

# A Framework for Mask-Wearing Recognition in Complex Scenes for Different Face Sizes

Hanan A. Hosni Mahmoud, Amal H. Alharbi and Norah S. Alghamdi\*

Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, 11047, KSA

\*Corresponding Author: Norah S. Alghamdi. Email: NOSAlghamdi@pnu.edu.sa

Received: 05 August 2021; Accepted: 06 September 2021

**Abstract:** People are required to wear masks in many countries, now a days with the Covid-19 pandemic. Automated mask detection is very crucial to help identify people who do not wear masks. Other important applications is for surveillance issues to be able to detect concealed faces that might be a safety threat. However, automated mask wearing detection might be difficult in complex scenes such as hospitals and shopping malls where many people are at present. In this paper, we present analysis of several detection techniques and their performances. We are facing different face sizes and orientation, therefore, we propose one technique to detect faces of different sizes and orientations. In this research, we propose a framework to incorporate two deep learning procedures to develop a technique for mask-wearing recognition especially in complex scenes and various resolution images. A regional convolutional neural network (R-CNN) is used to detect regions of faces, which is further enhanced by introducing a different size face detection even for smaller targets. We combined that by an algorithm that can detect faces even in low resolution images. We propose a mask-wearing detection algorithms in complex situations under different resolution and face sizes. We use a convolutional neural network (CNN) to detect the presence of the mask around the detected face. Experimental results prove our process enhances the precision and recall for the combined detection algorithm. The proposed technique achieves Precision of 94.5%, and is better than other techniques under comparison.

**Keywords:** Mask detection; deep learning; CNN; small faces; Covid-19

## 1 Introduction

Masks play an important role in defending people against Covid-19. Many countries forced laws to wear masks in public. Some people may not adhere to such protective laws or may forget to wear masks. Nevertheless, wearing a mask might lessen the risk of corona infection [1]. A report of the WHO confirmed over 120 million confirmed cases with more than 2 million death by February 2021. Many of these cases could be avoided by taking precaution steps such as wearing masks and social distancing. Respiratory drops from sneezing or talking loudly can transmit covid-19 as depicted in [2,3]. A proven way to reduce covid-19 spreading is to enforce mask wearing in public as advised by WHO [2].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we studied the micro-graphs of different types of masks by considering the texture features of the images of those masks. Texture features extraction can be performed by fractal analysis methods, Fourier transform techniques, also, gray co-occurrence matrix. Authors at [4], presented a study that indicated that gray co-occurrence matrix (GLCM) can quantify the spatial association among adjacent pixels in a digital image. GLCM is utilized in texture detection [5], skin texture identification [6], defect recognition [7], fabric cataloging [8] and egg fertility detection [9]. Authors in [10], categorized fabrics utilizing Support Vector Machine with GLCM and Component Analysis. The experimental results of [11] show that the technique of combining GLCM and Back propagation deliver 93% accuracy. Since GLCM is efficient in texture analysis, we decided to utilize GLCM for the texture analysis of masks. Also, K Nearest Neighbor technique is simple and possess high performance in classification process [12], it is successfully used in text medical domain [13], ingestion patterns examination [14]. The authors in [15], concluded that k-nearest neighbor's algorithm (KNN) variant technique could identify the Coronavirus patients more accurately. Therefore our research utilizes the KNN algorithm to detect the time-wearing of masks based on the texture features mined from the micro-images at different period time.

CNN is a neuron machine model that uses convolution construction [16]. LeNet was one of the first CNN to be proposed. Rectified Linear machines gave a boost to CNN by year 2011. Object detection methodologies are established using Region-based Convolutional Neural Networks (R-CNN) [17]. R-CNN is an upgraded process based on CNN. Authors in [18], projected an epoch R-CNN system that was utilized mainly for object detection. It utilized search selective technique to identify borders from the rest of the image and normalize it to the CNN input size, followed by object identification by the SVM using linear regression model. The drawbacks were complicated training phase and slow testing speed.

An enhancement of the training phase and the testing speed of R-CNN was performed by the proposed Fast R-CNN. Fast R-CNN utilizes less layers with pooling layer. It also used the Softmax classifier instead of the SVM for better and faster classification [19]. However, the utilized selective search technique used in the fast R-CNN makes the computation speed not suitable for large data sets. Faster R-CNN introduces the combined feature extraction and bounding box regression to speed up the process. Faster R-CNN executes well for large objects (faces), but cannot identify small objects or faces [20]. We have to utilize other techniques for smaller objects that can optimizes face detection with respect to scale invariance and image resolution. Scale invariance is a central part of all current object detection methods.

Observing mask-wearing manually in large crowds can deem impossible and costly, especially in the Pilgrimage season. Reducing manual observation while making certain that all people are wearing masks all the time is an urgent problem. Image analysis tools can reduce the labor force and material costs, and can considerably protect people in many areas. Computer vision procedures and can be utilized for mask detection. Deep neural networks also are utilized greatly in object classification and recognition [21].

There are very few computer vision research that was done for mask detection and is usually utilized in very simple scenes but not from surveillance images. For this reason we will survey other object detection techniques that can help us in our research. For object detection, researchers utilized a histogram of gradient and a SVM machine to detect persons and a Hough transform to detect helmets and located faces by background subtraction. They also used background modeling and face classification method called C4 to detect faces, and detect whether masks using color transformation and recognition [22]. But it was not suitable for complex scenes and energetic backgrounds, such as busy streets and Pilgrimage season. Authors in [23], utilized a shot object detector process to detect faces with RetinaNet with a multiple features to overcome the limitation in accuracy. They also utilized the (YOLO) algorithm to detect helmet wearing in images with less than four people.

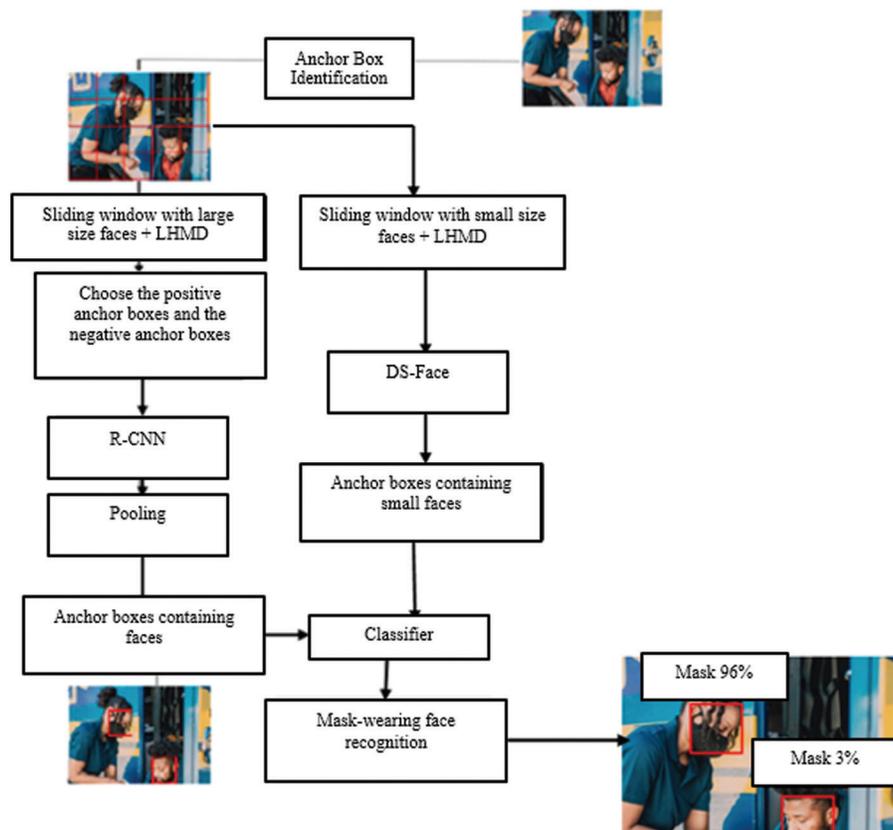
In this research, we propose a framework to incorporate two deep learning procedures to develop a technique for mask-wearing recognition especially in complex scenes and various resolution images. In

this paper, we propose a mask-wearing detection algorithms in complex situations under different resolution and face sizes. Testing data are utilized for validation of our model.

The paper is organized as follows, method is described in Section 2. Experiments are described and results are analyzed in Section 3. While conclusions are depicted in Section 4.

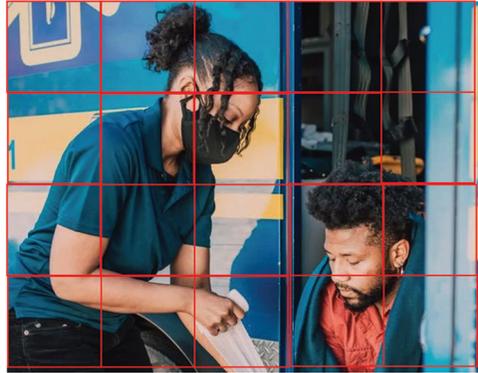
## 2 Method

The proposed technique is for mask-wearing detection purpose. The proposed model is comprised of two streams; the first one is to detect anchors for faces and the second stream is utilized to detect the mask-wearing objects. The block diagram of the proposed technique is depicted in Fig. 1. The face detection stream identify and compute the anchor boxes which contains the faces using sliding window. Fig. 2. Display anchor boxes over the whole pictures, while Fig. 3. Displays anchor boxes over two faces with different orientation.

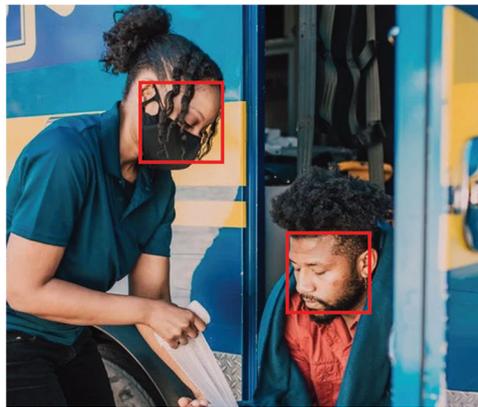


**Figure 1:** The block diagram of the proposed technique

Mask-wearing detection can be addressed as object detection algorithm. Many object detection algorithms are established in the literature. Neural Networks (CNN), support vector machine (SVM) and naive Bayes classifier, are examples of such object detection algorithms. CNNs are found to be more suitable for scenes with complex surroundings that have many objects with different orientations. CNNs have superior accuracy and dependability for such cases. We choose to utilize convolution neural networks to resolve this matter.

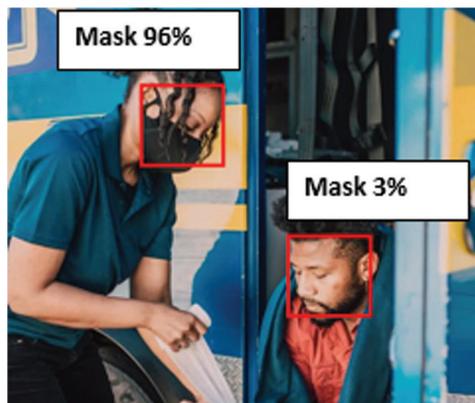


**Figure 2:** Anchor boxes over the whole picture



**Figure 3:** Anchor boxes over two faces with different orientation

In this paper, we are using Region-Based CNN for face detection (R-CNN). R-CNN performs feature extraction via a several layers. Namely: Input, Convolution, relu and pooling layers. Region-Based CNN uses a large number of anchor boxes to the input image. 256 positive anchor boxes and 256 negative boxes will be randomly selected in the training phase. Softmax layer will utilize these anchor boxes to extract candidate regions and their bounding boxes. R-CNN scans images utilizing a sliding windows above the anchors. At the end it yields the anchors with the maximum probability of enclosing a face object as shown in [Fig. 4](#).



**Figure 4:** The result of the Complete learning method (CLM)

To calculate the similarity of an anchor box with the surrounding anchor boxes, using a sliding window, we use the metric in Eq. (1).

$$\text{Similarity} = \frac{\text{Area of overlap of anchor boxes}}{\text{Area of union of compare boxes}} \quad (1)$$

Positive anchor boxes have similarity  $> 0.75$ , while Negative anchor boxes have similarity measures  $< 0.3$ . All other anchor boxes with similarity measures between 0.3 and 0.75 will be ignored in the training.

For an input image  $Im$ , the ground-truth anchor are represented as  $G$ .  $G_A$  represents the selected anchor boxes with Similarity  $> \varepsilon$ , where  $\varepsilon$  is a predetermined threshold that found to give better classification when it is equal to 0.75 as stated before. The symbol  $c_A$  represents the presence of the mask confidence score, calculated by the R-CNN algorithm. The term  $A$  represents the used algorithm,  $W_A$  is the weight of the CNN, and they are depicted in the following Eqs. (2)–(4)

$$G_A, c_A = A(Im, W_A) \quad (2)$$

where  $G_A = A(Im, W_A)[0]$  and  $c_A = A(Im, W_A)[1]$ ,

Therefore, the Loss function can be defined as follows:

$$\text{Loss}(A(Im, W_A)[0], A(Im, W_A)[1], G) \quad (3)$$

To train the model we use the minimization function in Eq. (4).

$$W_A^* = \min_{W_A} \text{Loss}(A(Im, W_A)[0], A(Im, W_A)[1], G) \quad (4)$$

### 2.1 Different Size Face Detection Algorithm (DS-Face)

A one-stage face detection method is proposed that incorporates the fusion of multi-scale features. We present an algorithm called DS-Face to facilitate the detection for different small-sized faces. The training data are utilized on multiple scales to enhance the prediction accuracy. For an input image  $Im$ , we use  $G_D$  to represent the bounding boxes and  $c_D$  to represent the confidence scores computed by the Different small-Size Face Detection algorithm. Where,  $D$  represents the Different Size-Face Detection algorithm and  $W_D$  represents the weight, we can represent that in Eq. (5).

$$G_{D,c_D} = D(I, W_D) \quad (5)$$

### 2.2 Low and High Resolution Mask-Wearing Detection: LHMD

For the face anchors  $G_A$  and  $G_D$  identified by the proposed algorithm, we perform a merging procedure using the following scheme:

---

#### LHMD Algorithm

---

##### Start

- 1) Create an empty list  $G_S$  (list of anchor)
  - 2) Create the score vectors  $c_{SA}$  and  $c_{SC}$ .
  - 3) For anchor(i) in  $G_A$  and the scores  $c_A$ , namely,  $G_A[i]$  and  $c_A[i]$ 
    - a. Insert  $G_A[i]$  and  $c_A[i]$  into  $G_S$  and  $c_{SA}$ .
    - b. Compare  $G_A[i]$  with all the anchors in  $G_D$ .
    - c. If any anchor  $G_n$  in  $G_D$  possess more than 70% overlapping area with  $G_A[i]$ :
-

---

**Algorithm:** (continued)

---

**Then**

- Remove  $G_n$  from  $G_D$
- Remove the corresponding entries from  $Score_p$ .
- Calculate the average (Avg) value of the removed entries in  $Score_p$
- Insert Avg into  $c_{SC}$ .

**Else**

Insert zero into  $c_{SC}$ .

**EndIf****EndFor**4) **For each anchor(j) in  $G_s$ ,**

- a. Compute the covering area of anchor (j): Area (i) ;
- b. Store Area(i) in the vector  $\beta$

**Endfor****End**

---

**2.3 CNN for Mask-Wearing Detection**

The three algorithms: R-CNN, DS-Face and LHMD are able to determine that the image encloses a face but cannot determine if a mask is present. Therefore, we are adding the usage of a CNN to identify mask-wearing. For anchor boxes identified by the Different Size-Face Detection algorithm, we will use a CNN for mask-wearing detection. The face region is enlarged and used as an input to the CNN for prediction. The confidence score designates there is a mask-wearing on the identified face through the CNN, as shown in Eq. (6).

$$Score_p = P(G_D, Im, W_P) \quad (6)$$

P represents the forward propagation function of CNN. Where P denotes the composition of two convolutions followed by a fully connected (FC) layer. And the Loss function is presented in Eq. (7).

$$minimize (Loss(G_D, P(G_D, Im, W_P), G)) \quad (7)$$

**3 Experimental Results**

A large dataset of mask-wearing faces is essential for the training phase of deep learning networks. Classification of mask-wearing faces and non-wearing masks. In [18], they build a public dataset of mask-wearing faces (CMFD), and non-wearing masks (IMFD) in the MaskedFace-Net. 137, 016 face images that include different sizes faces and are accessible at [18]. Half of them are mask-wearing faces and the other half of non-wearing mask faces.

The performance of our proposed method is evaluated using multiple criteria. We compute the true positive (TP), false positive FP, false negative FN, the recall and precision rates. Six-fold cross validation model is utilized for the experiments. The dataset is randomly partitioned into six partitions. The Training set contains half of the images (both faces with masks and without masks). The validation subset contains 2/6 and the testing subset contains 1/6 of the whole database.

### 3.1 Analysis

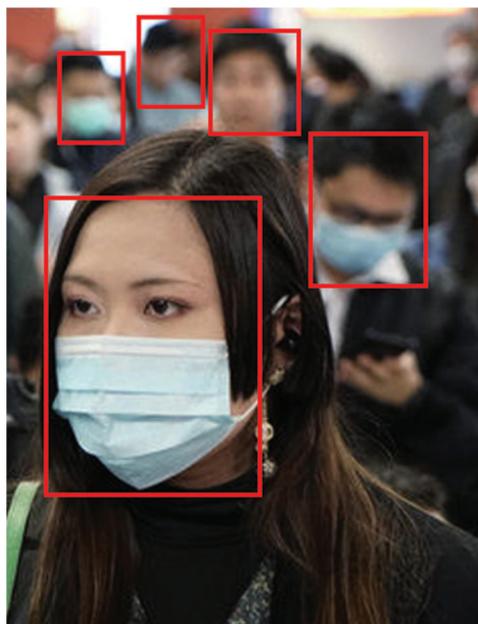
The experimental results of the R-CNN for identifying faces are depicted in Figs. 5 and 6. From the results, it is concluded that the R-CNN is fit to identify large faces, but not small faces. While the combined DS-Face algorithm combined with LHMD algorithm plus CNN can identify small faces with high and low resolution as shown in Fig. 6, while Fig. 7 depicts the results of combing the three algorithms DS-Face + LHMD + R-CNN.



**Figure 5:** R-CNN for identifying big faces



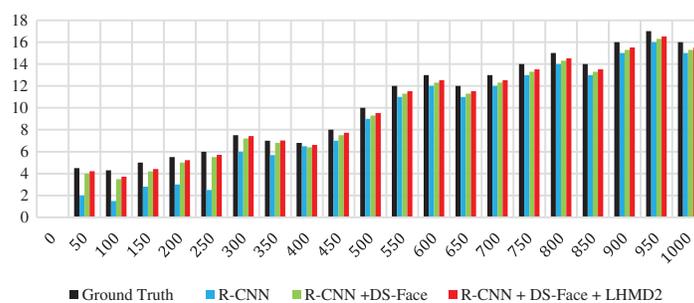
**Figure 6:** Detecting small faces with high and low resolution by DS-Face combined with LHMD + CNN



**Figure 7:** DS-Face + LHMD + R-CNN

A comparison of the ability of the three models to detect faces in a given area is depicted in Fig. 8. The test included 2000 images from the data set. The number of faces of different sizes and different resolution are counted for each model. We compare the ground truth (the actual number of faces in a given area (the area is given by regions of images of area  $P \times P$ , where  $p$  is the number of pixels) with the three models:

- 1) The R-CNN alone
- 2) The R-CNN combined with the DS-Face
- 3) The R-CNN with both the DS-Face and the LHMD models



**Figure 8:** Statistics of the ground truth against R-CNN, R-CNN+DS-Face and R-CNN+DS-Face+LHMD algorithms

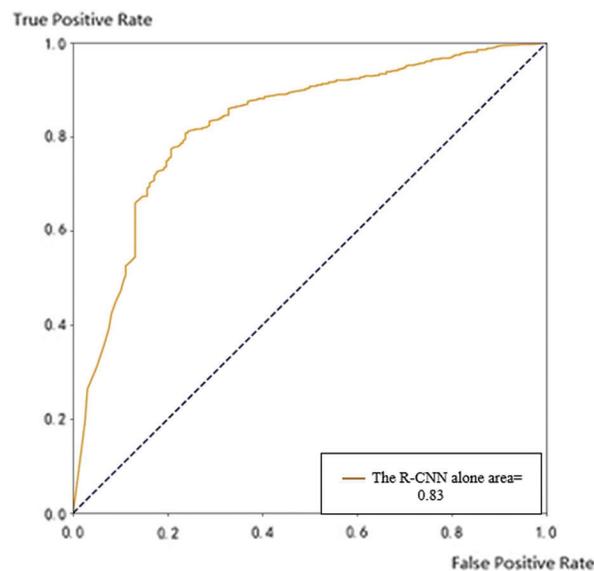
The ground truth specified that there are 7883 faces with large size and 3170 faces with small sizes among the 2000 image dataset. The R-CNN model shows the ability to detect 7700 large size faces with positive rate of 97.5% and 1050 small size faces with true positive rate of 33%. It was also observed that the false negative rate (FN) of large size faces is 3.2%, while the FN rate of small size faces is around 60.5%. The results were expected because R-CNN is trained mainly on large size faces.

### 3.2 Accuracy of the Mask Detection Algorithm

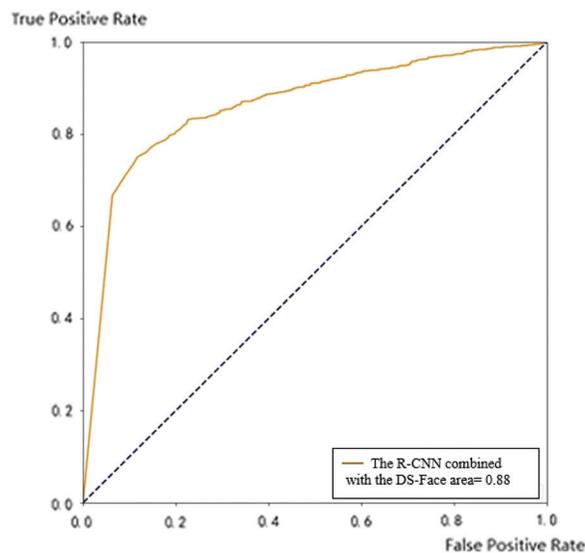
In this section, we are going to evaluate the mask detection Algorithm accuracy. We found that increasing the number of training steps in the model will avoid overfitting. We made different sets of experiments to observe the accuracy with various number of training steps. At the training phase, we labeled the face in the training dataset with wearing or not wearing mask labels. We utilized images from the dataset as training subset to assess the accuracy of the model. We used 10,000 steps in the training procedure and tested the images in the training subset, the accuracy was not satisfactory with missing most of the small faces. We can state that the precision of the model is evaluated by the number of detected faces (large and small). We repeated the experiment by increasing number of training steps to 20,000 steps. We found that the number of detected faces in 1500 images as testing subset is kept at about 1900 target. With increasing the number of training steps, the count of detected faces increased slightly. with 50,000 training steps, the detected faces increased to 2511 faces. We measured the precision and recall rate at 50,000 training step milestone and found them to be 88.3% and 86.9% respectively. By increasing the training steps to 75,000 steps, the number of detected faces reached around 3000 faces with precision rate of 93.2%, and recall of 91.3%. The results for different number of training steps is depicted in [Tab. 1](#). We also display the ROC curve on the training image set in [Figs. 9–11](#).

**Table 1:** Number of training steps and accuracy of the model

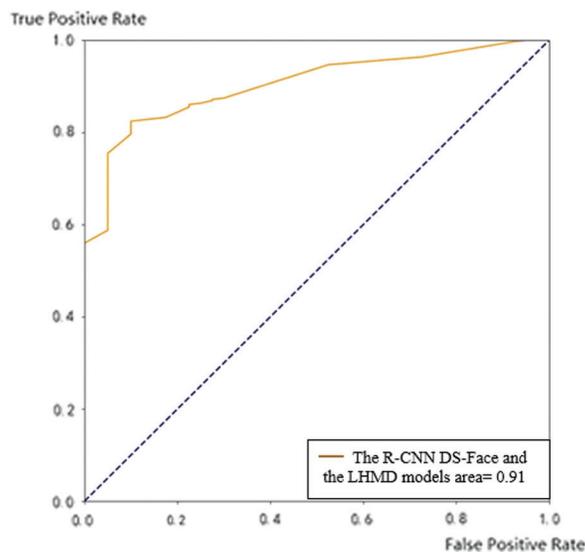
Number of training steps	Precision (%)	Recall (%)
10,000	80.0	72.4
20,000	83.5	81.9
50,000	88.3	86.9
75,000	93.2	91.3



**Figure 9:** ROC of mask detection algorithm preceded by the R-CNN alone



**Figure 10:** ROC of the mask detection algorithm preceded by both R-CNN and DS-Face



**Figure 11:** ROC of the mask detection algorithm preceded by R-CNN with both the DS-Face and the LHMD algorithms

The accuracy of the trained small Face model of the DS-Face algorithm is aided by decreasing the scoring threshold to 0.5, to increase the capabilities of the Small Face model to identify the locations of the small faces. DS-Face face achieved detection with precision of 88.6% for small faces, and recall rate of 68.9%. To measure the accuracy of the R-CNN we cropped the images to include only one target. We applied the cropping to 7,000 images from the training image set. We got over 30,000 images with only one face in it after the cropping phase. We utilized 25,000 images as training input for the R-CNN, and the 5,000 images for the validation phase to measure the accuracy of the R-CNN. The 30,000 cropped images are partitioned into people with and without masks.

We utilized the cross-validation method [24,25], and the CNN composed of six convolution layers followed by pooling layers. The first layer and the second convolution kernel layer are defined as [5,5], while the size of the third and fourth convolution kernel layers is [3,3]. With this configuration, the precision of the classifier at the last layer reached 90.3%. The ROC curve on the training set using the R-CNN alone is shown in Fig. 9.

The ROC curves of the R-CNN combined with DS-Face proves that such combination can enhance the accuracy of the classification result. The area under its ROC curve coverage increased to be 0.88, as shown in Fig. 10.

The ROC curve of the combined R-CNN and DS-Face has been increased from 0.83 to reach 0.88. The ROC curve covered of the model that combined R-CNN, DS-Face and the LHMD model has the largest area under the curve than the previous two models. This model is the best among the three models and has coverage area of 0.91.

### 3.3 Comparison of the Three Models

We performed a comparison between the three models using the metrics TPR, FPR, FNR, precision, and recall. The results are demonstrated in Tabs. 2 and 3. The tables summarize the same results showing the third model that combines the R-CNN and our propose algorithms DS-Face and LHMD works better in terms of all the metrics.

**Table 2:** The metrics TPR, FPR, FNR

Model	True positive rate (%)	False positive rate (%)	False negative rate (%)
Mask detection algorithm preceded by R-CNN alone	84.7	43.1	52.7
Mask detection algorithm preceded by R-CNN and DS-Face algorithms	89.8	25.8	21.9
Mask detection algorithm preceded by R-CNN with DS-Face and LHMD algorithms	95.6	4.7	4.7

**Table 3:** Comparison of precision and recall

Model	Precision (%)	Recall (%)
Mask detection algorithm preceded by R-CNN alone	85.4	27.3
Mask detection algorithm preceded by R-CNN and DS-Face algorithms	93.5	48.1
Mask detection algorithm preceded by R-CNN with DS-Face and LHMD algorithms	95.5	57.5

## 4 Conclusion

In this research, we propose a framework to incorporate two deep learning procedures to develop a technique for mask-wearing recognition especially in complex scenes and various resolution images. An R-CNN is used to detect regions of faces, which is further enhanced by introducing a different size face detection even for smaller targets. To increase the accuracy of our R-CNN we integrated two more algorithms. DS-Face to increase the accuracy of detecting different size faces especially small faces.

While LHMD algorithm is to increase accuracy for low resolution targets. The combined model improves the accuracy to a higher extent. Our experiments proved that the DS-Face with the mask classifier CNN overcome the inadequacies that the R-CNN model alone have when trying to detect small faces. The accuracy of a single deep learning model R-CNN combined with the CNN for mask detection did not meet perform well for mask detection because it misses a lot of small faces and low resolution targets. We reached a conclusion of combining two other algorithms to the R-CNN to detect more faces to achieve better results. We combined the base algorithm R-CNN with other algorithms for better mask detection algorithm to improve the detection accuracy. Our model can be faster and performs in real time by utilizing GPU and distributed computing for faster processing speed especially in complex scenarios.

**Funding Statement:** This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] D. Hui, I. Azhar and T. Madani, "The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in wuhan China," in *Proc. of Int. Journal of Infectious Diseases*, Cleveland, Ohio, vol. 91, no. 1, pp. 264–266, 2020.
- [2] Y. Liu, P. Sun and M. Highsmith, "Performance comparison of deep-learning techniques for recognizing birds in aerial images," in *Proc. IEEE Third Int. Conf. on Data Science in Cyberspace (DSC)*, Athens, Greece, vol. 3, pp. 317–324, 2018.
- [3] X. Chen, K. Kundu and Y. Zhu, "3d object proposals for accurate object class detection," *Advances in Neural Information Processing Systems*, vol. 2, no. 1, pp. 424–432, 2015.
- [4] W. Liu, D. Anguelov and D. Erhan, "SSD: Single shot detector," in *Proc. of European Conference on Computer Vision*, Paris, France, vol. 4, pp. 21–37, 2016.
- [5] F. Yang, W. Choi and Y. Lin, "Exploit all the layers: Fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Cairo, Egypt, vol. 13, pp. 210–221, 2019.
- [6] C. Szegedy, S. Ioffe and V. Vanhoucke, "Inception-v4 inception-resnet and the impact of residual connections on learning," *International Journal of Artificial Intelligence*, vol. 2, no. 1, pp. 123–134, 2020.
- [7] M. Sadiq, X. Yu, Z. Yuan, Z. Fan, A. Rehman *et al.*, "Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform," *IEEE Access*, vol. 7, no. 6, pp. 127678–127692, 2019.
- [8] G. Cao, X. Xie and W. Yang, "Feature-fused SSD: Fast detection for small objects," in *Proc. Ninth Int. Conf. on Graphic and Image Processing*, New York, NY, vol. 9, pp. 106–113, 2018.
- [9] M. Sadiq, X. Yu, Z. Yuan, F. Zeming, A. Rehman *et al.*, "Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–Computer interfaces," in *IEEE Access*, vol. 7, no. 9, pp. 171431–171451, 2019.
- [10] G. Takacs, V. Chandrasekhar and S. Tsai, "Unified real-time tracking and recognition with rotation-invariant fast features," in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Bern, Germany, vol. 9, pp. 934–941, 2020.
- [11] Z. Zou and Z. Shi, "Ship detection in spaceborne optical image with SVD networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 5832–5845, 2020.
- [12] G. Cheng, P. Zhou and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7405–7415, 2019.

- [13] W. Ouyang, X. Wang and X. Zeng, "Deepid-net: Deformable deep-convolutional neural networks for object detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Ostrava, CZ, vol. 10, pp. 2403–2412, 2020.
- [14] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *Proc. of Int. Joint Conf. on Neural Networks*, Napoli, Italy, vol. 7, pp. 2809–2813, 2019.
- [15] Z. Zhu, D. Liang and S. Zhang, "Traffic-sign detection and classification in the wild," *Computer Vision and Pattern Recognition*, vol. 2, no. 3, pp. 2110–2118, 2016.
- [16] J. Jin, K. Fu and C. Zhang, "Traffic sign recognition with hinge loss trained convolutional neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1991–2000, 2020.
- [17] C. Sagonas, E. Antonakos, G. Tzimiropoulos and S. Zafeiriou, "300 faces in-the-wild challenge: Database and results," *Image and Vision Computing*, vol. 47, no. 2, pp. 3–18, 2016.
- [18] A. Cabani, K. Hammoudi, H. Benhabiles and M. Melkemi, "Masked face-net—A dataset of correctly/incorrectly masked face images in the context of COVID-19," *Smart Health*, vol. 19, no. 1, pp. 125–137, 2020.
- [19] W. Liu, D. Anguelov and D. Erhan, "SSD: Single shot multibox detector," *Computer Vision*, vol. 1, no. 2, pp. 21–37, 2019.
- [20] S. Zhang, G. He and H. Chen, "Scale adaptive proposal network for object detection in remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 864–868, 2019.
- [21] A. Howard, M. Zhu and B. Chen, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *Computer Vision*, vol. 2, no. 1, pp. 214–224, 2020.
- [22] I. Freeman, L. Roesse-Koerner and A. Kummert, "Effnet: An efficient structure for convolutional neural networks," *Image Processing*, vol. 2, no. 1, pp. 310–321, 2019.
- [23] J. Redmon and A. Farhadi, "YOLO9000: Better faster stronger," *Computer Vision and Pattern Recognition*, vol. 2, no. 1, pp. 117–129, 2019.
- [24] M. Sadiq, X. Yu, Z. Yuan and Z. Aziz, "Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform," *Electronics Letters*, vol. 12, no. 2, pp. 1367–1369, 2020.
- [25] M. Sadiq, X. Yu and Z. Yuan, "Exploiting dimensionality reduction and neural network techniques for the development of expert brain-computer interfaces," *Expert Systems with Applications*, vol. 16, no. 1, pp. 123–134, 2021.