International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Three Classification Models for Masked Faces

[1]Mahmoud Hardan Fahim, [2]Mohamed Abdel-Nasser, [3]Moataz Abdelwahab, [4]Adel B. Abdel-Rahman

Submitted: 03/02/2024 **Revised**: 11/03/2024 **Accepted**: 18/03/2024

Abstract: In recent years, deep learning technology has grown quickly and made great progress in the field of computer vision. Convolutional neural networks (CNNs), Residual networks (ResNets), Inception networks, and MobileNet are among the deep learning classification models that have gained widespread use for picture classification applications. The COVID-19 pandemic in early 2020 led to the widespread use of face masks as a vital containment strategy. The use of masks has given rise to two problems. The first is whether or not the individual is donning a mask. The second problem is that these masks cover almost half of the human face, which obviously modifies facial appearance. In this paper, we propose three proposed classification models (Tree Models). The proposed models are tested in three distinct cases. Case 1 (2 Outputs), the models are evaluated for Masked/Unmasked classification. Case 2 (50 Outputs) is aiming for recognition of 50 different persons. The models are tested on an expanded set of subjects, from 50 to 85, in Case 3 (85 Outputs).

Keywords*: Convolutional Neural Networks (CNNs), Classification Models, Masked/Unmasked Classification, Identification*

Introduction

Computer vision using deep learning technology involves the application of advanced neural network architectures to analyze and interpret visual data. Deep learning models, particularly Convolutional Neural Networks (CNNs), have proven effective in tasks such as image classification, object detection, facial recognition, and image segmentation. These models can automatically learn hierarchical representations from raw image data, enabling them to understand complex patterns and features.

Various deep learning architectures, including not only CNNs but also networks like Residual Networks (ResNets), Inception Networks, and MobileNet, cater to different requirements and challenges within computer vision tasks. These technologies have significantly advanced the capabilities of computer vision systems, leading to improved accuracy and efficiency in tasks that traditionally relied on handcrafted features and algorithms. Image classification, the task of assigning images to predefined classes, is a fundamental challenge in computer vision, forming the basis for tasks like localization, detection, and segmentation. While humans perform this task effortlessly, it poses significant challenges for automated systems, including viewpoint-dependent object variability and high in-class variability. Traditionally, a dual-stage approach involving handcrafted features and a trainable classifier was used, but this approach heavily depended on feature extraction design, proving challenging [4].

In recent years, deep learning models, particularly Convolutional Neural Networks (CNNs), leveraging multiple layers for nonlinear information processing, have shown significant progress in overcoming these challenges. The rise of deep CNNs can be attributed to the deep learning renaissance, fueled by GPUs, larger datasets, and improved algorithms. Advances like the first GPU implementation and the application of maximum pooling have contributed to the popularity of deep CNNs in image recognition, classification, and detection tasks.

The term "masked facial recognition" describes the capacity to identify people even while they have on masks. The identification of facial characteristics like the mouth, nose, and eyes is the mainstay of traditional face recognition systems. But these systems have faced difficulties due to the increasing usage of face masks, particularly in reaction to incidents like the COVID-19 epidemic. Modern facial recognition algorithms and technologies have been developed to overcome this difficulty by recognizing faces even when a significant portion of them are hidden by masks. These systems might make use of a number of strategies, such as:

- Mask Detection: By integrating facial recognition with mask detection technology, the system

[1]*Electrical Department, Faculty of Engineering South Valley University,Qana, Egypt*
engmahfah2005@gmail.com
[2]*Faculty of Engineering, Aswan University*
Aswan , Egypt
mohamed.abdelnasser@aswu.edu.eg
[3]*School of Electronics, Communications and Computer Engineering*
E-JUST,Egypt

moataz.abdelwahab@ejust.edu.eg

[4]*Electrical Department, Faculty of Engineering, South Valley University,Qana, Egypt School of Electronics, Communications and Computer Engineering, E-JUST,Egypt*

can determine whether a person is wearing a mask and modify its recognition strategy appropriately [7-17].

- Database enhancement: To help the system learn and detect masked faces, databases used to train facial recognition models are updated with pictures of people wearing masks [18-20].

The development of masked facial recognition technologies has become crucial for various applications, including security, access control, and public health. However, ethical considerations and privacy concerns related to facial recognition, even with masks, continue to be subjects of debate and scrutiny. As a result, the implementation and regulation of masked facial recognition systems often involve careful consideration of privacy laws and ethical guidelines.

This paper explores a deep learning-based models and testing it on 3 cases of face recognition. The paper's structure includes a section on related work in Sect. 2, which provides insights into history of the face recognition. Section 3 presents Study cases and datasets. The proposed deep learning models are also explored in this section 4. Finally, we show the results in section 5.

Related Work

The face recognition (FR) is an active area of work in the field of image processing. Face recognition universally used in the number of applications, including social networking, access control, video monitoring, and law enforcement. In 2021, Raji, Inioluwa Deborah, and Genevieve Fried. [4] divided the lifetime of face recognition, from 1964 to now, into four time periods. The first period, Early Research Findings period, (1964-1995), 31 years, can be considered the foundation period for facial recognition. Tomkins and Mc Carter [24] can be considered among the first researchers in the field of facial recognition, as they presented in 1964 research to identify human emotions, such as anger and happiness, using 69 photos of 24 firefighters. In 1986, Bruce and Young [26] developed a theoretical model and set of terms to understand and discuss how familiar faces are recognized and the relationship between recognition and other aspects of face processing.

The second period identified by Raji and Fried is a stretch from 1996 to 2006 and has been termed commercial viability as "New Biometrics". In this framework, the researchers sought ways to solve the problems of face recognition and to increase the efficiency of the existing algorithms. P. J. Phillips, Hyeonjoon Moon, S. A. Rizvi, and P. J. Rauss [28] presented the Face Recognition Technology (FERET) program, which addresses critical requirements for reliable face recognition systems by creating a large database of facial images and performing tests. The FERET database includes 14,126 images from 1,199 individuals. The third FERET test for face recognition in 1996 assessed the state-of-the-art, identified areas for future research, and measured the performance of the algorithm. Yang [29] investigated the use of kernel Principal Component Analysis and kernel Fisher linear discriminant for face recognition, focusing on Kernel Eigenface and Kernel Fisherface methods. These methods consider higher-order correlations and provide better representation. The results show that kernel methods achieve lower error rates and better performance than Eigenface, Fisherface, and ICA-based methods. Yu, Hua, and Jie Yang [30]) proposed A direct LDA algorithm for high-dimensional data classification, specifically face recognition. The algorithm discards the null space of S, which lacks discriminative information, while maintaining the null space of S. Computational techniques are introduced to handle large scatter matrices efficiently. The result was a single LDA algorithm with Fisher’s exact criterion solution. Juwei Lu, K.N Plataniotis, and A.N. Venetsanopoulos [31], proposed an algorithm for face recognition systems that addresses the shortcomings of linear discriminant analysis (LDA) methods. The algorithm is efficient and cost-effective compared with other commonly used two-face database methods, and outperforms traditional methods such as Eigenfaces, Fisherfaces, and D-LDA.

The period from 2007 to 2013 was the third period of Raji and Fried timeline. Raji and Fried [4] called this period "Mainstream Development for Unconstrained Settings". The development of the Internet has provided researchers with a great opportunity to develop available data for testing programs and algorithms. This was demonstrated by the creation of Internet-based datasets. Huang et al [32] introduced Labeled Faces in the Wild database designed to study the unconstrained face recognition problem, encompassing labeled face photographs from everyday life. The database exhibited natural variability in factors such as pose, lighting, race, accessories, occlusions, and background. It provides specific experimental paradigms for consistent and comparable research, including baseline results from state-of-the-art face recognition systems and parallel databases, including an aligned version. HMDB [33]) is a large action video database collected from digitized movies to YouTube, containing 51 action categories and 7,000 manually annotated clips. Some have sought another direction: the development of existing algorithms. Tan and Triggs [34] studied a combination of robust luminance normalization, local texture-based face representations, and distance-transformation-based matching metrics to improve face recognition. It introduces a simple preprocessing chain, local three-patterns (LTP), and demonstrates that replacing the local histogram with a local distance transform-based similarity meter boosts performance. This method outperforms popular datasets such as Grand Challenge v1, Extended Yale-B, and CMU PIE facial recognition. Tan and Triggs [35], returned in 2010 to improve the same point by adding Kernel PCA feature extraction and incorporating Gabor wavelets and LBP, resulting in more accurate local appearance cues. It outperforms existing preprocessors on three challenging data sets, achieving a Face Verification Rate of 88.1% at 0.1% False Accept Rate

The last period proposed by Raji and Fried [4] extends from 2014 until now, when deep learning broke into the field of face recognition. However, this must be divided into two periods. From 2014 to before the end of 2019, face recognition researchers worked with the introduction of deep learning in the field of face recognition to improve existing algorithms and create artificial intelligence models capable of solving problems facing this field. Sun, Yi, et al. [36] proposed Two DeepID3 deep neural network architectures for face recognition, reconstructed from the VGG net and GoogLeNet layers. These architectures achieved 99.53% LFW face verification accuracy and 96.0% LFW rank-1 face identification accuracy, with joint supervisory signals added during training. Parkh et al [37] focused on face recognition using a CNN and large-scale training datasets. It presents a combination of automation and human-in-the-loop to create a large dataset of 2.6M images, discussing data purity and time trade-offs. They also presented methods and procedures for achieving comparable state-of-the-art results on the standard LFW and YTF face benchmarks. Xiang et al. [38] proposed a light CNN model to solve the problem of noisy labels in datasets. Yaniv et al. [39] converted 2D images to 3D images to increase accuracy to 97.35%. Cui et al [40] proposed a 2D to 3D model to decrease the face recognition challenges. Abul Hasent et al. [41] proposed a single CNN model that can be trained on different datasets. Wing et al [42] increased the performance of CNN by using a constrained triplet loss layer. Guben et al [43] updated the ResNet architecture to increase accuracy. Chunrui et al [42] presented a CNN model which compares 2 images. Ye et al [44] improved a strategy to detect a tiny face on images with an accuracy of 93.7% on FDDB and 82.6% on the Wider Faces Dataset.

And the end of 2019, the Covid-19 disease appeared. Many countries compelled citizens and visitors to wear medical face masks. New challenges have appeared in face recognition. Therefore, wearing a face mask is a new challenge. The lack of a dataset for masked faces and identifying persons is another challenge. Venkateswarlu et al [45] presented a MobileNet model that can detect mask parts of faces. Sanjaya et al [46] updated the MobileNet model to the MobileNetV2 model and increased the performance accuracy to 96.85%. Anwarn et al [47] developed an open-source MaskFace tool used to convert the face dataset to the masked face dataset. Chowdary et al [48] provided a deep learning model to detect whether faces are masked or not and trained using Simulated Masked Face Dataset. Militant and Dionisio [49] proposed an IOT system which can detect masked or not and test on its 25K person dataset. Chuvada et al [50] introduced a multi-stage CNN which could detect masked/ unmasked faces in images or video frames. Joshi et al [51] presented a deep learning frame which used MTCNN to detect the faces and MobileNetV2 to find masked parts. Din, Nizam Ud, et al. [52] proposed a method to remove masks from the face automatically and to rebuild corrupted regions while preserving the initial face structure. Nieto-Rodriguez et al [1] introduced a system to detect medical masks in operating rooms, aiming to minimize false positives and trigger alarms for healthcare personnel not wearing the mask. The system uses two face detectors, one for the face and the other for the mask, and color processing to enhance the true positive to false positive ratio. It achieved 95% recall and a false positive rate of less than 5%. The system also offers real-time image processing at 10 fps VGA resolution, with background subtraction techniques up to 20 fps. Dewantara and Rhamadhaningrum [2] used Adaptive Boosting and Cascade Classifier based methods to detect mask wearers on various facial poses. OpenCV's models are effective for frontal and profile faces but not for mask-wearers. To detect multipose mask-wearers, a new nose- and mouth-based model was trained independently. The Caspeal and AISL face databases were used to train the Haar-like, LBP, and HOG features. Haar-like features achieved the best detection accuracy of 86.9%, whereas LBP features outperformed them in computation time, achieving less than 30 ms from image loading to detection. Wang et al [19] presented three masked face datasets: Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD), and Simulated Masked Face Recognition Dataset (SMFRD). RMFRD is the world's largest offering free access to industry and academia. The multi-granularity model achieved 95% accuracy.

Study Cases and Dataset

Over the past decade, face recognition (FR) has become increasingly vital for applications such as security and access control. The performance of FR technology has approached human levels, yet the advent of the Covid-19 pandemic has introduced setbacks. The widespread use of medical masks, driven by concerns about the dangers of Covid-19, has posed challenges by obscuring a significant portion of facial data. Mandates in many countries for citizens and visitors to wear medical face masks have created new hurdles for face recognition systems. Distinguishing individuals with or without masks and the lack of datasets containing images of masked faces are among the emerging challenges in face recognition technology.

This section presents three case studies. The first case, Case 1 (two outputs), involved evaluating the models for masked/unmasked classification, resulting in two output classes. The second and third cases were designed to examine the proposed models as identification models. Case 2 (50 outputs) involves evaluating the models on the Small Masked/Unmasked Identify dataset with the aim of identifying 50 different classes. Case 3 (85 outputs) tested the models on the Small Masked/Unmasked Identify dataset with an extended set of 50 output classes. This comprehensive testing across different scenarios demonstrates the versatility and applicability of the proposed models for diverse image classification tasks

The "Real-World Masked Face Recognition Dataset" (RMFRD) [18] is utilized, featuring 5,000 images of 525 individuals with masks and 90,000 images of the same individuals without masks. The main dataset is further divided into three smaller datasets for specific purposes.

The "2 Outputs Dataset," or "Mask/Unmask Dataset," consists of 268 masked images and 15,748 unmasked images, tailored for binary classification into two classes: masked or unmasked. During testing, this dataset was used in Case 1 to assess the performance of the proposed models.

In the second and third cases, two smaller datasets were created by dividing the selected RMFRD subsets. The first set comprised 50 subjects, whereas the second set consisted of 85 subjects. The main advantage of the RMFRD is its construction, which involves subjects divided into masked and unmasked categories. This structuring allows for more specific and targeted training and testing scenarios, especially in the context of face recognition tasks, where the presence or absence of masks significantly impacts the classification process. Examples of dataset illustrated in Fig.(1).

Fig.(1) Examples of Masked/Unmasked Dataset

The proposed models

A Convolutional Neural Network (CNN) is a specialized neural network that combines two essential layers: the convolution layer and the pooling layer, forming a type of feedforward neural network Fig. (2). The convolution layer's role is to train fewer parameters, extracting feature information from input data in a locally connected manner, offering advantages in training efficiency due to lower data requirements. In contrast to fully connected layers, convolution layers focus on local connectivity. Following the convolution layer, the pooling layer is employed to reduce data volume, mitigate overfitting, and enhance overall performance.

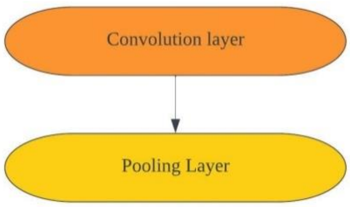
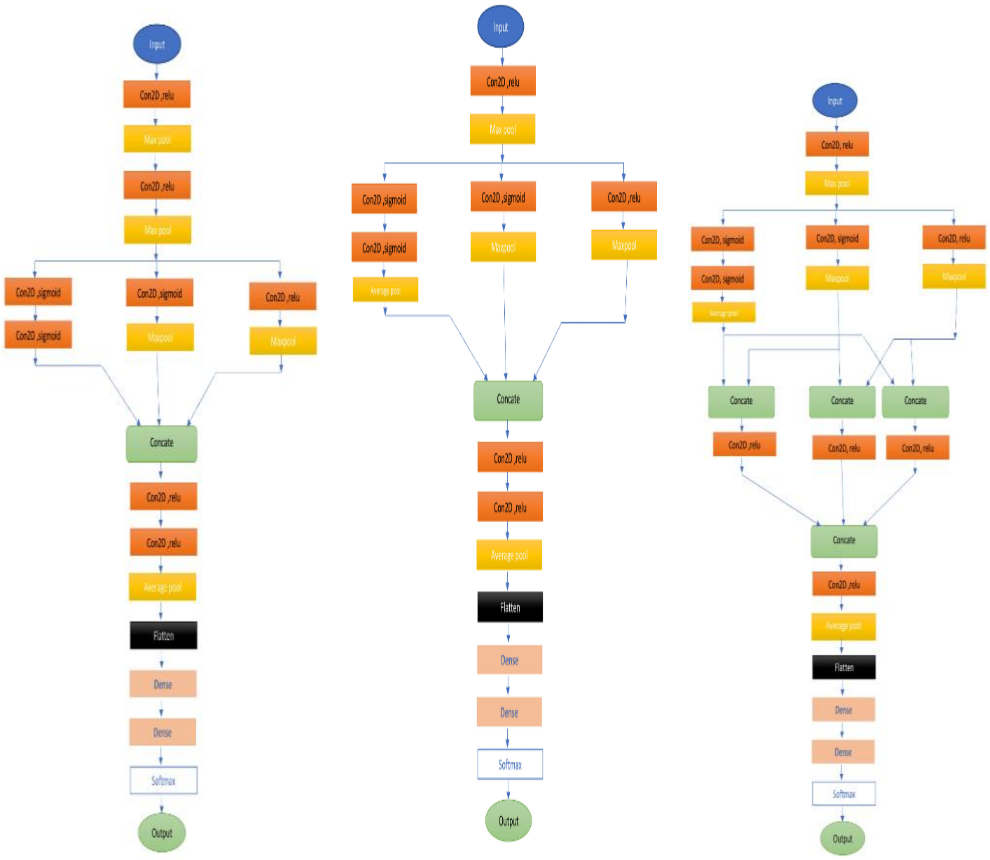


Fig.(2) The Constructure Of Convolution Neural Network (CNN)

In this section, we introduce three proposed models denoted as Fig. (3) (a, b, c), employing a technique that structures the model as a small tree with three branches. Each branch is composed of distinct layers, including convolutional and pooling layers with diverse kernels, aimed at capturing varied image features. Proposed Model No. 1 is constructed with two Convolutional Neural Networks (CNNs) featuring convolution layers, ReLU activation functions, and Max pooling layers. These are distributed into three branches: two with CNN layers (convolution, ReLU activation, and Max pooling) and one with convolution layers using a sigmoid activation function and Max pooling. The outputs are concatenated, followed by additional convolution layers, ReLU activation, a CNN layer with Average pooling, a Fully connected layer, two Dense layers, and a softmax activation function (Fig. 3-a). The Proposed Model No. 2 differs by removing one CNN before branching and adding an Average pooling layer in the first branch (Fig. 3-b). The Proposed Model No. 3 is an advancement of The Proposed Model No. 2, featuring four concatenate layers, with three settling after the three branches. The outputs from pairs of branches are concatenated into each layer, and after the final Concatenate layer, the three branches are merged. The Proposed Model No. 3 follows a similar structure to the previous models, with the exception of removing one convolution layer after the final Concatenate layer. This adjustment may impact the network architecture, potentially influencing the model's performance in terms of features extraction and classification.



(a) Proposed Model No.1 *(b) Proposed Model No.2* *(C) Proposed Model No.3*

Fig. (3) The Proposed Models

Results

In this section, we delve into the training and testing processes for each proposed model. In the first case, Case 1 (2 Outputs), training utilizes the Small Mask/Unmask Dataset with an 80:20 ratio for the train and test sets, focusing on the lower half of images. In the second and third cases, training involves the upper half of unmasked images from two smaller datasets derived from the RMFRD. Testing, in turn, employs the upper part of masked pictures as shown in Fig. (4).

Several factors were used to measure the efficiency of the proposed model. These factors include efficiency, error rate, and the time taken to achieve results, which, in our case, were measured in milliseconds (ms). Table (1) shows the values of the three factors, depending on the case and the proposed models.

(a) The Full Face

(b) The Upper half Of Face

(c) The Lower half Of Face

Fig (4) The Image Half's Of Face

Table (1)

The Accuracy, Error, And Execution Time After Training The Three Proposed Models In The Three Study Cases

Study case		Accuracy	Error	Execute Time
Case (1)	Proposed Model No.1	100%	0.0014%	39.00 ms
	Proposed Model No.2	98.8%	1.2%	64.22 ms
	Proposed Model No.3	100%	0.0011%	35.01 ms
Case (2)	Proposed Model No.1	99.882%	0.018%	39.75 ms
	Proposed Model No.2	92.506%	7.494%	65.54 ms
	Proposed Model No.3	99.988%	0.012%	39.76 ms
Case (3)	Proposed Model No.1	94.91%	5.09%	41.9 ms
	Proposed Model No.2	90.81%	9.19%	66.86 ms
	Proposed Model No.3	95.13%	5.86%	99 ms

From Table (1), comparing the three proposed models in terms of efficiency and error rate, the third proposed model appears to be the best in terms of efficiency in all cases. Even in the case of equality between the third and first proposed models in the first scenario, the error rate favors the third proposed model. One of the key factors we studied to determine the efficiency of our models is the execution time (time to reach a decision).

In the first case, the third model was the fastest, closely approaching equality with the first model in the second case, but becoming slower in the third scenario. This is attributed to the third model's reliance on multiple concatenated layers in its design.

Figure (5) illustrates the change in execution time for different study cases (number of output classes), with the first case as the baseline for comparison. The figure shows that the second model had the least change in execution time, with a change rate not exceeding 4.11% from the first case (two outputs) to the third case (85 outputs). This was the first model with a change rate that did not exceed 7.5%. However, the third model experiences the most significant change in execution time, with a change between the second case (50 outputs) and the third case (85 outputs) of approximately 148%.

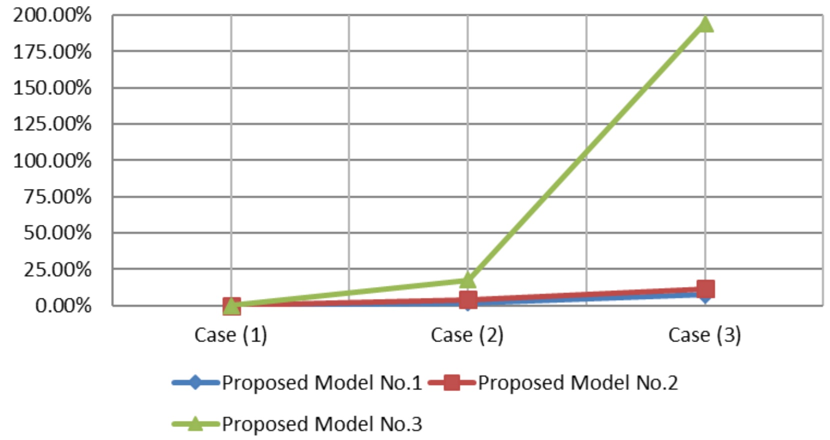


Fig. (5) Execute Time vs. No. of outputs

Recall, Specificity, Precision, and F1-score serve as standard metrics for assessing the performance of classification models, each offering distinct insights:

Recall (Sensitivity or True Positive Rate): Recall = TP / (TP + FN). These metrics gauge a model's ability to capture all relevant instances, representing the ratio of actual positive instances that are correctly identified. Specificity (True Negative Rate): Specificity = TN / (TN + FP). It evaluates a model's capacity to correctly identify negative instances, indicating the proportion of actual negative instances that are correctly recognized. Precision (Positive Predictive Value): Precision = TP / (TP + FP). Precision measures the accuracy of positive predictions, revealing the ratio of instances predicted as positive that were genuinely positive. F1-score = 2 × (Precision × Recall) / (Precision + Recall). The F1-score is the harmonic mean of the precision and recall, offering a balanced measure between the two. These metrics are especially valuable when handling imbalanced datasets or scenarios in which different error types carry varying degrees of importance. Although high values for Recall, Specificity, Precision, and F1-score generally signify robust model performance, the choice of the most relevant metric depends on the specific objectives of the classification task. Table 2 presents the recall, specificity, precision, and F1-score values of the proposed models. Table 2 illustrates that a significant advantage of our proposed models in all study cases is their performance, which consistently exceeds 0.90.

TABLE 2
RECALL, SPECIFICITY, PRECISION, AND F1-SCORE IN CASE OF TRAIN MODELS

Study case	Model	Recall	Specificity	Precision	F1-score
Case (1)	Proposed Model No.1	0.999993	0.999972	0.999986	0.99999
	Proposed Model No.2	0.993964	0.976285	0.988	0.990973
	Proposed Model No.3	0.999995	0.999978	0.999989	0.999992
Case (2)	Proposed Model No.1	0.9999099	0.99963970	0.99981982	0.99986485
	Proposed Model No.2	0.9610713	0.86056896	0.92506	0.94272189
	Proposed Model No.3	0.9999399	0.99976002	0.99988	0.99990999
Case (3)	Proposed Model No.1	0.9738854	0.903131	0.949100	0.96133296
	Proposed Model No.2	0.9518319	0.831654	0.908091	0.92944698
	Proposed Model No.3	0.9701203	0.890314	0.941974	0.95584024

Table 3 reveals that the images utilized in our approach are larger than those in other methods. Following a comprehensive analysis of various efficiency indicators and a comparison with alternative methods, the results underscore the superior efficiency and reduced execution time of our proposed models in face recognition scenarios, whether with or without masks. This suggests a noteworthy enhancement in accuracy and speed compared to existing solutions, rendering our models well-suited for practical applications in real-world settings.

Conclusion

In summary, this paper introduces three proposed classification models based on Convolutional Neural Networks (CNNs), tailored for diverse face recognition tasks. The models are evaluated in three cases: Masked/Unmasked classification, identification among 50 different classes, and an extended set of 85 classes. Despite challenges posed by widespread mask-wearing due to Covid-19, the models show case versatility. The Real-World Masked Face Recognition Dataset (RMFRD) is leveraged, containing masked and unmasked images. Three smaller datasets are derived for binary classification and identification cases with 50 and 85 output classes. The proposed models' efficiency and error rates are compared, revealing the third model's superior efficiency. Even when the third and first models are equal in the first case, the error rate favors the third model. Efficiency, measured by execution time, positions the third model as the fastest in the first case, but it slows in the third case due to its reliance on concatenate layers. The second model exhibits the least change in execution time, while the third model undergoes a significant 148% change from the second to the third case.

TABLE 3

COMPARISON BETWEEN OUR PROPOSED MODELS AND OTHER METHODS IN CASE 1

Model	No. Of Images Used To Train	Accuracy	Recall	Precision	F1-Score	Execution Time
Proposed Model No.1	15781	100%	99.9993%	99.9986%	99.999%	39.00 ms
Proposed Model No.2	15781	98.8%	99.3964%	0.988%	99.0973%	64.22 ms
Proposed Model No.3	15781	100%	99.9995%	99.9989%	99.9992%	35.01 ms
Nieto et al [1]	677 test case	--	95%	--	--	34 ms
Dewantara et al [2]	1000	86.9%	--	--	--	30 ms
Jiang et al [3]	7950	--	--	--	93.73%-97.95%	--
Line et al [5]	992	94.6%-95.8%	--	--	--	1.791 sec
Petrović et al [6]	video	84-91%	--	--	--	--
Zereen et al[21]	5504	97.13%	--	--	--	--
Fang et al [22]	6024	--	--	96.4%	--	83.1 ms
Merculdo et al [23]	4095	98%	--	--	--	4.7 sec
Rudraraju et al [25]	1270	90%	--	--	--	--
Talahua et al [27]	13359	99.65	--	--	--	840 ms
Wang et al [19, 20]	7804	--	--	--	99.49%-99.65%	112.5 ms

References

[1] A. Nieto-Rodriguez, M. Mucientes, and V. M. Brea, "System for medical mask detection in the operating room through facial attributes," in *Pattern Recognition and Image Analysis: 7th Iberian Conference, IbPRIA 2015, Santiago de Compostela, Spain, June 17-19, 2015, Proceedings 7*, 2015: Springer, pp. 138-145.

[2] B. S. B. Dewantara and D. T. Rhamadhaningrum, "Detecting multi-pose masked face using adaptive boosting and cascade classifier," in *2020 International Electronics Symposium (IES)*, 2020: IEEE, pp. 436-441.

[3] 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), 2021: IEEE, pp. 832-837.

[4] I. D. Raji and G. Fried, "About face: A survey of facial recognition evaluation," *arXiv preprint arXiv:2102.00813,* 2021.

[5] H. Lin, R. Tse, S.-K. Tang, Y. Chen, W. Ke, and G. Pau, "Near-realtime face mask wearing recognition based on deep learning," in *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, 2021: IEEE, pp. 1-7.

[6] N. Petrović and Đ. Kocić, "IoT-based system for COVID-19 indoor safety monitoring," *IcETRAN Belgrade,* 2020.

[7] O. P. Gupta, A. P. Agarwal, and O. Pal, "A study on Evolution of Facial Recognition Technology," in *2023 International Conference on Disruptive Technologies (ICDT)*, 2023: IEEE, pp. 769-775.

[8] P. Zhang, F. Huang, D. Wu, B. Yang, Z. Yang, and L. Tan, "Device-Edge-Cloud Collaborative Acceleration Method Towards Occluded Face Recognition in High-Traffic Areas," *IEEE Transactions on Multimedia,* 2023.

[9] X. Yang, M. Han, Y. Luo, H. Hu, and Y. Wen, "Two-Stream Prototype Learning Network for Few-Shot Face Recognition Under Occlusions," *IEEE Transactions on Multimedia*, vol. 25, pp. 1555-1563, 2023.

[10] P. Terhörst, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, and A. Kuijper, "Pixel-level face image quality assessment for explainable face recognition," *IEEE Transactions on Biometrics, Behavior, and Identity Science,* 2023.

[11] A. Siddiqua, A. Ramachandra, K. Pavan, N. Aparna, and K. Bhaumik, "Real Time Face Mask Detection and Monitoring System (RFMDM)," in *2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2023: IEEE, pp. 1-6.

[12] S. A. Sari and W. F. Al Maki, "Masked Face Images Based Gender Classification using Hybrid Bat Algorithm Optimized Bagging," in *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2023: IEEE, pp. 091-096.

[13] M. Pudyel and M. Atay, "An Exploratory Study of Masked Face Recognition with Machine Learning Algorithms," in *SoutheastCon 2023*, 2023: IEEE, pp. 877-882.

[14] C.-Y. Lin, F.-J. Chen, H.-F. Ng, and W.-Y. Lin, "Invisible Adversarial Attacks on Deep Learning-based Face Recognition Models," *IEEE Access,* 2023.

[15] M. Dosi *et al.*, "Seg-DGDNet: Segmentation based Disguise Guided Dropout Network for Low Resolution Face Recognition," *IEEE Journal of Selected Topics in Signal Processing,* 2023.

[16] L. Cimmino, D. Freire-Obregón, and A. F. Abate, "Image Similarity between Masked and Unmasked Face for Consumer Electronics Applications," in *2023 IEEE International Conference on Consumer Electronics (ICCE)*, 2023: IEEE, pp. 1-2.

[17] S. Banerjee, W. Scheirer, K. Bowyer, and P. Flynn, "Analyzing the Impact of Shape & Context on the Face Recognition Performance of Deep Networks," in *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, 2023: IEEE, pp. 1-8.

[18] M. Dalkıran. "Real World Fasked Face Recognition Dataset (RMFRD)." https://www.kaggle.com/datasets/muhammeddalkran/masked-facerecognition (accessed.

[19] Z. Wang *et al.*, "Masked face recognition dataset and application," *arXiv preprint arXiv:2003.09093,* 2020.

[20] Z. Wang, B. Huang, G. Wang, P. Yi, and K. Jiang, "Masked face recognition dataset and application," *IEEE Transactions on Biometrics, Behavior, and Identity Science,* 2023.

[21] A. N. Zereen, S. Corraya, M. N. Dailey, and M. Ekpanyapong, "Two-stage facial mask detection model for indoor environments," in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020*, 2021: Springer, pp. 591-601.

[22] T. Fang, X. Huang, and J. Saniie, "Design flow for real-time face mask detection using PYNQ system-on-chip platform," in *2021 IEEE International Conference on Electro Information Technology (EIT)*, 2021: IEEE, pp. 1-5.

[23] F. Mercaldo and A. Santone, "Transfer learning for mobile real-time face mask detection and localization," *Journal of the American Medical Informatics Association,* vol. 28, no. 7, pp. 1548-1554, 2021.

[24] S. S. Tomkins and R. McCarter, "What and where are the primary affects? Some evidence for a theory," *Perceptual and motor skills,* vol. 18, no. 1, pp. 119-158, 1964.

[25] S. R. Rudraraju, N. K. Suryadevara, and A. Negi, "Face mask detection at the fog computing gateway," in *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, 2020: IEEE, pp. 521-524.

[26] V. Bruce and A. Young, "Understanding face recognition," *British journal of psychology,* vol. 77, no. 3, pp. 305-327, 1986.

[27] 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,* vol. 13, no. 12, p. 6900, 2021.

[28] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, no. 10, pp. 1090-1104, 2000.

[29] M.-H. Yang, "Face recognition using kernel methods," *Advances in Neural Information Processing Systems*, vol. 14, 2001.

[30] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data—with application to face recognition," *Pattern recognition,* vol. 34, no. 10, pp. 2067-2070, 2001.

[31] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Transactions on Neural networks,* vol. 14, no. 1, pp. 195-200, 2003.

[32] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[33] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: a large video database for human motion recognition," in *2011 International conference on computer vision*, 2011: IEEE, pp. 2556-2563.

[34] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Analysis and Modeling of Faces and Gestures: Third International Workshop, AMFG 2007 Rio de Janeiro, Brazil, October 20, 2007 Proceedings 3*, 2007: Springer, pp. 168-182.

[35] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE transactions on image processing,* vol. 19, no. 6, pp. 1635-1650, 2010.

[36] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873,* 2015.

[37] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015.

[38] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security,* vol. 13, no. 11, pp. 2884-2896, 2018.

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

[40] J. Cui, H. Zhang, H. Han, S. Shan, and X. Chen, "Improving 2D face recognition via discriminative face depth estimation," in *2018 International Conference on Biometrics (ICB)*, 2018: IEEE, pp. 140-147.

[41] A. Hasnat, J. Bohné, J. Milgram, S. Gentric, and L. Chen, "Deepvisage: Making face recognition simple yet with powerful generalization skills," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 1682-1691.

[42] C. Han, S. Shan, M. Kan, S. Wu, and X. Chen, "Face recognition with contrastive convolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 118-134.

[43] H. W. F. Yeung, J. Li, and Y. Y. Chung, "Improved performance of face recognition using CNN with constrained triplet loss layer," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017: IEEE, pp. 1948-1955.

[44] F. Ye, M. Ding, E. Gong, X. Zhao, and L. Hang, "Tiny face detection based on deep learning," in *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2019: IEEE, pp. 407-412.

[45] I. B. Venkateswarlu, J. Kakarla, and S. Prakash, "Face mask detection using mobilenet and global pooling block," in *2020 IEEE 4th conference on information & communication technology (CICT)*, 2020: IEEE, pp. 1-5.

[46] S. A. Sanjaya and S. A. Rakhmawan, "Face mask detection using MobileNetV2 in the era of COVID-19 pandemic," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, 2020: IEEE, pp. 1-5.

[47] A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," *arXiv preprint arXiv:2008.11104,* 2020.

[48] G. Jignesh Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Face mask detection using transfer learning of inceptionv3," in *Big Data Analytics: 8th International Conference, BDA 2020, Sonepat, India, December 15–18, 2020, Proceedings 8*, 2020: Springer, pp. 81-90.

[49] S. V. Militante and N. V. Dionisio, "Real-time facemask recognition with alarm system using deep learning," in *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, 2020: IEEE, pp. 106-110.

[50] A. Chavda, J. Dsouza, S. Badgujar, and A. Damani, "Multi-stage CNN architecture for face mask detection," in *2021 6th International Conference for Convergence in Technology (i2ct)*, 2021: IEEE, pp. 1-8.

[51] A. S. Joshi, S. S. Joshi, G. Kanahasabai, R. Kapil, and S. Gupta, "Deep learning framework to detect face masks from video footage," in *2020 12th international conference on computational intelligence and communication networks (CICN)*, 2020: IEEE, pp. 435-440.

[52] N. U. Din, K. Javed, S. Bae, and J. Yi, "A novel GAN-based network for unmasking of masked face," *IEEE Access,* vol. 8, pp. 44276-44287, 2020.