Masked Face Detection and Recognition System Based on Deep Learning Algorithms

Hayat Al-Dmour [1, *], Afaf Tareef [1], Asma Musabah Alkalbani [2], Awni Hammouri [1], and Ban Alrahmani [1]

[1] Faculty of Information Technology, Mutah University, Mu'tah, Al Karak, Jordan;
Email: a.tareef@mutah.edu.jo (A.T.), hammouri@mutah.edu.jo (A.H.), banalrhmani7@gmail.com (B.A.)
[2] Department of Information Technology, University of Technology and Applied Sciences, CAS IBRI, Muscat 516;
Email: asmam.ibr@cas.edu.om (A.M.A.)
*Correspondence: Hdmour@mutah.edu.jo (H.A.)

***Abstract*—Coronavirus (COVID-19) pandemic and its several variants have developed new habits in our daily lives. For instance, people have begun covering their faces in public areas and tight quarters to restrict the spread of the disease. However, the usage of face masks has hampered the ability of facial recognition systems to determine people's identities for registration authentication and dependability purpose. This study proposes a new deep-learning-based system for detecting and recognizing masked faces and determining the identity and whether the face is properly masked or not using several face image datasets. The proposed system was trained using a Convolutional Neural Network (CNN) with cross-validation and early stopping. First, a binary classification model was trained to discriminate between masked and unmasked faces, with the top model achieving a 99.77% accuracy. Then, a multi-class model was trained to classify the masked face images into three labels, i.e., correctly, incorrectly, and non-masked faces. The proposed model has achieved a high accuracy of 99.5%. Finally, the system recognizes the person's identity with an average accuracy of 97.98%. The visual assessment has proved that the proposed system succeeds in locating and matching faces.**

***Keywords*—COVID-19, facemask detection, face recognition, AI, deep learning, Convolutional Neural Network (CNN)**

I. Introduction

Scientists have yet to discover a treatment to combat Coronavirus (COVID-19); therefore, it is necessary to follow the imposed prevention instructions, which include social distancing, wearing masks and gloves, and consistently sterilizing with medical alcohol [1, 2]. Wearing a face mask is one of the essential recommendations made by the World Health Organization and imposed by governments on their citizens. Masks conceal a significant section of the face; hence, wearing masks may have a significant influence on identifying human identity. This issue has created concerns in many industries and areas that need a person's identification, such as the authentication required to unlock a phone or attendance registration in businesses and numerous public security systems, such as airports and railway stations [1]. Using standard means of unlocking, such as passwords and fingerprints, has the potential to propagate the infection. As a result, facial recognition is the optimum method for unlocking or authenticating security systems to limit the danger of transmission [3].

Scientific attempts to address and respond to the epidemic are increasing. To combat this pandemic, researchers are turning to artificial intelligence (AI), particularly deep learning, which has already shown improved performance in medical applications and can find patterns in big complicated datasets. More specifically, deep learning has lately made substantial advances in various domains, including object finding, segmentation, picture categorization, detection, and identification. As a result, it was important to use these powers to combat the epidemic [4]. During the COVID-19 pandemic, AI techniques assisted in predicting the number of illnesses or the spread of the virus, resulting in warnings and suitable preventative actions. Deep Neural Network (DNN) techniques give several models that aid coronavirus image processing, detection, and classification. Convolutional Neural Network (CNN) is one of the most significant networks [5–7]. There is also the convolutional graph networks model and the most recent models that have demonstrated their efficacy.

The CNN process is a method of tracking information and data that uses the natural sensory systems function. For instance, the brain has multiple interconnected preparation segments, such as when neurons collaborate to explain specific tasks. It is not subject to task-specific rules. In this work, CNN technology was used to learn facial features. Deep neural networks must be used to learn facial features from images. Consequently, a CNN is a deep learning system specializing in image processing. Each image comprises a chain of pixels, each with a value that may be used to identify the color and brightness level of the pixel as the images are represented as graphical data [8].

In facial recognition systems, a person's identity is identified by a unique code obtained from numerous places on the face, including the nose, chin, lips, eyes, and jaw. When a person wears a mask, several of these essential

Manuscript received September 21, 2022; revised November 9, 2022; accepted December 7, 2022; published March 14, 2023.

areas become obscured, making identification difficult owing to a lack of face information. This study presents a deep learning-based approach for detecting face masks. This study suggests a method for distinguishing persons who wear masks, do not wear masks, or who have their faces covered by something other than a mask. The suggested technology will be used to address an issue induced by the coronavirus epidemic, namely, face masks that conceal people's identities.

This study recommends employing feature selection and deep learning for masked face identification. The key contributions of this study include increasing detection confidence and establishing the accuracy score for the three main techniques for recognizing face masks using the most current advances in deep learning. Furthermore, the model employs multi-classification and improves mask face detection by training the model on various covers that can conceal the faces.

The rest of the paper is organized as follows: Section II presents an extensive literature review of neural networks, deep learning, and face mask detection methods. The proposed method for face-mask detection using multi-class classification is described in Section III. Experimental results to evaluate the effectiveness of the proposed method are shown in Section IV. Finally, the conclusion is given in Section V.

II. Literature Review

A recap of the preceding ways is offered at the conclusion of each approach, along with our recommended method approach for each problem and how the proposed method solves the difficulties. In 2000, Luo and Eleftheriadis [9] architected an algorithm that was capable to detect faces via a texture-based system. This system was unique owing to its process of compressing the domain with the help of existing works that were designed in the pixel domain. However, this research figured out problems encountered in the compressed DCT domains. Feature vector selection, preprocessing design, block quantization problems, and multi-model structure systems were a few of these problems. This research also figured out that in the block quantization problem, face detection was followed by a shorter feature vector. Therefore, a texture-based algorithm was not suitable for the pixel domain, and a combined texture-based system was found to be more useful that was built-in with the feature of face color detection. Hjelmås and Low [10] asserted that the human face is more prone to higher variability which becomes difficult for computers to detect it. Since a plethora of varieties of detection systems has been proposed by numerous studies, Hjelmås and Low [10] proposed an image-based algorithm that was more technical. Another study by Zhang and Zhang [11] identified that digital cameras are equipped with built-in face detection systems that may auto-focus and auto-exposure. However, this study has recommended using digital photo management software such as Windows Live Photo Gallery, Apple's iPhone, and Google's Picasa to detect and tag more people in the images. Similarly, Viola and Jones [12] developed a system of face detection that was to be installed on the desktop and was able to detect 15 frames in a single second aiming to decrease computational time with more detection accuracy. Notably, this proposed system was more efficient than its counterparts. In addition to it, Yang and Luo *et al.* [13] also proposed the WIDER FACE dataset which was 10 times larger and more efficient than other comparative datasets. This system was featured with rich annotations such as face bounding boxes, poses, occlusions, and categories.

Moreover, Lin and Zhao *et al.* [14] suggested a face recognition system based on deep learning and quantization approaches; they retrieved features using a CNN algorithm, then quantized the feature maps using the Bag-of-Features paradigm. Finally, in the classification step, the Multilayer Perceptron (MLP) technique is applied. The results demonstrate a high level of recognition accuracy [14]. Lin and Zhao *et al.* [14] described a method for recreating 3D face shapes with high-fidelity textures without the need to capture a large face texture library from single-view photos. The main idea is to use a 3D Morphable Model (3DMM) approach to finalize the first texture built using face information from the input image. In qualitative and quantitative comparisons, the results demonstrated that the strategy outperformed cutting-edge methods [5].

Mohamed and Mohamed *et al.* suggested a hybrid system that identifies the facemask by using machine learning and deep learning approaches. The system is divided into two parts: the first collects feature from the three datasets used: RMFD, SMFD, and LFW via Resnet50, and the second classifies the facemask using Support Vector Machines (SVM), decision trees, and the ensemble technique. The outcome demonstrates that the SVM model attained a high accuracy rate of 100% [6]. Loey and Manogaran *et al.* [7] proposed deep learning models, such as the Inception-v3 CNN, that used deep learning architectures with training parameters, like inception-v4, Mask R-CNN, Faster R-CNN, YOLOv3, Xception, and DenseNet, to recognize face masks with 99.9% accuracy [7].

Nagrath and Jain *et al.* [15] created three types of data sets: the Masked Face Detection Dataset, the Masked Face Detection Dataset (MFDD), and the Real-world Masked Face Recognition Collection (RMFRD), the biggest real-time dataset of masked faces. They built a face-eye-based multi-granularity model for face identification using the Simulated Masked Face Recognition Dataset (SMFRD). They were 95% accurate. In the classification step, they employed the VGG16 CNN model. If a person does not wear a face mask, the Raspberry Pi-based system raises an alarm; the findings reveal that the system has an accuracy of up to 96% [15]. Bartlett and Littlewort *et al.* [16] also developed a system of frontal face detection from videos that were equipped to detect fear, sadness, joy, disgust, neutrality, surprise, and anger. This study adopted a novel approach of combining the features of Adaboost and SVM to ameliorate the results. This system was used on several platforms such as ATR's RoboVie, Sony's Aibo pet robot, and CU animator. Yang and Jiachun *et al.* [17] employed a YOLO system that was equipped with fast and target

detection features that were built for a real-time working environment. The study has reported the efficiency and robustness of the YOLO system. Voulodimos and Doulamis *et al.* [18] also conducted a review study in order to investigate the levels of efficiency of different deep learning models that had been adopted by different studies for solving the issues of computer vision. These different techniques included Deep Boltzmann Machines and Deep Belief Networks, Convolutional Neural Networks, and Stacked Denoising Autoencoders. According to the findings of this study, Deep Boltzmann Machines and Deep Belief Networks brought about an increase in performance in terms of object verification, estimation of human pose, the retrieval of images, recognition of images, and semantic segmentation. Also, Guo and Liu *et al.* [19] in their review study, characterized different schemes of deep learning into four classes namely Autoencoder and, Convolutional Neural Networks, Restricted Boltzmann Machines, and Sparse Coding. The results of this review study endorsed the vitality of these four schemes; however, it emphasized to development of further schemes and models for increasing the accuracy of CNN-based algorithms with reference to human raters. Din et al. developed a method that consists of two stages: detecting the facemask and removing it with the GAN. The first step consists of a map module that recognizes the person wearing the mask, and the second stage consists of an editing module that works to finish the face when the mask is removed. The technology was quite efficient in detecting and completing faces [20, 21]. Sandesara and Vijh *et al.* [22] created a face mask identification framework based mostly on deep learning computer vision techniques, with the CNN algorithm utilized to distinguish those who wear masks from those who do not. The technology was built into a camera to follow people's movements in real-time. The technique detected disguised faces with 96% accuracy [22].

EyesGAN was a face recognition system presented by Mata et al. based on the composition of people's faces from their eyes. By using the perceptual loss and self-attentional mechanisms in GANs, the system attained a stunning accuracy rate of 96.10% when wearing a face mask [22]. Qi and Jia *et al.* [23] created a deep facial clustering method based on a residual graph Convolutional Network (RCNN) with additional hidden layers. The k-Nearest Neighbor (kNN) technique is used to create subgraphs for each node. The ResNet idea was then applied to CNNs, and RCNN was created to learn how to connect two nodes. The suggested technique is more accurate and produces better clustering results than other current approaches to facial clustering. Furthermore, the suggested RCNN clustering approach discovers clusters automatically and scales to huge datasets [23].

Aswal and Tupe *et al.* [24] considered a single-step pre-trained YOLO-face/trained YOLOv3 model on a set of known individuals, and a two-step method based on a pre-trained one-stage feature pyramid detector network RetinaFace to propose proposed a single camera masked face detection and identification method based on two approaches. This proposition was for localizing masked faces and VGGFace2 for generating facial feature features for an efficient mask. In trials, RetinaFace and VGGFace2 achieve state-of-the-art results of 92.7% overall performance and 98.1% face detection [25].

The preceding techniques rely on a pre-trained model that has been trained on human faces. This pre-trained model allows the classifiers to recognize human faces and their characteristics without requiring the classifiers to extract the real aspects of the human face. The technique suggested in the present study will train and evaluate human face traits, as well as extract faces with varying angles and postures. Furthermore, the model will be able to recognize several faces in a single picture. Following an examination of previous researchers' methodologies and gaps, the face detection models were constructed by modifying photos and applying masks to original images that already recognized the model's properties.

III. Methodology

The proposed approach comprises three phases. The first phase is preprocessing phase to prepare the dataset and detect the region of interest (ROI), i.e., face area. This research used the fine-tunes Residual Neural Network for constructing the model and trained the model to extract the Region of Interest (ROI) from each image present in the dataset. The second step divides the detected faces into three categories: unmasked faces, masked faces, and wrongly masked faces. The final phase detects the person's identity to allow authentication.

At the sub-stage of this study, a Deep Learning model network will be trained and examined for detecting a face. This training database will be employed for the training network. The proposed approach is illustrated in Fig. 1.

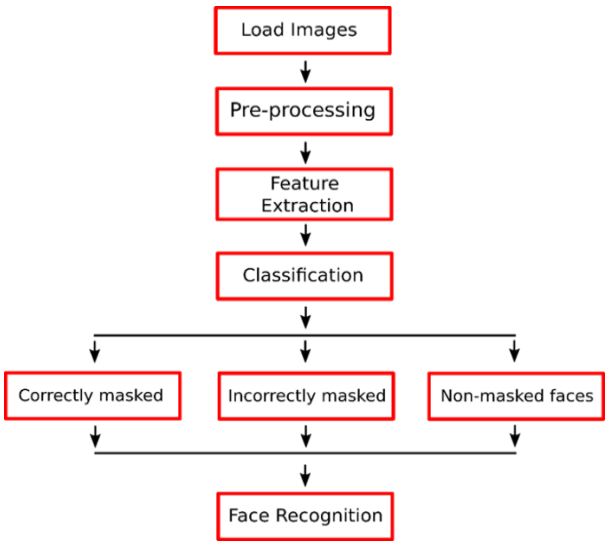

Figure 1. The proposed model.

A. Pre-processing

In this phase, data preparation is performed by resizing the image to a reasonable size of 224×224 pixels as the images in the training dataset were of varying sizes. Then, it is followed by data augmentation which involves resizing and cropping input images and altering them, such

as by shifting, flipping, and changing the color (see Fig. 2 (a)). Data augmentation is a commonly used technique for improving the generalizability of an over-fitted model. Providing more training data and exposing the model to different types of data within a class will make the training process more robust and increase its chances of generalization. The idea behind data augmentation is that a real-world dataset can consist of only a few images taken under limited conditions. Nevertheless, our target application may exist in various conditions, including different orientations, locations, scales, and brightness levels. Therefore, these situations are taken into account by training our model with synthetically modified data such as horizontal and vertical reflection, slight rotation or magnification, and color inversion. Training of the data comprised numerous changes such as using different operations of image manipulation i.e., flips, zooms, shifts, and mean subtraction. Finally, the region of interest is detected using Haar feature-based cascade classifiers by cropping the square in the middle to eliminate the face background effect as shown in Fig. 2 (b). This defined region will categorize and comprehend the facial state with a mask, without a mask, or inadequately masked. Also, in the facial recognition section, to recognize the individual.

Figure 2. Data preprocessing phase: (a) Data augmentation, (b) (Region of interest) ROI detection.

B. Masked Face Classification

The second phase is masked face classification. The model is trained with different datasets to classify face images into three categories: mask, covered face, and without a mask. The model was evaluated in various settings with a variety of hyperparameters, and the results are detailed in the results section. The CNN network is used to identify face masks in Fig. 3.

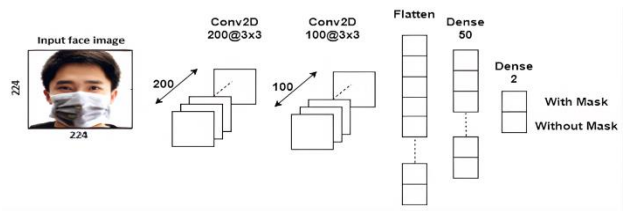

Figure 3. Face Mask Detection Using CNN network Architecture.

1) Feature extraction

Face detection relies on feature extraction to provide a discriminative collection from which to identify and learn the basic facial qualities. Our designed classification system uses a pre-trained VGG-16 extractor to get a discriminative feature vector from an image. VGG is a CNN model for image recognition proposed by the Visual Geometry Group at the University of Oxford. VGG-16 has 16 convolutional layers, Max Pooling, Activation, and Fully Connected (FC) layers. There are 13 convolutional layers, 5 Max Pooling layers, and 3 Dense layers totaling 21, yet only 16 are weight layers.

The conv1 layer receives a 224×224 RGB picture as input. The picture is processed through a stack of typical layers, with the filters set to capture left/right, up/down, and center with a minimal receptive field of 3×3 (the smallest size to catch left/right, up/down, and center). The 1×1 convolution filter is also employed in one of the settings, which may be thought of as a linear transformation of input channels followed by non-linearity. The standard space padding and the convolution stage are set to one pixel. Layer feedback is the spatial resolution kept during convolution; for 33 convolutions, the padding is one pixel. Five maximum pools are used for spatial pooling.

Three FC layers employ a stack of convolution layers (changing in depth depending on architecture): The first and second layers each have 4096 channels. The third, on the other hand, has a 1000-way ILSVRC grouping and 1000 channels (one for each class). The soft-max layer is the last layer. For all networks, the eventually connected layers are created similarly. Both hidden layers are subjected to non-linearity correction (ReLU). Except for one, the networks do not apply Local Response Normalization (LRN), which does not increase ILSVRC data collection efficiency but does lead to higher memory and computational use. This study used extremely efficient detectors and features, the computational cost of the detector at every location and scale is faster. This study chose the 512-d feature vectors for better results.

2) CNN-based classification

For our face detection, a CNN classifier is utilized, as shown in Fig. 4. This strategy reduces the number of network parameters and speeds up the picture template search. The first three layers of the CNN were connected in parallel to the first three layers of the second CNN, yielding a seven-layer two-stream CNN capable of recognizing faces and facial characteristics in a single forward pass.

Once the combined model has been trained end-to-end using RGB pictures of size 32×32, the gradients are transmitted to both streams of the combined network. One-quarter of the 32×32 training face pictures are 16×16 images of facial components. The idea behind this method is that the data generated by local face component detection is critical for detecting face regions and should be integrated into the detection process from the beginning. 104.280 free parameters must be optimized to train this network.

Once the training process is completed, we run the network on the set of images from which we collect a subset of false positives F1, which is added to the original set of negative examples N0. The set F1 is chosen based on the network output. We filter the false positives based on their score during each training cycle and choose a specified number of samples. We increase the number of

positive instances after each training cycle to maintain the same ratio of positive to negative samples. The CNNs were trained using Stochastic Gradient Descent (SGD) with a learning rate of 0.001 for the first 200.000 iterations, and then we reduced it to 0.0001.

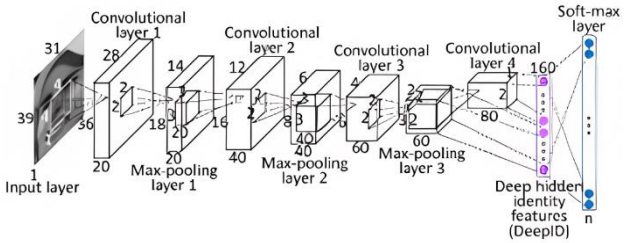

Figure 4. Face detection using CNN network.

C. Face Mask Recognition

During community closure, masked face identification algorithms are used to identify and recognize individuals. Face recognition gates, facial attendance machines, and facial security checks at railway stations have also been upgraded to accommodate people wearing masks.

In the last step, the system is responsible for distinguishing between the identities of different people's faces. In order to achieve an automatic recognition process, a face database can be built by taking several images of each individual's face, and then the features of these images are extracted and stored in the system database. Then, the system performs face detection and feature extraction whenever an input image appears. Later, the system compares the image features to each stored face print in the database to either grant or deny access.

Face recognition is divided into two different applications: identification and verification. As part of the face identification process, the system identifies an image for a particular user. On the other hand, the system asks whether an identified image is true or false for face verification.

For masked face recognition, the VGG-16 feature extractor and CNN have performed again on the defined face area, i.e., ROI, where the target here is the person's identity. The CNN network will be constructed based on the dataset features passed by the VGG-16 using a cross-validation approach with a k-fold value of 10. The neurons and layers will vary dependent on the length of the feature vectors. The layers are based on the extract percentage of the features, and neurons are based on the formula (2*length of features+ 1) to obtain the best performance of CNN. Still, all datasets will have one Flattened layer and two Dense layers.

The classification is performed based on mathematics. To calculate the layers' number, we used L = K×((W−F+2P)/S+1) L = INT (0.001×((88,592 – 48384) /3)) = 13, where L is the number of layers, K is a constant number, W is weighted, F is features, P is padding and S is Stride. The parameters were assigned as follows: width 98, height 113, channels 8, Filter count: 18, spatial extent = 3, Stride =2, and padding=0. The input vector layer is based on the number of the selected features, and it was 88,592 (98×113×8) in this present case. The model will start to classify the features and decrease the number of features considering the weights and most affected features. The data will be filtered using 16 pooling kernel filter matrices to have 48384 (48×56×18) features from the 88,592 features from the first input layer.

The next layer will calculate the best features after decreasing data by using 16 pooling kernel filter matrices to reduce the features to 5460 (26×30×7). The next layer is more selective on features, selecting the most critical features that lead to the type of traffic and decreasing the data to almost 630 (9×10×7) features. In the following two layers, the classification deals with the FP and TN to increase the accuracy and minimize the error at the end 252 (6×6×7) features.

The flattened Layer and dense Layer make the type of traffic to provide the best prediction for the data type, determines the traffic classes, and give results by showing the confusion matrix and classification report. The dense layer is configured as 50, dropout 0.25, 30; dropout 0.25 then ClassNum size with no dropout.

IV. Result and Discussion

A. Dataset Description

The datasets used in the proposed method are as the following (Table I):

TABLE I. Dataset Descriptions

Dataset	Images	Format	Faces
RMFD	4,400	jpg/png	Single
CFR	2,549	jpg	Single
Masked-FaceNet	35,511	jpg/png	Single

(1) RMFD *(https://github.com/X-zhangyang):* There are around 2500 images of 460 people wearing masks in the collection and around 90,000 images of the same 460 people without masks. This is a large real-world masked/non-masked face dataset. However, 4000 images (2000 masked ones and 2000 non-masked ones) were selected for training to have balanced datasets. Afterward, 400 images (200 masked ones and 200 non-masked ones) were selected for testing the trained model.

(2) Masked-FaceNet *(https://github.com/cabani/MaskedFace-Net)*: The study used 10,000 images for training for each of the classes of correctly masked and incorrectly masked faces. As mentioned above, an extra 10,000 images were collected from the RMFD database for the non-masked faces class. As a result, 30,000 images were utilized for training across all three classes. In addition, 5511 images from all three classes were included in the testing.

(3) Celebrities Face Recognition (CFR) *(https://www.kaggle.com/vasukipatel/face-recognition-dataset):* This face recognition dataset includes 31 well-known Hollywood stars. Each celebrity has more than 50 images, for a total of 2549. Face traits and images from various

perspectives are included in the collection. Any faulty images were removed from the dataset. The mask faces images were tested after training the model on the dataset.

As the designed model is applied on streams of images containing different persons with variable degrees of masking, it is important to correctly identify the ROI from these images before feeding them into the trained model for prediction. The visual results recognize the faces and display findings for each face in the image.

The dataset was employed in this study utilizing a cross-validation strategy. The k-fold value was set to 10. One rationale was that the dataset had a restricted amount of images, and cross-validation is a useful strategy in such a situation.

B. Evaluation Metrics

The accuracy of the classifiers with each feature extraction approach was measured using the metrics described below. Along with the description of recall, precision, and accuracies, several terminologies are widely employed, such as true positive (TP), true negative (TN), false negative (FN), and false-positive (FP). If a patient has the syndrome, the test also shows that the illness exists, and the diagnostic test findings are true and positive. Similarly, if a patient lacks the condition, the diagnostic evaluation reveals that the disease does not exist (TN). Positive and negative outcomes will arise from a good result between the diagnostic test and the established condition (the standard of truth). When the test reveals that the patient is healthy, the findings indicate the presence of disease (FP). The diagnostic test result is erroneous if it suggests that a patient with the condition is not present for certain (FN). The findings of the tests are contradictory to the actual conditions: false positives and false negatives. The confusion matrix t gives an output matrix representing the full model performance. There are several measurements were used for performance evaluation:

- Precision = TP/ (TP + FP)
- Recall = TP/ (TP + FN)
- Accuracy = (TN + TP)/(TN+TP+FN+FP)
- F1 Score= 2×((precision × recall)/ (precision + recall)).

C. Experimental Results

1) Masked and non-masked face classification

Using the RMFD database, the binary model was utilized to recognize masked and non-masked faces. As a baseline model, the model was successful at recognizing masked and non-masked faces. Cross-validation with a k=10 value was employed. While tracking the validation accuracy, an early stop with a patience value of 5 was also utilized to prevent the training procedure from overfitting. 3600 images were used for training and 400 for validation with this value of cross-validation sets. The sets changed as a result of the cross-validation procedure. The classification results using the binary model are shown in Table II. The model was run several times more, with no discernible difference in accuracy.

Table II demonstrates that the model can recognize masked and non-masked faces with a mean validation accuracy of 98.68% and a standard deviation of 0.423.

TABLE II. CNN CLASSIFICATION CROSS-VALIDATION ITERATIONS (BINARY MODEL)

Iterations for k-fold=10	Validation Accuracy (%)
1	98.96
2	99.22
3	99.22
4	98.96
5	98.18
6	98.44
7	98.44
8	98.96
9	98.22
10	98.22
Mean with a standard deviation	98.68 (+- 0.423)
Final Run	99.77

For testing, 400 images were used, as indicated in the RMFD dataset description (200 images for each class). The confusion matrix for the binary class for test data is shown in Fig. 5. The first class (masked face) was predicted correctly 194 times and incorrectly six times. Similarly, the second class (non-masked) was properly predicted 199 times and incorrectly predicted one time. Fig. 6 depicts the variance inaccuracies in the k-fold runs and Fig. 7 depicts the final model training accuracies.

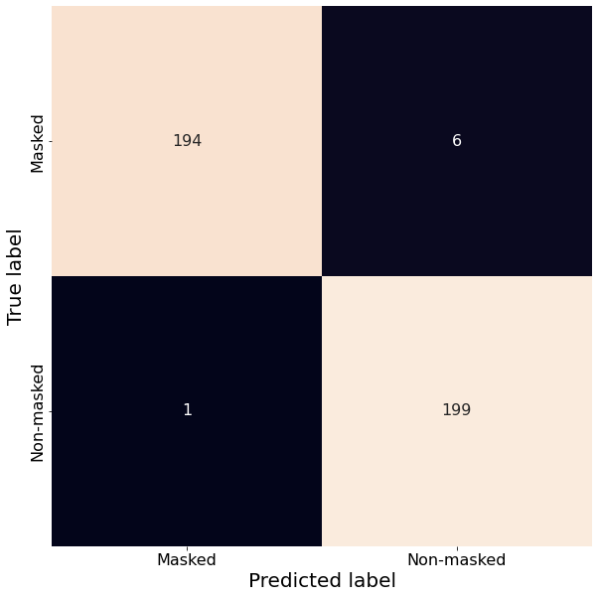

Figure 5. CNN classifier confusion matrix for test data (binary model).

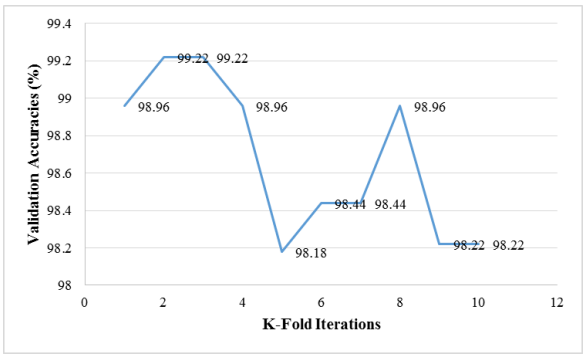

Figure 6. Classification results of cross-validation with validation accuracies.

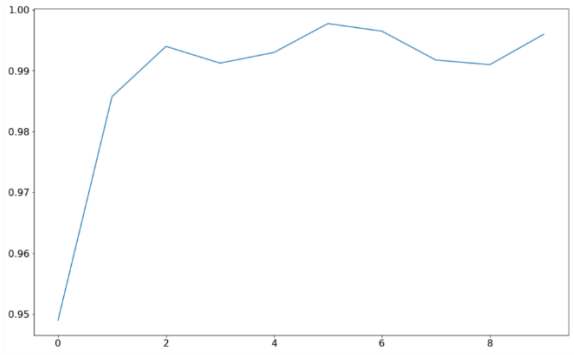

Figure 7. Training accuracies for the final model run for binary classification.

2) Multi-class results (correctly masked, incorrectly masked, and non-masked faces)

The model's findings are displayed in Table III, using the Masked-FaceNet dataset and three groups of correctly masked (CMFD), incorrectly masked (IMFD), and non-masked faces. According to the workings of cross-validation, the images were shuffled as validation sets in training sets in each iteration. The study utilized cross-validation with a k-fold value of 10. Early halting was implemented to track validation accuracy and avoid model overfitting. Using ten divisions of the dataset for training, the entire set of 30,000 images was separated into 27,000 images for training and 3000 images for validation.

TABLE III. CNN CLASSIFICATION CROSS-VALIDATION ITERATIONS (MULTI-CLASS MODEL)

Iterations for k-fold=10	Validation Accuracy (%)
1	99.63
2	99.6
3	99.76
4	99.6
5	99.7
6	99.4
7	99.7
8	99.66
9	99.73
10	99.76
Mean with a standard deviation	99.65 (+- 0.1)
Final Run	99.5

A total of 5511 images from the three classes were utilized for testing, and the confusion matrix indicated a very good accuracy as shown in Fig. 8. Table IV shows the accuracy, recall, and F1 score values for the multi-class model's last run.

The performance of the proposed model is compared to that of other relevant techniques. Table V summarizes the experiment's findings and indicates that the suggested model outperforms previous techniques in terms of accuracy. Additionally, a multi-class classification was also implemented in the proposed model, whereas all other techniques were binary class classifications.

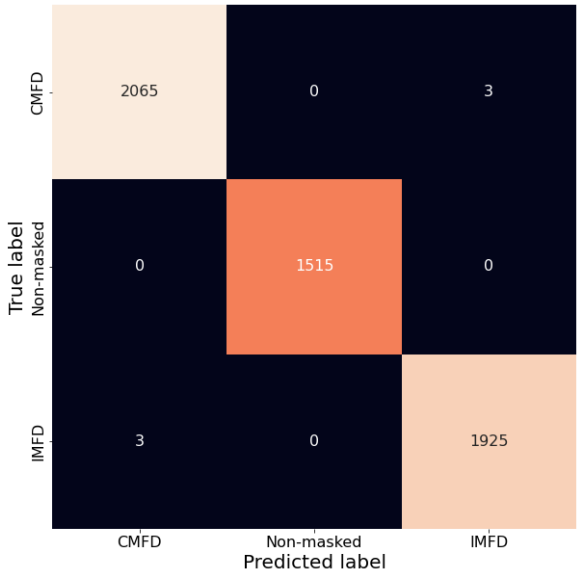

Figure 8. Confusion matrix for the multi-class model.

TABLE IV. PRECISION, RECALL, AND F1 SCORES FOR MULTI-CLASS FINAL MODEL EVALUATION

	Precision	Recall	F1-Score
CMFD	0.9985	0.9985	0.9985
IMFD	0.9984	0.9984	0.9984
Non-masked	1.0	1.0	1.0

TABLE V. FACE MASK DETECTION ACCURACY RESULTS

Paper title and the Authors	Main Technique	Single\ Hybrid	Data Set	Accuracy results
The Proposed method	CNN	Single	RMFD, CFR, Masked-FaceNet	**99.5**
Loey *et al.*, [7]	Machine Learning	Hybrid	Masked Face Dataset (SMFD, RMFD)	94.64
Nagrath *et al.*, [15]	DNN	Single	Masked Face Dataset (SMFD, RMFD)	92.64
Chawda *et al.*, [25]	CNN	Single	Masked Face Dataset (SMFD, RMFD)	93.05
Sandesara *et al.*, [22]	CNN	Single	RMFD (Real World Masked Face Dataset) and Kaggle Dataset	96

3) Face recognition results

The Celebrities Face Recognition (CFR) dataset is used for the face recognition evaluation. The Dataset has been cleared of any corrupted images. Mask faces images have been tested after the model has been trained on the Dataset.

The model understood the facial features and predicted the right person even with the masks on. The visual results show that the model was able to verify and recognize each face even with the masks on, the accuracy reached 98.83% due to the multiclass, and some photos are similar in some features in general. On the other hand, the model recognized each person's visual and facial characteristics. Fig. 9 shows a sample of our face recognition results.

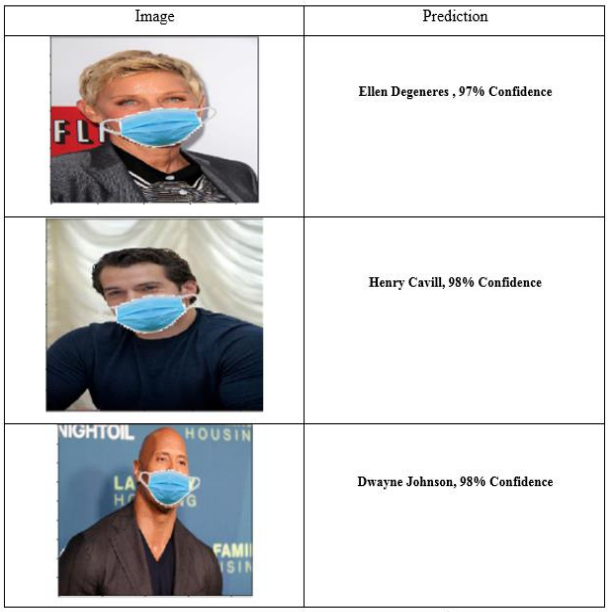

Figure 9. Sample of our face recognition results.

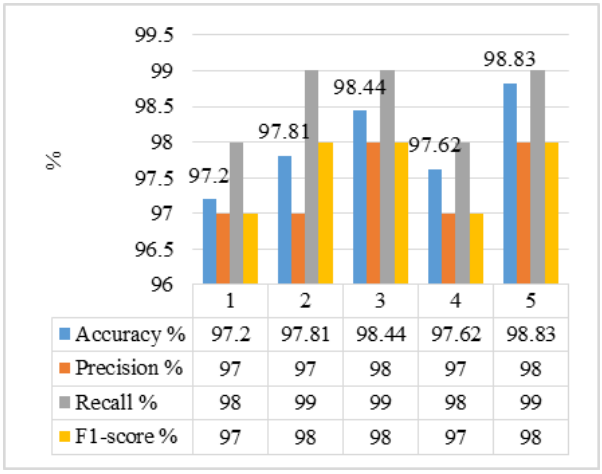

Figure 10. CNN proposed method results for face prediction.

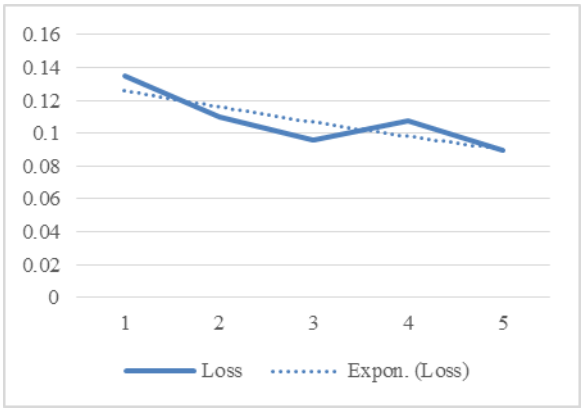

Figure 11. Proposed method loss results for face prediction as a function of five runs.

The face recognition phase in the proposed method yields a high performance with an average accuracy of 97.98%. Further results are provided in Fig. 10 and Fig. 11. As shown in the figure, the proposed method achieves high accuracy, precision, recall, and F1-score for five different images, whilst the loss was reduced as shown in Fig. 11.

V. Conclusion

Masked face recognition has become increasingly important because of the COVID-19 pandemic. Future studies need to involve sensing social separation and alerting if someone is not correctly wearing a face mask. This paper proposes a deep-learning-based method for accurate face mask detection and masked facial recognition. The test findings demonstrate a high level of accuracy in recognizing persons who are wearing, are not wearing, or are wearing a face mask improperly. The model achieved a performance accuracy of greater than 99% for masked face detection, and 97% for masked face recognition. Furthermore, the visual assessment was successful in locating and matching faces. However, a real-time detection model with substantial improvements in performance and run-time processes needs to be provided in future studies, emphasizing the FP and TN rates to reduce the recognition error for detecting and identifying a person's faces

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors contributed to the study's conception and design. Hayat Al-Dmour, Awni Hammouri, and Ban Alrahmani conducted the research by defining research frameworks and designing a research methodology. Afaf Tareef and Asma Musabah Alkalbani conducted and conceived the experiments and performed analysis. The initial draft of the manuscript was written by Hayat Al-Dmour and Afaf Tareef and all authors commented on previous versions of the manuscript. All authors discussed the results and approved the final manuscript.

References

[1] M. F. Arefi and M. Poursadeqiyan, "A review of studies on the COVID-19 epidemic crisis disease with a preventive approach," *IOS Press*, vol. 66, no. 4, 2020. doi: 10.3233/WOR-203218

[2] S. Mukherjee, S. Boral, H. Siddiqi, A. Mishra, and B. C. Meikap, "Present cum future of SARS-CoV-2 virus and its associated control of virus-laden air pollutants leading to potential environmental threat–a global review," *J. Environ. Chem. Eng*, vol. 9, no. 2, 104973, 2021.

[3] S. Sethi, M. Kathuria, and T. Kaushik, "Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread," *J. Biomed. Inform*, vol. 120, 103848, 2021.

[4] Y. Pan and L. Zhang, "Dual attention deep learning network for automatic steel surface defect segmentation," *Comput. Civ. Infrastruct. Eng*, 2021.

[5] J. Lin, Y. Yuan, T. Shao, and K. Zhou, "Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5891–5900, 2020.

[6] S. S. Mohamed, W. A. Mohamed, A. T. Khalil, and A. S. Mohra, "Deep learning face detection and recognition," *Int. J. Adv. Sci. Technol*, vol. 29, no. 2, pp. 1–6, 2020.

[7] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods

for face mask detection in the era of the COVID-19 pandemic," *Measurement*, vol. 167, 108288, 2021.
[8] Z. Kolar, H. Chen, and X. Luo, "Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images," *Autom. Constr*, vol. 89, pp. 58–70, 2018.
[9] H. Luo and A. Eleftheriadis, "On face detection in the compressed domain," in *Proc. the eighth ACM international conference on Multimedia*, pp. 285–294, 2000
[10] E. Hjelmås and B. K. Low, "Face detection: A survey," *Comput. Vis. Image Underst.*, vol. 83, no. 3, pp. 236–274, 2001.
[11] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Microsoft Publication, MSR-TR-2010-66, 2010.
[12] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis*, vol. 57, no. 2, pp.137–154, 2004.
[13] S. Yang, P. Luo, C.C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5525–5533.
[14] K. Lin, H. Zhao, J. Lv, C. Li, X. Liu, R. Chen, and R. Zhao, "Face detection and segmentation based on improved mask R-CNN," *Discret. Dyn. Nat. Soc*, vol. 2020, 2020.
[15] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "SSDMNV2: A real-time DNN-based face mask detection system using single shot multibox detector and MobileNetV2.," *Sustain. Cities Soc.*, vol. 66, 102692, 2021. doi: 10.1016/j.scs.2020.102692
[16] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction," in *Proc. Conference on Computer Vision and Pattern Recognition Workshop*, no. 5, pp. 53–53, 2003.
[17] W. Yang and Z. Jiachun, "Real-time face detection based on YOLO," in *Proc. 1st IEEE International Conference on Knowledge Innovation and Invention* (ICKII), 2018, pp. 221–224.
[18] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci*, 2018.
[19] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, and S. M. S. Wu, "Lew Deep learning for visual understanding: A review," *Neurocomputing*, vol. 26, no. 187, pp. 27–48, 2016
[20] 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.
[21] A. G. Sandesara, D. D. Joshi, and S. D. Joshi, "Facial mask detection using stacked CNN model," *Int. J. Sci. Res. Comput. Sci. Eng. Inform. Technol*, 2020.
[22] V. J. Srivastava and S. Vijh, "Face mask detection using convolutional neural network," in *Proc. Conflu. 2022 - 12th Int. Conf. Cloud Comput. Data Sci. Eng.*, vol. 21, no. 1, 2022, pp. 26–30. doi: 10.1109/Confluence52989.2022.9734156.
[23] C. Qi, J. Zhang, H. Jia, Q. Mao, L. Wang, and H. Song, "Deep face clustering using residual graph convolutional network," *Knowledge-Based Syst*, vol. 211, 106561, 2021.
[24] V. Aswal, O. Tupe, S. Shaikh, and N. N. Charniya, "Single camera masked face identification," in *Proc. 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 57–60, 2020.
[25] S. Chawda A. Patil, A. Singh, and A Save, "A novel approach for clickbait detection," in *Proc. 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, 2019, pp. 1318–1321.

Hayat Al-Dmour received the B.Sc. degree in computer science from Mutah University, Jordan in 2005, the M.S degree in computer science from Yarmouk University, Jordan in 2007, and a Ph.D. degree from the University of Technology Sydney, Australia in 2018. She is currently an assistant professor in the faculty of Information Technology at Mutah University, Jordan. She has many publications in several international conferences and journals. Her research interests over information hiding, image processing and machine learning, mainly for medical image segmentation.

Afaf Tareef received the B.Sc. degree in computer science from Mutah University, Jordan in 2008, M.Phil. degree from the University of Jordan in 2010, and a Ph.D. degree from the University of Sydney, Australia. She is currently an assistant professor in the faculty of Information Technology at Mutah University, Jordan. Her research interests include image processing and medical image analysis.

Asma Musabah Alkalbani received the B.S. degree in computer engineering from the Caledonian College of Engineering, Muscat, Oman, in 2004, the M.S. degree in information technology from La Trobe University, Melbourne, Australia, in 2010, and the Ph.D. degree in software engineering from the University of Technology Sydney (UTS), Sydney, NSW, Australia. From 2010 to 2014, she was a Lecturer with the College of Applied Sciences, Oman, and from 2014 to 2018, and also a Casual Academician with the Computer Science School, UTS, where she supervised 12 Master Graduation projects (data analytics and service discovery). Since 2018, she has been an Assistant Professor with the Information Technology Department, College of Applied Sciences, Oman. Her research interests include service discovery, web data mining, and business data analytics to better understanding business need especially the data analytics of biomedical data, text mining, and social network analysis.

Awni Hamouri is currently a professor of computer science at Mutah University, Jordan. Prof. Hamouri earned his PhD degree in computer science from Illinois University in USA, 1994. His research interest is AI, Computer Networks and Software Engineering.