



A Hybrid ROI Extraction Approach for Mask and Unmask Facial Recognition System using Light-CNN

Ahmed Ahmed¹ and Faris S. Alghareb²

¹Department of Computer Networks and Internet, College of Information Technology, Ninevah University, Mosul, Iraq

²Department of Computer and Informatics Engineering, College of Electronics, Ninevah University, Mosul, Iraq

Received 26 Feb. 2024, Revised 24 May 2024, Accepted 26 May 2024, Published 7 Sep. 2024

Abstract: In recent years, deep learning-based algorithms have been immensely employed and tested in a variety of real-world applications. The efficacy of such algorithms has been thoroughly examined in a practical setting. In this paper, CNN-based deep learning approaches are utilized to recognize faces in real-time to identify faces with and without mask. We employ pre-trained algorithms (YOLOv2 and SSD) to identify people wearing a face mask, which enables a machine to perform recognition tasks while evolving through a learning method. Meanwhile, if there is more than one person in the scene, the one with the max score will be selected for classification. Thus, a hybrid approach that combines YOLOv2 and SSD algorithms to work in parallel is developed for masked-face extraction. Likewise, the Viola-Jones algorithm is used here to detect faces without mask and randomly select a single region of interest (ROI) to be stored for classification. All pre-processing algorithms work separately in parallel as reconstruction steps for preprocessing to crop the ROI and store images for training and testing dataset. Followed by developing a lightweight computational complexity CNN model for face mask recognition to identify whether the selected person's face is wearing a mask or not. The dataset contains numerous variations in appearance and viewpoint to capture different scenarios with and without mask faces. On average, the proposed face mask detection architecture realizes recall and F1 score of 98.3% and 98.31%, respectively. The training performance, on the other hand, has improved by 32.1% for training time as compared to AlexNet and 40.43% of storage space (model size) reduction compared to EfficientNet-B0. The presented framework architecture is an efficient face mask and unmask detector and can be employed as a robust medical assistant face detector in the healthcare sector for automated tracking of a patient, visitor, or staff member wearing a mask or not.

Keywords: Hybrid ROI extraction, deep learning, CNN, face mask recognition, YOLOv2, SSD.

1. INTRODUCTION

Face mask detection system is a computer vision application designed to identify whether or not individuals are wearing face masks [1, 2]. It can be applied in a wide range of fields and scenarios, such as hospitals, schools, public transportation, and other establishments for ensuring public safety while maintaining strict health guidelines [3, 4]. Nevertheless, it has been more particle in medical buildings since the hit of covid-19 pandemic as a tool to enforce mask-wearing policies in the healthcare sector. The technology typically involves the use of object detection and classification to identify whether the detected person's face is covered with a mask or not [5]. Image classification takes an input image and predicts what object exists in the image. Object detection, on the other hand, is not only concerned with the prediction of an object but also determines its location via surrounding predicted objects with bounding boxes [6, 7]. Mask face detection and identification techniques are powered by deep learning algorithms working alongside each other in parallel to release powerful systems operating in real-time [8].

In the last five years, deep neural networks (DNNs) with dense layers and intensive training data have led the research and development of facial features recognition techniques [9, 10]. Enormous effort has been dedicated to developing efficient and lightweight face-recognition techniques. These techniques were designed to achieve high prediction accuracy while constructed with optimized computational complexity; and therefore, incurring low power [8, 11]. Compared to numerous state-of-the-art DL-based algorithms, CNN is the dominant approach because of its natural ability to learn and discover features directly from raw images dataset rather than these features being explicitly identified and structured by human beings [9, 12]. Constructing a CNNbased model from scratch is considered a time extensive process and might not realize accurate prediction. Consequently, several multiple layers of Artificial Neural Networks (ANNs) for different tasks and applications, such as LeNet, AlexNet [13], VGG16 [14], ResNet, DenseNet, LightCNN, GoogleNet, DeepMaskNet [5], FaceNet, MobileNet, DarkNet, etc. have been developed for a variety of deep learning applications



in the last two decades [11, 15]. To make use of these pre-designed multiclass classification and recognition networks, transfer learning has emerged [16, 17]. Transfer learning is the technique in which pre-constructed models are used to train related tasks instead of building them from scratch, leading to reducing the training time and optimizing the work through building deep learning models in short development cycles. Transfer learning allows for the use of pre-trained models to transfer their acquired knowledge to develop new models with modified tasks [18]. Additionally, these pre-designed networks can be directly employed as pre-trained models to function for a specific task while saving significant time and effort required to build an equivalent network from scratch. Therefore, they have been embedded in deep learning IDE environments such as MATLAB, TensorFlow, PyTorch, etc. for easy access and flexible use [11].

Furthermore, preprocessing for the input images and/or fine-tuning the classifier have been essential to prevent undesired degradation in the accuracy rate [11]. This preprocess step prepares the dataset images by cropping, resizing, or enhancing them to improve analysis by removing irrelevant information that might deceive the model prediction accuracy [19]. Therefore, herein, we introduce a real-time hybrid approach for detecting the region of interest (ROI), i.e., faces with and without mask, cropping them, and then storing these preprocessed images in the dataset for further precise analysis and classification of face mask detector. The primary contributions of this manuscript are highlighted as follows.

- We propose a hybrid approach for ROI selection that achieves improved face mask detector architecture.
- A lightweight CNN-based face mask classification model with reduced complexity is developed. It delivers a significant reduction in Mbytes storage space (95.9% and 42.5%) for the model size, compared to AlexNet and EfficientNet-B0, respectively.
- The findings illustrate the efficacy of the presented face mask architecture to accurately classify faces with and without mask while containing various appearance and information of the dataset.

The rest of the article is structured as follows. Section 2 surveys selected relevant work. Section 3 presents our proposed face mask detector architecture. Findings and discussions are analyzed in Section 4. Also, the description of the dataset is covered in Section 4-A. Section 5 draws the conclusion of the paper.

2. RELATED WORK

In this section, several related prior works that have been published in the literature are reviewed [5, 8, 20-23]. Face mask detection has been thoroughly investigated by the researchers in the field of face detection [23]. More importantly, Convolution Neural Networks (CNNs)

are widely employed to evaluate face mask detection on automatic deep face recognition techniques in recent state-of-the-art deep learning-based approaches [24]. The selected contemporary approaches are listed based on their concepts and paradigms for the model in terms of the model complexity of computations and size in Megabytes, prediction accuracy, and dealing with various illuminations and expressions of masked and unmasked face images.

Naeem et al. [5] introduced a novel deep face mask detection and recognition, namely DeepMaskNet. The proposed framework is capable of detecting a large scale of diverse faces with and without mask dataset, which the authors developed. The created dataset (MDMFR) can be used for face mask detection and masked-face recognition. The model achieves 100% accuracy for mask face detection and 93.3% facial recognition. Meanwhile, the achieved superiority of the model was realized at the expense of incurring six convolutional layers, and thus, a higher number of parameters need to be stored for model deployment. Diaz et al. [8] provided a comprehensive survey for face recognition considering the trade-off between accuracy and efficiency for lightweight model architecture regarding computational complexity. A face mask detection system (FMDS) for lightweight computation cost to be implemented on resource-limited devices, i.e., Raspberry Pi 4B, intruded in [25]. The authors employed a modified version of the Single Shot MultiBox Detector (SSD), namely Pruned-SSD, for face detection whereas the pre-trained models of MobileNetV2 and ResNet50 were utilized for mask and face recognition, respectively. The face detector system was evaluated based on a limited dataset containing 200 images (160 for training and 40 for testing). The model achieves 92.5% mask detection accuracy while it requires 14.8 MB of storage space. Rusli et al. [26] used LeNet for masked and unmasked faces, however, to mainly train the LeNet while focusing on the face and ignoring other details of the image, preprocessing steps were conducted first to prepare the images dataset. The Multi-task Cascaded Neural Network (MTCNN) was employed to crop the region of interest (faces). The dataset used was collected from Kaggle for face mask detection and contains 339 samples for both masked and unmasked faces. The reported findings depict that the proposed approach achieves low prediction accuracy for unmasked faces (78.6%), on average. Similarly, the authors in [27] used the MTCNN along with the MobileNet for face mask recognition, and the study was done using a small dataset that contains limited samples, 313 with mask and 443 without mask. It is worth pointing out that there were limited images in the datasets containing faces with and without mask. This has been reported in several prior works conducted to detect whether a person putting a mask or not [25-27].

On the other hand, the MobileFaceNet was employed as benchmark for constructing lightweight CNN [28]. Diaz et al. [8] examined the impact of developing lightweight face architectures of different real-world applications to

serve as guidance for researchers in the community of face detection and recognition. The researchers in [3] introduced a deep CNN face mask detection technique. The proposed technique utilizes two convolutional layers to realize an optimized face mask detected system that can predict whether a person wearing a mask or not in an image or video streaming. Similarly, the authors in [23] introduced a lightweight face recognition model based on assessing top three lightweight pretrained networks including: ShuffleFaceNet, MobileFaceNet, and VarGFaceNet, using datasets for both masked and periocular face recognition. The authors evaluated the accuracy of the trained models based on measuring different three datasets (LFW, AgeDB-30, and CALFW). On average, the highest accuracy obtained from the validated models are 98% and 100% for masked and periocular faces, respectively. Moreover, the AlexNet and VGG16 networks are also tested and realized accuracy of 96.8% and 97.6%, respectively, for masked images. The size and the number of parameters of AlexNet were reported to be 244MB and 61M, respectively. Meanwhile, the MobileFaceNet model realizes the best optimized performance in terms of size (8.2MB), number of parameters (2M), and prediction accuracy (98.5%). In a similar manner, an accuracy and computational complexity trade-off was conducted recently in [29]. The authors have reported that in case of achieving a moderate complexity while maintaining high accurate prediction, the VGG16 network delivers the best trade-off.

Examining the scene of wearing a sunglass, mask, scarf, etc. on the performance of Face Recognition (FR) algorithms was conducted in [20], which the authors referred to as faces with occlusion. The findings concluded that an occluded image could highly impact the prediction accuracy of face recognition. Also, the study in [30] employed Multi-task Cascaded Convolutional Neural Networks (MTCNN) to recognize a face with mask as the mask could greatly affect the accuracy and robustness of FC algorithms. The authors in [21] proposed a facial recognition system based on MobileNetV2 architecture for features extraction (object detection) along with OpenCV for face detection. The presented approach achieves 99.65% accuracy to decide if an individual wearing a mask or not, however, the used dataset (Real-World-Masked-Face-Dataset) does not contain wide variations in mask types, alternation in appearance and viewpoint (frontal faces and view faces), as these variations can have a strong influence on degrading the accuracy. Voila-Jones together with Haar Cascade and Principal Component Analysis (PCA) were used in [31] for achieving improved face detection system. Furthermore, a Masked Facial Recognition (MFR) approach is proposed in [22] for masked and unmasked face detection system. The Inception-ResNet V1 architecture is employed to train the model using the CASIA dataset while the LFW (Labeled Faces in the Wild) dataset utilized for performance assessment. An accuracy of 96% is achieved. However, the structure of the Inception-ResNet roughly requires 30 convolutional layers, therefore, the proposed technique incurs

higher complexity due to intensive computations.

Considering the transfer learning, an AlexNet-CNN architecture framework was proposed based on transfer learning [12]. The developed CNN model composes of five convolutional layers, three maxPooling layers, three Fully Connected (FC) layers, and a SoftMax classifier. After training the CNN model based on transfer learning, the testing processes are carried out and an accuracy rate of 98% and 99% is achieved for datasets ORL and CUHK, respectively. Nevertheless, AlexNet requires numerous numbers of parameters (60 million and 0.65M neurons) to construct its deep network architecture. Therefore, less complicated networks that require reduced number of neurons and parameters are sought for more optimized and sold solutions. Replacing the last layer of AlexNet with a fully-connected layer to detect masks and helmets was presented in [17]. Similarly, a facemask detector was presented in [32] based on YOLOv5. The system is able to detect mask and unmask accurately for real-time data stream. Also, in [33], the pretrained algorithms of YOLOv3 and SSD were compared for mask face detection. It is reported that YOLOv3 realizes better accuracy (91.28% vs. 86.65% for SSD algorithm).

To summarize, most of these prior works mainly concentrate on evaluating the performance of deep FR technology models while ignoring the model complexity as the proposed models incur high computational cost. To achieve favorable attributes in terms of high prediction accuracy, low computational complexity, optimized model size on disk, and capability of coping with diverse appearances and viewpoints in the face mask dataset, we introduce a lightweight CNN model based on a hybrid ROI approach that achieves excellent prediction accuracy for face mask and without mask detection. Furthermore, the architecture involves preprocessing to prepare the input data samples for training and testing dataset. The preprocessing devotes pre-trained models for YOLOv2/v4 [34] and [35] SSD to implement a hybrid approach for face mask cropping, combined with the Voila-Jones algorithm to prune faces without masks. These algorithms were involved to function alongside each other thereby providing an improved face mask detector system while meeting design constraints for the aforementioned attributes.

3. METHOD OVERVIEW

In this work, we present our approach to recognize human faces with and without mask. An improved face mask detection system that concentrates on preprocessing the dataset through a hybrid ROI approach followed by a lightweight CNN-based model for features extraction and classification is presented. The architecture of the presented method is illustrated in Figure 1. As can be seen, the structure employs pipeline (staged steps) and parallelism techniques to speed up the preprocessing for the data preparation, thereby providing the system with capability of classifying a person's face wearing a mask or not in real-time. First, pre-trained face mask detection detectors, i.e.,

YOLOv2 [34] and SSD [35] an open-source real-time object detection algorithms, are employed to identify whether individuals wearing a mask or not. Each algorithm detects multiple facial masks if there is more than one person in the image wearing a mask. The person's face mask with the highest confidence score will be selected and stored. Secondly, Viola-Jones algorithm [36] performs in parallel to detect faces without mask. In case of multiple faces are detected in the scene, a single face is randomly selected and stored. Finally, these preprocessed images are then utilized to train the proposed CNN model for performance evaluation.

A. ROI Selection

To improve the performance, we deploy a combination of detection algorithms to select human faces with and without mask. Both You Only Look Once (YOLO v2/v4) and Single Shot MultiBox Detector (SSD) algorithms work in parallel to detect individuals wearing a mask. A face mask with the max score is selected from the masked faces. To detect face without mask, we use Viola-Jones algorithm. A single face is randomly selected, in case, when multiple faces are detected in the input image. The extracted ROI-based images are stored for training the model. Figure 2 shows the ROI extracted from YOLOv2, SSD functioning in parallel to localize face objects. For training and testing images, we specify the number of ROI extracted from YOLOv2 and SSD based on the max score as listed in Table I. The number of ROI extracted based on SSD algorithm is higher than that extracted based on YOLOv2 algorithm for both training and testing samples. It can be concluded that the SSD algorithm provides better score in case of small size faces whereas YOLOv2 achieves better max score for frontal view face mask images and fails to accurately localize far or small faces. Meanwhile, there was a considerable number of images both algorithms failed to detect and crop a human face with a mask, last row of Figure 2. This is due to the fact that the dataset contains a variety of images with different masks, face alignment, and face distance or projection in the image (frontal faces or far view faces). The number of undetected samples for training and testing sets is 184 and 36, respectively. Furthermore, both algorithms fail to detect decorated or nonsurgical masks, i.e., masks with a logo, an animals' face, a human, respirator, etc. These masks show high variation in the pixels depicting the mask, i.e., noise, that deceives the working mechanism of the mask detection leading to not detecting decorative, respirator, and cloth or homemade masks, Figure 3 depicts some common types of masks. On the other hand, both algorithms achieve comparable max scores for images containing a surgical mask since it consists of similar image details. Moreover, YOLOv4 was implemented to further examine the object detection accuracy as compared to YOLOv2. It was found that an enhancement of 0.4% for employing YOLOv4 instead of YOLOv2 (98.7% and 98.31%, respectively). Therefore, we maintain the YOLOv2 as our object detector since there is no noticeable improvement in the object prediction accuracy

TABLE I. Number of ROIS extracted based on YOLOV2 and SSD algorithms.

Samples	# ROI from YOLOv2	# ROI from SSD	# of Undetected
Training	432	915	184
Testing	98	249	36

and also to keep a minimized model size of the object detector.

Figure 2 illustrates the testing results of the developed ROI scheme. It is obvious that the proposed hybrid ROI-based is able to detect faces with mask associated with large variations in appearance while most of the existing face detection techniques work well on frontal faces. On the other hand, there is a massive intention and demand to reduce the model size on disk. As shown in Figure 2, the developed ROI approach significantly removes unimportant pixels to help lower the unnecessary parameters in CNN, thus obtaining a lightweight structure while delivering competitive accuracy. In this case, the model size on disk will highly be shrunk, Section 4.2 further discusses the model storage space.

B. The Proposed CNN Base Model

The diagram of the network shown in Figure 4 represents the base model depicted in Figure 1, which is the last processing stage that provides the output. It consists of three pairs of convolutional layers and one FC layer. The output of each convolutional layer is passed through an activation function, rectified linear unit (ReLU), crossnorm, and max-pooling layers. The last max-pooling layer is followed by the FC layer. Lastly, a SoftMax layer is added after FC layer to achieve accurate predicting for the output class. The convolutional and maxPool layers are used for filtering and shrinking features from the dataset. While the fully connected layer including the SoftMax is essential for assuring probabilistic decision-making process. Note that hereafter, we refer to the developed CNN as the base model whereas the general structure is the proposed face mask detection architecture.

C. Algorithm Summary and Complexity Analysis

We summarize the proposed method in Algorithm 1. First, the pretrained YOLOv2 and SSD algorithms are used independently to detect a face mask. A face mask with the highest confidence score is stored for training the model. Similarly, Viola-Jones algorithm is employed to detect a face without mask. A single face is randomly selected among the detected faces in the input image and stored for model training. The complexity of the proposed method is measured depending on the number of layers and parameters. In Table II, we calculate the training time, number of layers, and number of parameters for the AlexNet, Vgg16, GoogleNet, ResNet-50, DenseNet-201,

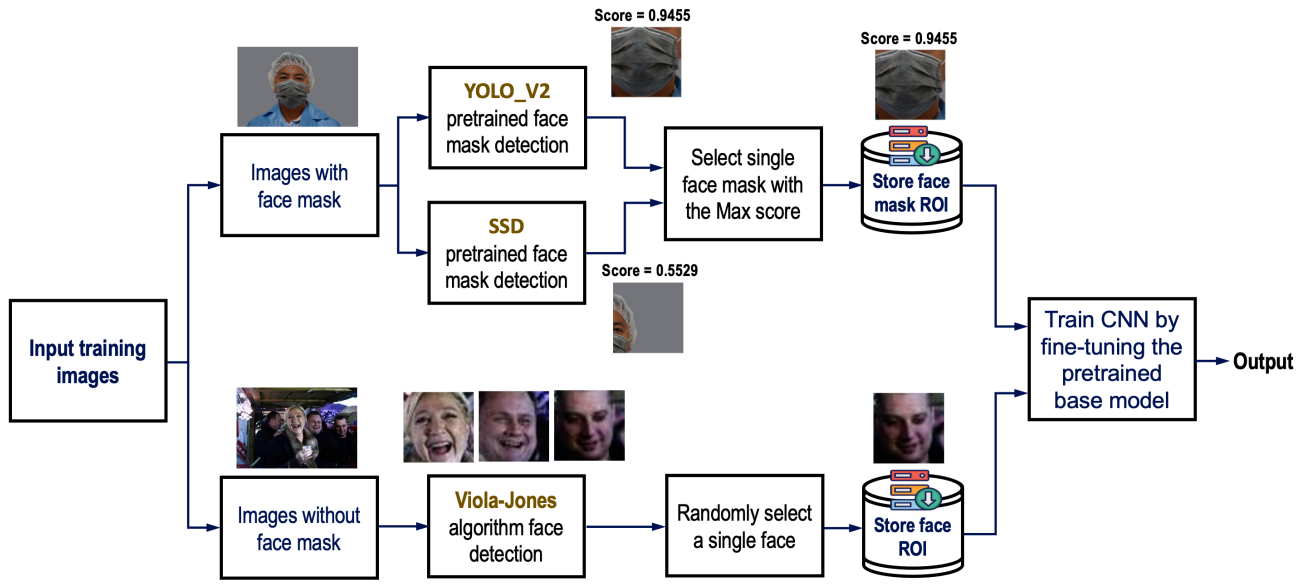


Figure 1. Block diagram of the proposed face mask detector architecture.




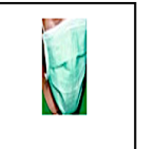












Original image	YOLOv2-based ROI	SSD-based ROI	Max Score ROI/Original
			
			
	Undetected		
		Undetected	
	Undetected	Undetected	

Figure 2. ROI extracted by YOLOv2, SSD, and max score.

Algorithm 1 Training Network with and without face mask

Input: dataset D1_with_faceMask,

D2_without_faceMask,

Pre-trained YOLO_v2 model, Pre-trained SSD model, and Viola-Jones algorithm

Output: Learned model

1: **Initialize:** epoch $m \leftarrow 1$, Max epoch $M \leftarrow 25$,

Batch size $\leftarrow 64$, Learning rate $\leftarrow 1 \times 10^{-3}$

2: **for** sample x_1 in D1_with_faceMask **do**

3: Use pre-trained YOLO_v2 model to extract ROI of x_1

4: Use pre-trained SSD model to extract ROI of x_1

5: Select a single ROI with max score

6: Store face mask ROI

7: **end for**

8: **for** sample x_2 in D2_without_faceMask **do**

9: Use Viola-Jones algorithm to extract ROI of x_2

10: Randomly select single ROI

11: Store face ROI

12: **end for**

13: **repeat**

14: train CNN on Stored ROI with and without mask

$m \leftarrow m + 1$

15: **Until** $m = M$

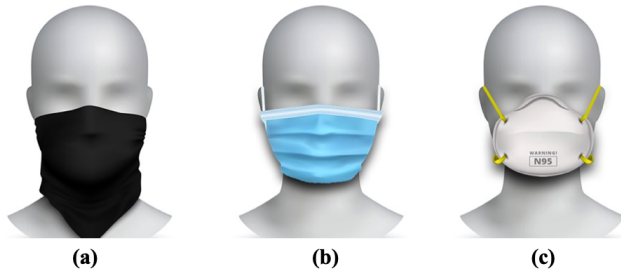


Figure 3. Common facemask types; (a) Cloth homemade, (b) Surgical, and (c) N95 mask.

EfficientNet-B0, and the proposed light-CNN model. It is concluded that the training time of the proposed model is decreased by 73.23 sec to realize 32.1% improvement of the training performance compared to AlexNet and 66.81% and 93.85% over GoogleNet and efficientNet-B0, respectively. In addition, the number of parameters is also reduced to 2.3M, which achieves a reduction of 95.9% in the model size for optimal model parameters. This is due to the fact that the classifier of the developed model only involves two classes at the output, and the structure of the CNN-based model incurs 16 layers instead of 25 layers in AlexNet, 144 layers in GoogleNet, and 290 lightweight layers in EfficientNet-B0, which in turns required reduced number of parameters (2.3M) compared to (58.2M, 5.9M, and 4M) according to AlexNet, GoogleNet, and EfficientNet-B0 network architecture.

4. EXPERIMENTAL SETUP AND RESULTS

We present a detailed description of the datasets and demonstrate the results of the proposed method to recognize human faces with and without mask. The training and testing performance were carried out on a computer with the following specifications: an Intel core *i7* processor running at a clock speed of 2.9 GHz, 8 GB RAM, and GPU (Graphics processing units) Nvidia GeForce RTX 3060 with 16 GB display memory. Additionally, MATLAB version R2022b was used as the IDE (Integrated Development Environment) and programming language for preparing the dataset (preprocessing for cropping images) and model training and testing processes.

A. Dataset Description

We utilized the dataset from Kaggle [37] to train the developed model and evaluate its performance. It includes faces with and without mask associated with significant variations in illumination and expression. The dataset encompasses 3,832 images where 1,914 images with mask and 1,918 without mask. The dataset is randomly partitioned into 80% images for training and 20% images for testing. Table III lists the number of training and testing samples for each class.

It is worth noticing that the dataset contains many alterations in appearance and viewpoint and diverse types of masks to capture different scenarios as the accuracy and

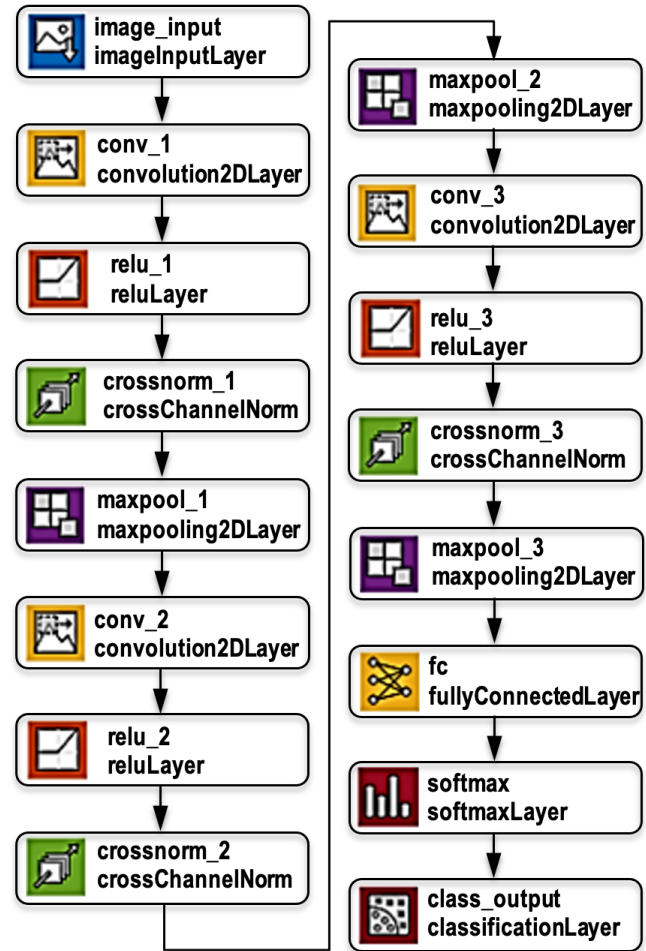


Figure 4. Proposed CNN base model.

precision of a face mask detection method can be impacted by these variations. And the size of the images was different, therefore, in the pre-processing step we resize the cropped images to be 227×227 pixels.

B. Results and Discussions

The efficacy of the proposed method is evaluated on face dataset with and without mask. The performance of our proposed face mask detection architecture is compared with the developed CNN as the baseline and with AlexNet as well. To train the proposed light CNN model for optimization, stochastic gradient decent with momentum (SGDM) of 0.9, learning rate of 1×10^{-3} , batch size 64 (due to limited display memory for the GPU), and maximum epoch 25, were configured for model set up.

As listed in Table II, the proposed light-CNN achieves favorable attributes compared to VGG16, GoogleNet, ResNet-50, DenseNet-201, and EfficientNet-B0. It realizes the highest accuracy while incurring the lowest model size on disk. Moreover, it terms of number of training parameters, the proposed light-CNN model achieves the

TABLE II. Performance analysis of the proposed light-CNN compared to some selected well-known deep CNN networks.

Network	# of Layers	Year	Accuracy (%)	Model Size on Disk (MB)	# of Parameters (M)	Training Time (sec)	Test Time (sec)	Inference Time (msec)
AlexNet [13]	25	2012	96.87	209.51	58.2	228.13	1.5	1.95
VGG16 [14]	41	2014	96.74	473	134.2	1983.81	8.67	11.3
GoogleNet [38]	144	2015	97.52	21.1	5.9	466.77	2.8	3.65
ResNet-50 [39]	177	2016	96.09	83.1	23.5	1173.07	5.16	6.72
DenseNet-201 [40]	708	2017	96.87	64.8	18.1	4996.65	18.92	24.66
EfficientNet-B0 [41]	290	2019	95.18	14.4	4	2518.76	7.3	9.51
Proposed Light-CNN	16	2024	98.31	8.578	2.3	154.9	1.22	1.59

TABLE III. The training and testing samples of a face dataset with and without mask.

Label	Training set	Test set
With mask	1531	383
Without mask	1534	384

TABLE IV. Performance comparison for face dataset with and without mask.

Method	Accuracy (%)
Base CNN	95.83
AlexNet	96.87
Proposed Architecture	98.31

best optimized number of trained parameters 98.28% and 42.5%, compared to the VGG16, which requires the highest number of parameters since it uses only 3×3 kernel size of the convolutional filters, and EfficientNet-B0, respectively. In terms of testing time, AlexNet provides 1.5 sec, however, the proposed model outperforms AlexNet by 18.67%, realizing the lowest testing time. This is contributed to the fact that our presented CNN network consists of only 16 layers, thereby incurring reduced computing operations, and thus shorter propagation delay and execution time. Furthermore, to demonstrate the efficacy of the proposed face mask and unmask detection system in terms of floating point operations (FLPOS), we calculated the inference time for deployment. The proposed architecture provides an improvement of 18.46% compared to AlexNet, which realizes the second best inference time among all implemented deep CNN networks. This is due to the structure of the presented light-CNN incurs less cascaded convolutional layers, leading to shrinking the amount of performed floating point operations which in turn reduces the overall execution time, as listed in column 2 and 9 of Table II.

For model performance evaluation, accuracy (ACC), precision (PRE), recall (REC), and F1 score (F1) metrics are used and calculated in (1), (2), (3), and (4)

$$ACC = \frac{A1 + A2}{A1 + A2 + B1 + B2} \quad (1)$$

$$PRE = \frac{A1}{A1 + B1} \quad (2)$$

$$REC = \frac{A1}{A1 + B2} \quad (3)$$

$$F1 = 2 \times \frac{PRE \times REC}{PRE + REC} \quad (4)$$

Where, $A1$, $A2$, $B1$, $B2$ represent true positive, true negative, false positive, and false negative, respectively. As listed in Table IV, the proposed method outperforms the base CNN by 2.59%, AlexNet by 1.49%, and GoogleNet by 0.81%. Moreover, Figure 5 depicts the confusion matrix of the base CNN, proposed model architecture, and some selected (commonly used) deep CNN networks. The proposed face mask detection architecture achieves the best competitive accuracy for all evaluation metrics.

As seen, we compute the performance metrics PRE , REC , and $F1$ for each class to evaluate the performance of our proposed face mask detection architecture compared to the base CNN, developed herein. It is clearly noticed that the proposed structure for preprocessing and hybrid selection approach for ROI based on the max score leads face mask detector framework outperforms the develop base model and slightly better than AlexNet, VGG16, GoogleNet, and DenseNet-201, meanwhile these networks incur higher computational complexity. Table V and Figure 6 summarize the average PRE , REC , and $F1$ of the proposed architecture improved by 2.56%, 2.58%, and 2.59%, respectively, as compared to the base CNN model.

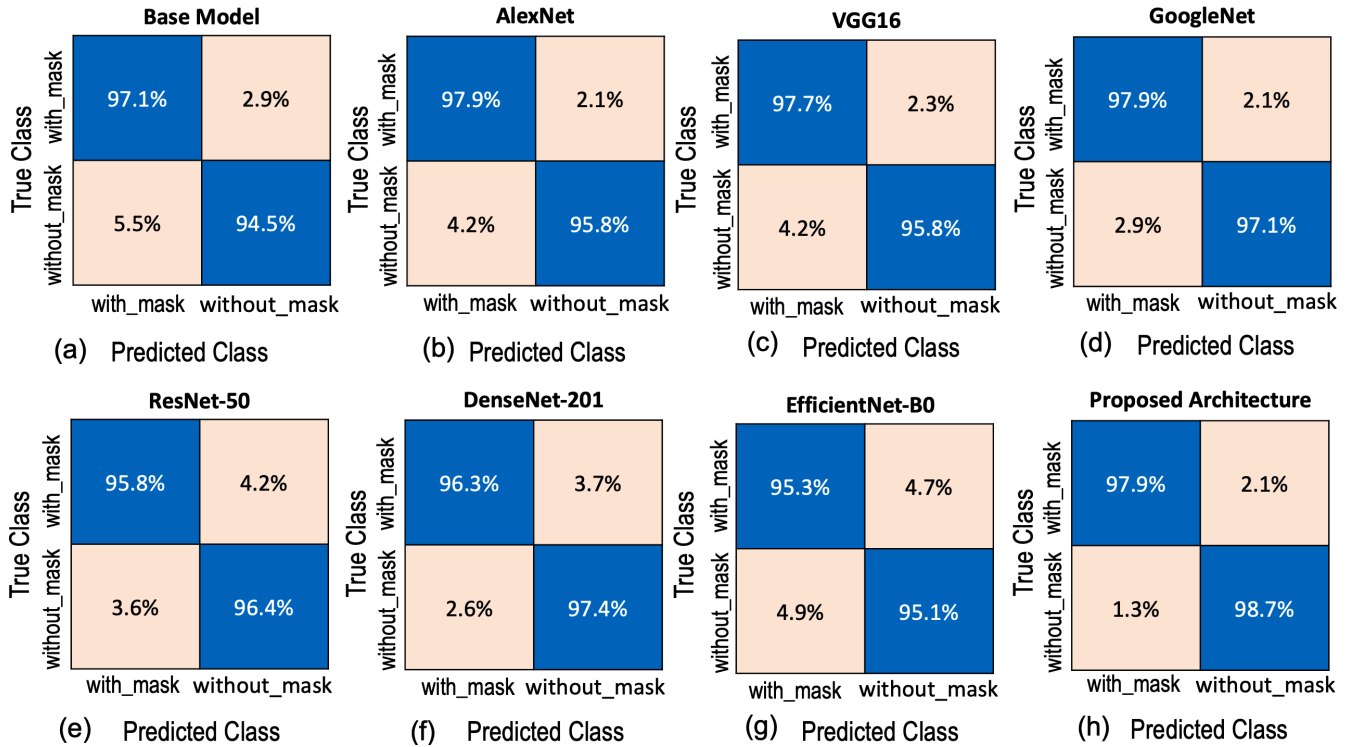


Figure 5. Confusion matrix for (a) the base model, (b) AlexNet, (c)VGG16, (d) GoogleNet, (e) ResNet-50, (f) DenseNet-201, (g) EfficientNet-B0, and (h) proposed architecture.

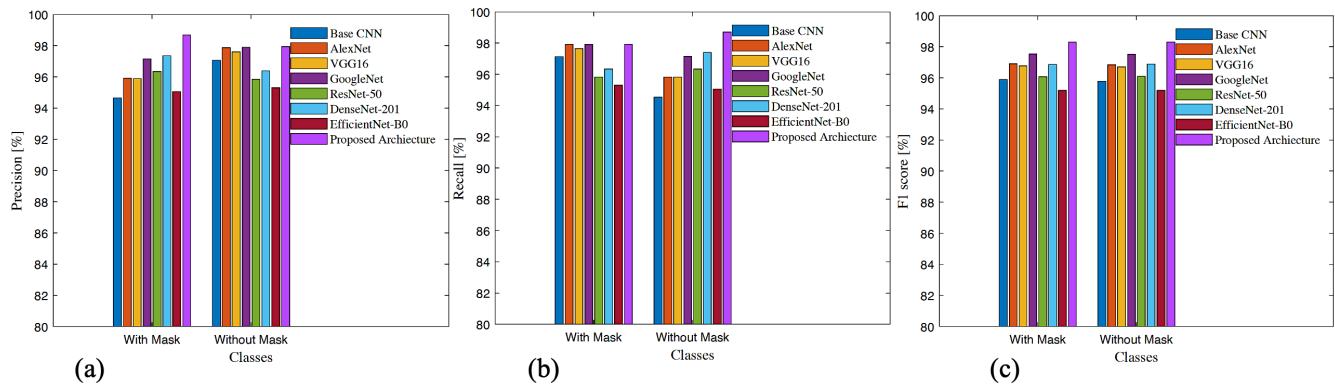


Figure 6. Evaluation metrics to assess performance of the developed approach for detecting a human face with and without mask; (a) Precision, (b) Recall, and (c) F1 score.

As can be noticed, the proposed architecture that implements a light-CNN structure delivers the highest prediction among all implemented CNN network structures based on the average of performance evaluation metrics. This depicts the effectiveness of the developed approach for face detection with and without mask.

The introduced hybrid ROI-based face mask detection outperforms most state-of-the-art approaches presented in the selected prior work. The superiority of the design in terms of performance and accuracy can be attributed to the approach of hybrid selection for ROI, besides the developed

model incurs less complexity which in turn shrinks the potential confusion. Additionally, there has been a burden for energy efficiency due to the intensive growth of big data. Therefore, a custom deep light-CNN network can significantly shrink the gap between big datasets of deep learning applications and the required computational energy.



TABLE V. The values of average pre, rec, and F1 on human face with and without mask.

Method	Average precision	Average recall	Average F1 score
Base CNN	95.86	95.83	95.83
AlexNet	96.89	96.87	96.87
VGG16	96.76	96.74	96.74
GoogleNet	97.53	97.52	97.52
ResNet-50	96.09	96.09	96.09
DenseNet-201	96.88	96.87	96.87
EfficientNet	95.18	95.18	95.18
Proposed Architecture	98.31	98.30	98.31

5. CONCLUSION

In this scholarly research study, a face mask prediction system is presented for detecting whether a person wearing a mask or not. The YOLOv2 and SSD are utilized as pre-trained models for input images to crop masked faces and store the one that has the maximum score whereas the Viola-Jones algorithm from MATLAB R2022b is employed to arbitrarily select and crop a single face without mask. The proposed structure employs a hybrid approach for cropping to select the face mask with the max score. The resilience of the developed face mask recognition architecture is validated for masked and unmasked face images under a variety of conditions such as face alignment and distance (frontal and far view), types of masks, and gender. Also, different evaluation metrics, such as REC, PRE, and F1-score, are calculated to evaluate the performance of the proposed method. The presented face mask detector architecture achieves competitive accuracy (98.31%) for both precision and F1 score. The training performance, on the other hand, has improved by 32.1%, 66.81%, and 93.85% combined with a reduction of 95.9%, 59.34%, and 40.43% for the model size compared to AlexNet, GoogleNet, and EfficientNet-B0, respectively. In short, the proposed architecture is a promising face mask detector that can be leveraged in healthcare systems for more accurate mask and unmask detection.

REFERENCES

- [1] Ganj, A., Ebadpour, M., Darvish, M. et al., "LR-Net: A Block-based Convolutional Neural Network for Low-Resolution Image Classification." *Iran J Sci Technol Trans Electr Eng* 47, 1561-1568 (2023).
- [2] P. C. Neto, J. R. Pinto, F. Boutros, N. Damer, A. F. Sequeira and J. S. Cardoso, "Beyond Masks: On the Generalization of Masked Face Recognition Models to Occluded Face Recognition," in *IEEE Access*, vol. 10, pp. 86222-86233, 2022.
- [3] Kaur, Gagandeep, et al. "Face mask recognition system using CNN model." *Neuroscience Informatics* 2.3 (2022).
- [4] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- [5] Ullah, Naeem, et al. "A novel DeepMaskNet model for face mask detection and masked facial recognition." *Journal of King Saud University-Computer and Information Sciences* 34.10 (2022).
- [6] Phil Kim, "MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence", Springer Science, 2017. ISBN-13: 978-1-4842-2844-9.
- [7] R. Li and J. Yang, "Improved YOLOv2 Object Detection Model," 2018 6th International Conference on Multimedia Computing and Systems (ICMCS), Rabat, Morocco, 2018, pp. 1-6.
- [8] Y. Martinez-Diaz, M. Nicolas-Diaz, H. Mendez-Vazquez, L. S. Luevano, L. Chang, M. Gonzalez-Mendoza, and L. E. Sucar, "Benchmarking lightweight face architectures on specific face recognition scenarios," *AI Review*, vol. 54, pp. 62016244, Feb. 2021.
- [9] Guodong Guo, Na Zhang, "A survey on deep learning based face recognition," *Computer Vision and Image Understanding*, Volume 189, 2019.
- [10] Ranjan Kumar Mishra, G. Y. Sandesh Reddy, Himanshu Pathak, "The Understanding of Deep Learning: A Comprehensive Review", *Mathematical Problems in Engineering*, vol. 2021, Article ID 5548884, 15 pages, 2021.
- [11] Qiangchang Wang, Guodong Guo, "Benchmarking deep learning techniques for face recognition," *Journal of Visual Communication and Image Representation*, Volume 65, 2019.
- [12] G. Lokku, G. H. Reddy and M. N. G. Prasad, "A Robust Face Recognition model using Deep Transfer Metric Learning built on AlexNet Convolutional Neural Network," 2021 International Conference on Communication, Control and Information Sciences (ICCIsc), Idukki, India, 2021, pp. 1-6.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. Curran Associates Inc., pp. 1097-1105, 2012.
- [14] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.
- [15] R. Ranjan, V. M. Patel and R. Chellappa, "HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 121-135, 1 Jan. 2019.
- [16] Weiss, K., Khoshgoftaar, T.M. and Wang, D, "A survey of transfer learning". *J Big Data* 3, 9 (2016).
- [17] K. A. Alshehhi, M. Y. Almansoori, M. K. Alnaqbi, Y. H. K. Aljewari and A. Desmal, "Mask and Helmet Detection using Transfer Learning," 2022 *Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, 2022, pp. 1-4.
- [18] S. Khan, E. Ahmed, M. H. Javed, S. A. A. Shah and S. U. Ali, "Transfer Learning of a Neural Network Using Deep Learning to Perform Face Recognition," 2019 *International Conference on*



- Electrical, Communication, and Computer Engineering (ICECCE), Swat, Pakistan, 2019, pp. 1-5.
- [19] M. S. Ejaz, "Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition," in 1st International Conference on Advances in Science, Engineering and Robotics Technology 2019 (ICASERT 2019), 2019.
- [20] Y. Su, Y. Yang, Z. Guo and W. Yang, "Face recognition with occlusion," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 2015, pp. 670-674.
- [21] Talahua, J.S.; Buele, J.; Calvopiña, P.; and Varela-Aldás, J., "Facial Recognition System for People with and without Face Mask in Times of the COVID-19 Pandemic," *Sustainability* 2021, 13, 6900.
- [22] Mishra, Saroj, and Hassan Reza, "A Face Recognition Method Using Deep Learning to Identify Mask and Unmask Objects." 2022 IEEE World AI IoT Congress (AIIoT). IEEE, 2022.
- [23] Y. Martínez-Díaz, H. Méndez-Vázquez, L. S. Luevano, M. Nicolás-Díaz, L. Chang and M. González-Mendoza, "Towards Accurate and Lightweight Masked Face Recognition: An Experimental Evaluation," in *IEEE Access*, vol. 10, pp. 7341-7353, 2022.
- [24] J. Deng, J. Guo, X. An, Z. Zhu and S. Zafeiriou, "Masked Face Recognition Challenge: The InsightFace Track Report," 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 1437-1444.
- [25] X. Li, "An Effective and Efficient Face Mask Recognition System for Edge Devices," 2022 5th International Conference on Data Science and Information Technology (DSIT), Shanghai, China, 2022, pp. 1-5.
- [26] M. H. Rusli, N. N. A. Sjarif, S. S. Yuhaziz, S. Kok and M. S. Kadir, "Evaluating the Masked and Unmasked Face with LeNet Algorithm," 2021 IEEE 17th International Colloquium on Signal Processing and Its Applications (CSPA), Langkawi, Malaysia, 2021, pp. 171-176.
- [27] H. -M. Tang and F. -C. You, "Face mask recognition based on MTCNN and MobileNet," 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, 2021, pp. 471-474.
- [28] Liu, W.; Zhou, L.; Chen, J., "Face Recognition Based on Lightweight Convolutional Neural Networks". *Info.* 2021, 12, 191.
- [29] A. Giri, D. S. Bisht, A. Chauhan and I. Kumar, "Computerized Face Mask Detection System Using Deep CNN and Transfer Learning," 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), Dehradun, India, 2023, pp. 457-462.
- [30] C. Qi and L. Yang, "Face recognition in the scene of wearing a mask," 2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI), Ottawa, ON, Canada, 2020, pp. 77-80.
- [31] J. Sikder, R. Chakma, R. J. Chakma and U. K. Das, "Intelligent Face Detection and Recognition System," 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 2021, pp. 1-5.
- [32] B. Bhagabati and K. K. Sarma, "Masked or Unmasked Face Detection from Online Video using Learning Aided Pattern Recognition Method," 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2022, pp. 1-4.
- [33] A. Shaik, R. T. Prabu and S. Radhika, "Detection of Face Mask using Convolutional Neural Network (CNN) based Real-Time Object Detection Algorithm You Only Look Once-V3 (YOLO-V3) Compared with Single-Stage Detector (SSD) Algorithm to Improve Precision," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-6.
- [34] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 2017, pp. 6517-6525.
- [35] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proceedings European Conference on Computer Vision*, 2016, pp. 21-37.
- [36] P. A. Viola and M. J. Jones, "Robust real-time face detection," *Internat. Journal of Comp. Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [37] <https://drive.google.com/drive/folders/1Dm2sV8UrMd6OKzjVkW859WznhfSXFZF8>.
- [38] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A, "Going deeper with convolutions." In *Proceedings of the IEEE conference on computer vision and pattern recognition 2015* (pp. 1-9).
- [39] He K, Zhang X, Ren S, Sun J, "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition 2016* (pp. 770-778).
- [40] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, "Densely connected convolutional networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition 2017* (pp. 4700-4708).
- [41] Tan M, Le Q, "EfficientNet: Rethinking model scaling for convolutional neural networks." In *International conference on machine learning 2019 May 24* (pp. 6105-6114). PMLR.



Ahmed Ahmed received the PhD degree in Electrical and Computer Engineering from the Department of Electrical and Computer Engineering at the University of Missouri, Columbia, MO, USA, in 2020. His research interests include computer vision, machine learning and deep learning. Email: ahmed.ahmed@uoninevah.edu.iq



Faris S. Alghareb received the Ph.D. degree in Computer Engineering from the Department of Electrical and Computer Engineering at the University of Central Florida, Orlando USA, in 2019. His research interests include soft error resilient computing architectures, reliable VLSI design with emphasis on low power and high performance, reconfigurable computing, imprecise signal processing, spin-based emerging Non-Volatile (NV) latching circuits, and applied machine and deep learning. He is a member of the IEEE. Corresponding author Email: faris.alghareb@uoninevah.edu.iq