): 1–10.
12. N. Ullah, A. Javed, M. Ali 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* 34, no. 10 (2022): 9905–9914.
13. M. Mahmoud, M. Kasem, and H. Kang, "A Comprehensive Survey of Masked Faces: Recognition, Detection, and Unmasking," preprint, arXiv, May 9, 2024, <https://doi.org/10.48550/arXiv.2405.05900>.
14. J. Zhang, D. An, Y. Zhang, et al., "A Review on Face Mask Recognition," *Sensors* 25, no. 2 (2025): 387.
15. A. Alzu'bi, F. Albalas, T. AL-Hadhrani, L. B. Younis, and A. Bashayreh, "Masked Face Recognition Using Deep Learning: A Review," *Electronics* 10, no. 21 (2021): 2666.
16. H. Al-Dmour, A. Tareef, A. M. Alkalbani, A. Hammouri, and B. Alrahmani, "Masked Face Detection and Recognition System Based on Deep Learning Algorithms," *Journal of Advances in Information Technology* 14, no. 2 (2023): 224–232.
17. M. Eman, T. M. Mahmoud, M. M. Ibrahim, and T. Abd El-Hafeez, "Innovative Hybrid Approach for Masked Face Recognition Using Pre-trained Mask Detection and Segmentation, Robust PCA, and KNN Classifier," *Sensors* 23, no. 15 (2023): 6727.
18. M. Hosen and M. Islam, "Himfr: A Hybrid Masked Face Recognition Through Face Inpainting," preprint, arXiv, September 19, 2022, <https://doi.org/10.48550/arXiv.2209.08930>.
19. M. Taha, T. Mostafa, and T. Abd El-Rahman, "A Novel Hybrid Approach to Masked Face Recognition Using Robust PCA and GOA Optimizer," *Scientific Journal for Damietta Faculty of Science* 13, no. 3 (2023): 25–35.
20. R. Raj and A. Kos, "An Extensive Study of Convolutional Neural Networks: Applications in Computer Vision for Improved Robotics Perceptions," *Sensors* 25, no. 4 (2025): 1033.
21. M. Faruque, M. Islam, and M. Islam, "Advanced Masked Face Recognition Using Robust and Light Weight Deep Learning Model," *International Journal of Computer Applications* 186 (2024): 42–51.
22. X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A Review of Convolutional Neural Networks in Computer Vision," *Artificial Intelligence Review* 57, no. 4 (2024): 99.
23. I. Q. Mundial, M. S. Ul Hassan, M. I. Tiwana, W. S. Qureshi, and E. Alanazi, "Towards Facial Recognition Problem in COVID-19 Pandemic," in *2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)* (IEEE, 2020), 210–214.
24. J. S. Talahua, J. Buele, P. Calvopiña, and J. Varela-Aldás, "Facial Recognition System for People With and Without Face Mask in Times of the Covid-19 Pandemic," *Sustainability* 13, no. 12 (2021): 6900.
25. S. Habib, M. Alsanea, M. Aloraini, H. S. Al-Rawashdeh, M. Islam, and S. Khan, "An Efficient and Effective Deep Learning-Based Model for Real-Time Face Mask Detection," *Sensors* 22, no. 7 (2022): 1–13.
26. Q. Ilyas and M. Ahmad, "An Enhanced Deep Learning Model for Automatic Face Mask Detection," *Intelligent Automation & Soft Computing* 31, no. 1 (2022): 241–254.
27. D. Akingbesote, Y. Zhan, R. Maskeliūnas, and R. Damaševičius, "Improving Accuracy of Face Recognition in the Era of Mask-Wearing: An Evaluation of a Pareto-Optimized Facenet Model With Data Preprocessing Techniques," *Algorithms* 16, no. 6 (2023): 292.
28. O. Abdelwhab, "Masked Faces Recognition Using Deep Learning Models and the Structural Similarity Measure," *Optoelectronics, Instrumentation and Data Processing* 59, no. 6 (2023): 766–771.
29. Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face Recognition Systems: A Survey," *Sensors* 20, no. 2 (2020): 342.
30. A. Kheaksong, P. Samothai, P. Sanguansat, T. Ngamloed, K. Srisomboon, and W. Lee, "Evaluation of Masked Face Recognition of Facenet Implemented With Machine Learning Algorithms," in *2022 6th International Conference on Information Technology (InCIT)* (IEEE, 2022), 475–478.
31. Y. Hung, "A Laborer's Mask-Wearing Behavior Detection Approach in the Manufacturing Field," *Processes* 11, no. 4 (2023): 1086.
32. W. Hariri, "Efficient Masked Face Recognition Method During the Covid-19 Pandemic," *Signal, Image and Video Processing* 16, no. 3 (2022): 605–612.
33. M. S. Islam, E. H. Moon, M. A. Shaikat, and M. J. Alam, "A Novel Approach to Detect Face Mask Using CNN," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)* (IEEE, 2020), 800–806.
34. G. Kaur, R. Sinha, P. K. Tiwari, et al., "Face Mask Recognition System Using CNN Model," *Neuroscience Informatics* 2, no. 3 (2022): 1–9.
35. M. Boulos, "Facial Recognition and Face Mask Detection Using Machine Learning Techniques," *Computer Science* (Montclair State University, 2021).
36. B. Mandal, A. Okeukwu, and Y. Theis, "Masked Face Recognition Using Resnet-50," preprint, arXiv, April 19, 2021, <https://arxiv.org/abs/2104.08997>.
37. Z. Arya and V. Tiwari, "An Implementation of Real-Time Automatic Masked and Unmasked Face Recognition Using Ibp Algorithm," *ADB Journal of Engineering Technology* 10, no. 1 (2021): 1–7.
38. S. Mhadgut, "Masked Face Detection and Recognition System in Real Time Using YOLOv3 to Combat COVID-19," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (IEEE, 2021), 1–7.

39. M. S. M. Suhaimin, M. H. A. Hijazi, C. S. Kheau, and C. K. On, "Real-Time Mask Detection and Face Recognition Using Eigenfaces and Local Binary Pattern Histogram for Attendance System," *Bulletin of Electrical Engineering and Informatics* 10, no. 2 (2021): 1105–1113.
40. V. Aswal, O. Tupe, S. Shaikh, and N. N. Charniya, "Single Camera Masked Face Identification," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)* (IEEE, 2020), 57–60.
41. H. Winahyu, E. Wardihani, and S. Beta, "Attendance System Based on Face Recognition, Face Mask and Body Temperature Detection on Raspberry Pi," in *2021 International Seminar on Application for Technology of Information and Communication (iSemantic)* (IEEE, 2021).
42. G. Yang, W. Feng, J. Jin, et al., "Face Mask Recognition System With YOLOV5 Based on Image Recognition," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)* (IEEE, 2020), 1398–1404.
43. N. Damer, S. McLaughlin, A. Dantcheva, F. Boutros, A. Hadid, and S. Marcel, "Masked Face Recognition: Human vs. Machine," preprint, arXiv, March 2, 2021, <https://arxiv.org/abs/2103.01924>.
44. P. C. Neto, F. Boutros, J. R. Pinto et al., "My Eyes Are up Here: Promoting Focus on Uncovered Regions in Masked Face Recognition," preprint, arXiv, August 2, 2021, <https://doi.org/10.48550/arXiv.2108.00996>.
45. M. Zhang, B. Ma, G. Song, Y. Wang, H. Li, and Y. Liu, "Towards Large-Scale Masked Face Recognition," preprint, arXiv, October 25, 2023, <https://doi.org/10.48550/arXiv.2310.16364>.
46. S. Jaiswal, A. Verma, and A. Mukherjee, "Mask-up: Investigating Biases in Face Re-identification for Masked Faces," preprint, arXiv, February 21, 2024, <https://doi.org/10.48550/arXiv.2402.13771>.
47. A. Al-Rammahi, "Face Mask Recognition System Using MobileNetV2 With Optimization Function," *Applied Artificial Intelligence* 36, no. 1 (2022): 2145638.
48. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted Residuals and Linear Bottlenecks," in *2018 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2018), 4510–4520.
49. L. Yong, L. Ma, D. Sun, and L. Du, "Application of MobileNetV2 to Waste Classification," *PLoS ONE* 18, no. 3 (2023): 1–6.
50. S. Bharadwaj, R. Jha, and J. Kumar, et al., "Comparative Study of Mobilenetv2, Simple CNN and Vgg19 For Image Classification," *Journal of Data Acquisition and Processing* 39, no. 1 (2024): 152–162.
51. I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, Present, and Future of Face Recognition: A Review," *Electronics* 9, no. 8 (2020): 1188.
52. T. Thaher, M. Mafarja, M. Saffarini, A. H. H. M. Mohamed, and A. A. El-Saleh, "A Comprehensive Review of Face Detection Techniques for Occluded Faces: Methods, Datasets, and Open Challenges," *Computer Modeling in Engineering & Sciences* 143 (2025): 2615–2673.
53. O. Gurav, "Face Mask Detection Dataset," 2020, Accessed 30 July 2023, <https://www.kaggle.com/datasets/omkargurav/face-mask-dataset>.
54. A. Jangra, "Face Mask Detection ~12K Images Dataset," 2020, Accessed 30 July 2023, <https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset>.
55. A. Singh, "Masked Face Recognition Dataset," 2023, Accessed 30 July 2023, <https://www.kaggle.com/datasets/nanimasaka/masked-face-recognition-dataset>.
56. D. Yi, J. Ahn, and S. Ji, "An Effective Optimization Method for Machine Learning Based on ADAM," *Applied Sciences* 10, no. 3 (2020): 1073.
57. G. Aryotejo, P. W. Adi, F. Ernawan, and M. Mufadhol, "Detecting and Counting Coin Using Opencv and Watershed Algorithm," in *AIP Conference Proceedings* (AIP, 2023), 060002-1–060002-6.
58. B. E. Granger and F. Perez, "Jupyter: Thinking and Storytelling With Code and Data," *Computing in Science & Engineering* 23, no. 2 (2021): 7–14.
59. Y. Hu, Y. Xu, H. Zhuang, Z. Weng, and Z. Lin, "Machine Learning Techniques and Systems for Mask-Face Detection—Survey and a New OOD-Mask Approach," *Applied Sciences* 12, no. 18 (2022): 1–37.
60. R. Bania, "Ensemble of Deep Transfer Learning Models for Real-Time Automatic Detection of Face Mask," *Multimedia Tools and Applications* 82, no. 16 (2023): 25131–25153.
61. J. V. B. Benifa, C. Chola, A. Y. Muaad, et al., "FMDNet: An Efficient System for Face Mask Detection Based on Lightweight Model During COVID-19 Pandemic in Public Areas," *Sensors* 23, no. 13 (2023): 6090.
62. S. Guo, L. Li, T. Guo, Y. Cao, and Y. Li, "Research on Mask-Wearing Detection Algorithm Based on Improved YOLOv5," *Sensors* 22, no. 13 (2022): 1–16.
63. H. Goyal, K. Sidana, C. Singh, A. Jain, and S. Jindal, "A Real Time Face Mask Detection System Using Convolutional Neural Network," *Multimedia Tools and Applications* 81, no. 11 (2022): 14999–15015.
64. P. Fazeli Ardekani, S. Tale, and M. Parseh, "Face Mask Recognition Using a Custom CNN and Data Augmentation," *Signal, Image and Video Processing* 18 (2023): 255–263.
65. E. Zhang, "A Real-Time Deep Transfer Learning Model for Facial Mask Detection," in *2021 Integrated Communications Navigation and Surveillance Conference (ICNS)* (IEEE, 2021), 1–7.
66. B. Chakma, M. A. Masud, T. Ahamed, and M. H. Tusher, *Identification of Face Mask Using Convolutional Neural Network-Based Efficientnet Model* (Khulna University Studies, 2022), 531–538.
67. C. Arora, R. Gupta, and D. Sridhar, *Face Mask Detection using Deep Learning CNN Architecture* (IJSER, 2022), 669–677.
68. 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* 167 (2021): 108288.
69. X. Fan and M. Jiang, "RetinaFaceMask: A Single Stage Face Mask Detector for Assisting Control of the COVID-19 Pandemic," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE, 2021), 832–837.
70. F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Elasticface: Elastic Margin Loss for Deep Face Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2022), 1578–1587.