Journal of Computer Science

Original Research Paper

Gender Classification from Face and Eyes Images Using Deep Learning Algorithm

[1]Zahraa Salam Abu Ghrban and [2]Nidhal Khdhair EL Abbadi

[1]*Department of Computer Science, Faculty of Education, University of Kufa, Najaf, Iraq*
[2]*Department of Computer Techniques Engineering, Al-Mustaqbal University College, Babylon, Iraq*

Article history
Received: 25-12-2022
Revised: 20-01-2023
Accepted: 26-01-2023

Corresponding Author:
Zahraa Salam Abu Ghrban
Department of Computer Science, Faculty of Education, University of Kufa, Najaf, Iraq
Email: zahraa.abu.ghrban@gmail.com

Abstract: Gender classification provides additional information about the individual's identity, which is crucial in surveillance, smart interface, and smart advertising, it is a very important aspect of face analysis that has piqued the interest of researchers in areas such as demographic information collection, surveillance, human-computer interaction, marketing intelligence, security, etc. Usually, facial images are used to extract features for classification. This study aims to determine gender based on the entire face or the eyes (for masked or occlusion faces). The proposed method consists of six main stages: (1) Remove the background using the Deep Labelling Version 3 plus (DeepLabV3+) method; (2) Skin detection by applying the combination of the two colors spaces models (Hue, Saturation, Value (HSV)) and (Luminance, Chrominance blue and Chrominance red (YCbCr)); (3) Face detection using the Haar Cascades classifier method; (4) Face alignment and cropping; (5) Classify the gender. Classification of the gender when the image contains the entire face is based on a deep wavelet. In the case of occlusion faces, we proposed to use a Convolution Neural Network (CNN) for classifying the gender from the eye (s). The dataset used for training this model is the celeb faces attributes dataset (CelebA) and some of the different datasets were also used to compare this model with the other previous works. In addition, we showed that our models perform well with images with some challenges, for example, some of the faces are not fully visible, not completely frontal, wearing glasses with different styles, have low quality and noise, have various lighting, children's and infants' faces, with and without makeup, open and closed mouths and closed eyes. Other than that, the accuracy achieved was 98% when classifying from face images and 98% from the eye. The proposed method was compared with previous works and was very promising. The contributions of this proposal were the use of the deep wavelet for gender classification and the proposal of a new method for skin detection, which enhances the performance of the Haar cascade method used for face detection. Also, we proposed a method for face alignment and, finally, the classification of genders with many challenges from the entire face or just from the eye (s).

Keywords: Color Space, Deep Wavelet, Eye Cropping, Haar Cascades, Human Face Alignment

Introduction

The smart device used has expanded along with technological advancements and social media has begun to catch everyone's attention. Nowadays, gender plays a significant role in how people interact in many apps implementing such methods has increased (Tilki *et al*., 2021).

Soft biometric modalities (ethnicity, gender, age, hair color, iris color, presence of mustache on the face, facial moles, and others) have demonstrated their utility in different applications, including reducing the search space significantly. This leads to improved recognition performance, reduced computation time, and faster processing of test samples (Singh *et al*., 2017). Identifying a person's gender is undoubtedly a simple task for humans. Hence, gender classification is still a hot topic for research in many areas, comprising computer vision and machine learning. Although human faces are an effective visual biometric trait, certain facial features, including semantic patterns, might deceive classification methods that rely on facial pictures (Afifi and Abdelhamed, 2019).

Gender classification has many applications, including intelligent human-computer interaction, social media, augmented reality, and demographic studies. In addition, it is involved in many security fields, such as control of access, individual authorization, and visual surveillance. Apart from that, facial photos are frequently used in these applications because they provide valuable information that could be used to extract human interaction (Mohammed, 2020).

However, due to the COVID-19 pandemic's widespread use of masks, the face appears obscured even in cooperative settings, leaving the ocular region (the eyes) as the only visible part. In this study, we proposed a method of many stages to classify gender from the face or eye images.

An accurate gender classification method can boost the performance of many other applications related to security, health, and smart devices, it is also an important task for many social activities. Gender classification may decrease the computational complexity of human identification systems which highly affects smartphone performance.

The remaining sections of this study first introduce the main contributions of the current paper, then discuss the related works on gender recognition from a face image and the eye region. The background is presented after related works to explain the main algorithms used in this study. The main steps of the proposed method (s) are introduced in the methodology section. The results and discussions are presented in a separate section after the methodology. Finally, we conclude the paper the conclusion.

Paper Contributions

There are several contributions in this study, and the main three contributions are.

The first one is using a new deep learning method based on a wavelet which is similar to a Convolution Neural Network (CNN), while the second contribution is introducing a new method for face detection based on a combination of two-color spaces (YCbCr and HSV). The third contribution, we suggested a new method for the alignment of the face by solving the problem of degrading the digital image due to rotation.

Gender From Face

Tilki *et al.* (2021) suggested Convolution Neural Networks (CNN) and AlexNet, known as two deep learning-based methodologies. Both models successfully classify people by gender (male or female). Moreover, CNN and AlexNet have accuracy scores of 92.40 and 90.50%, respectively. The Adience dataset was employed for the experiments.

Sumi *et al*. (2021) proposed a method based on CNN. Initially, each image is pre-processed using Multi-Task Cascaded Convolutional Neural Networks (MTCNNs). Note that CNN is applied to extract features. The authors built their system with various optimizers and deep learning techniques, including k-fold cross-validation, to get a better outcome. Two datasets were utilized from the Kaggle website and the Nottingham scan database. The maximum accuracy was 90% when using the Nottingham Scan database and 97.44% when using the Kaggle dataset. In a study by Yildiz *et al*. (2021), CNN's algorithm was trained using facial images that varied greatly regarding race, position, age, and illumination. The authors created low-resolution images using bilinear interpolation to create an approximate real-world photography scenario. The studies resulted in a 93.71% success rate on the VGGFace2 dataset and an 85.52% success rate on the Adience dataset. This project aims to recognize gender accurately from low-resolution, uncontrolled images.

Islam *et al*. (2020) established a system, by using Pareto Frontier deep learning networks for automatically classifying photos from the internet by gender. Three distinct Pareto frontier CNN models with ImageNet training were used to examine the experiment (GoogleNet, SqueezeNet, and ResNet50). The experimental outcomes seen by the implemented CNN models revealed their ability to automatically analyze face photos and strengthened their application in a similar type of classification challenge. When the network parameters were combined optimally, the CNN networks pre-trained on the Pareto frontier showed a classification rate of greater than 90% on the WIKI cleaned dataset.

Kamaru (2020) developed a fully customized, handcrafted CNN architecture that needs much fewer trainable parameters and lower input picture resolutions. Additionally, it makes use of batch normalization layers, which boosts computing performance. Based on tests employing datasets that are freely available to the public, such as the LFW and CelebA datasets, their CNN accurately predicted the outcome of 95 and 96%, respectively.

Gender from Eyes Region

Cimtay and Yilmaz (2021) classify gender from eye images based on the most recent deep CNN models InceptionV3, InceptionResnetV2, Xception, and NASNetLarge. Results indicate that accuracy is still quite good even if face data is restricted to the area around a single eye. Furthermore, each eye image in the dataset is normalized to eliminate the impact of various lighting conditions. The model accuracy based on NASNetLarge and Xception was 95.85 and 96.98%, respectively.

Alonso-Fernandez *et al*. (2020) investigated the estimation of age and gender using smartphone-taken ocular selfies. The requirement to use face masks has made partial facial occlusion problems. This was addressed by the adaptation of two already-existing lightweight CNNs presented in the ImageNet challenge and two more proposed architectures for mobile face recognition. In addition, they used networks already trained on ImageNet to combat overfitting because the datasets for soft biometrics prediction using selfie photographs are few. Their test on the Adience dataset revealed an accuracy of 76.6% for one eye image and 78.9% for both eyes.

Alonso-Fernandez *et al*. (2020), used pre-trained CNN architectures that were already in use, followed by SVM classifiers that were suggested in the context of the ImageNet large-scale visual recognition challenge, to predict gender, age, and ethnicity. Experiments using 12007 photos from the Labeled Faces in the Wild (LFW) database showed an accuracy of 92.6% for images of one eye and 93.4% for images of both eyes. Additionally, they examine pictures that simply show the mouth area. They evaluate the viability of employing only the ocular or mouth areas of incomplete pictures.

Tapia *et al*. (2019) introduced a Super-Resolution Convolutional Neural Networks (SRCNNs) method for improving the resolution of low-quality periocular iris pictures clipped from selfies of a subject's faces. This study used a Random Forest (RF) classifier, boosting image resolution by two or three times can increase the accuracy of gender categorization. Therefore, the best gender classification accuracy when upscaling images from 150 × 150 to 450 × 450 pixels was 90.15% for the right eye and 87.15% for the left. These outcomes demonstrate that the gender classification rate rises when the image resolution is increased with the SRCNN.

Rattani *et al*. (2018), the authors conduct a detailed investigation of gender prediction from ocular pictures captured by smartphone front-facing cameras. For gender prediction, the scientists employed pre-trained and customized convolutional neural network designs. The prediction accuracy has been increased by using multiclassifier fusion. Employing pre-trained residual networks as a feature extractor in conjunction with SVM and MLP classifiers, the highest accuracy of 90.0% was reached and by using custom CNN, 89.01% was obtained.

DeepLabv3+

It is a state-of-art deep learning model for semantic image segmentation, where the goal is to assign semantic labels (such as a person, dog, car, and so on) to every pixel in the input image. The Deep Labelling Version 3 plus (DeepLabV3+) model has encoding and decoding phases, Fig. 1. The encoding phase extracts the essential information from the image using a CNN. In contrast, the decoding phase reconstructs the output of appropriate dimensions based on the information obtained from the encoder phase Chen *et al*. (2018).

Haar-Cascades Classification Method

Haar-Cascades were first introduced by Viola and Jones (2001) and it was one of the most popular object detection algorithms. This method can detect different regions of the human image, like the face, eyes, nose, mouth, and body. Note that the haar-cascades method is done with the help of the Haar-like features, as shown in Fig. 2, which can be calculated efficiently using the Adaboost classifier and integral images in the cascade classifier. It is an algorithm that can detect objects, irrespective of their scale in the image and location.

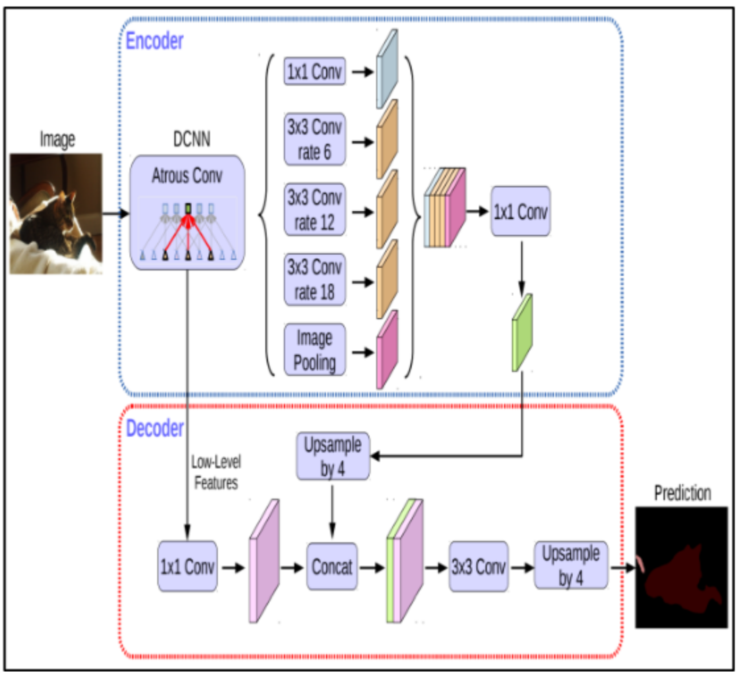


Fig. 1: The architecture of the DeepLabv3+ (Chen *et al*., 2018)

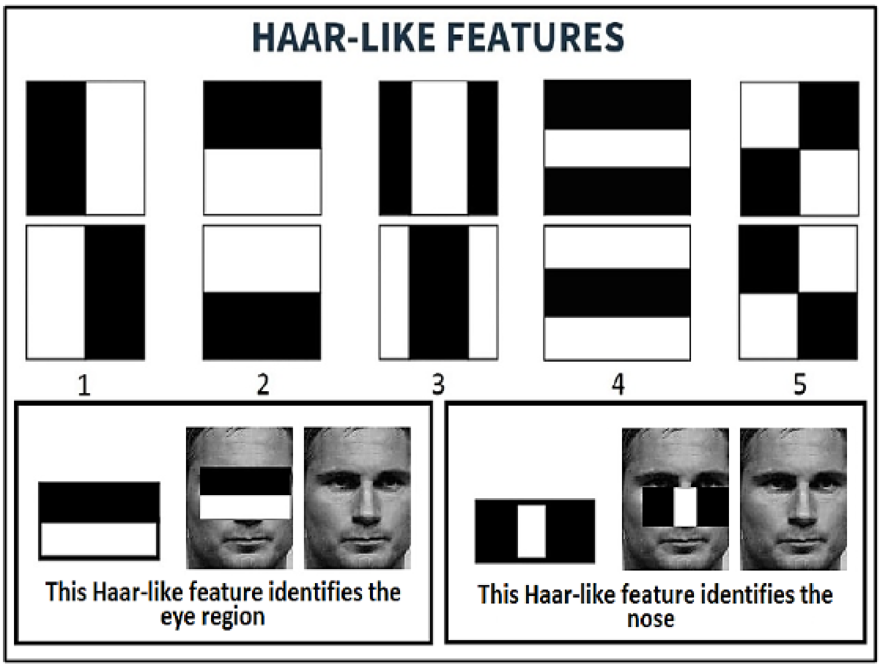


Fig. 2: The Haar-like features for face parts detection

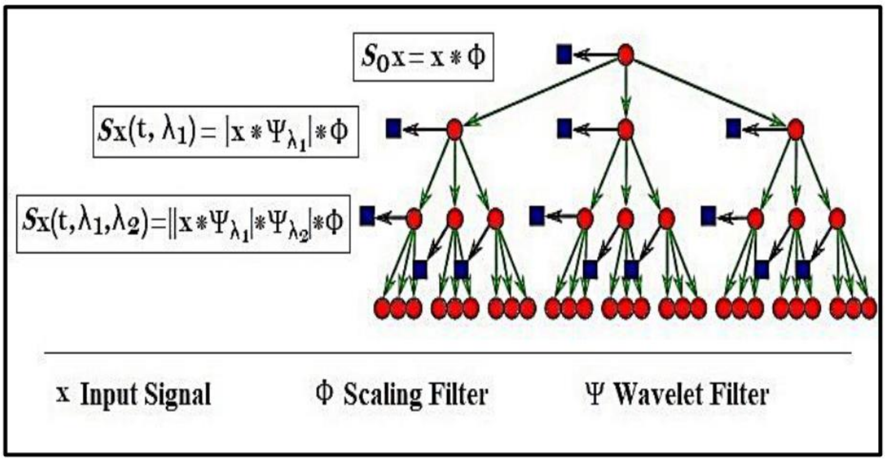


Fig. 3: The wavelet scattering process with hierarchical representation at multiple layers Mallat (2012)

Wavelet Scattering Transform Network

Wavelet scattering is a null-parameter convolution network originally proposed by Mallat (2012). A wavelet scattering network enables the derivation of low-variance features from image data for deep learning applications. Other than that, low-pass scaling filters and predefined wavelets are used in the network. The scattering transform generates data representations that reduce differences within a class while keeping discriminability across classes. The scattering can be utilized successfully in situations with a shortage of training data.

The input signal is first averaged using wavelet lowpass filters; this is the layer zero scattering feature and the high-frequency details are lost with the averaging operation. Subsequently, the details lost in the first step are captured at the next layer by performing a continuous wavelet transform of the signal to yield a set of scalogram coefficients. In this case, a modulus is applied to the scalogram coefficients and the output is filtered with the wavelet low-pass filter, producing a set of layers 1 scattering coefficients. The same process is repeated to obtain the layer 2 scattering coefficients. Correspondingly, the output of the scalogram coefficients in the previous layer always becomes the input to the operations in the next layer. This nonlinear process continues according to the number of layers defined by the user, Fig. 3.

Materials and Methods

The proposed method consists of six main stages: Removing background, skin detection, face detection, face alignment, and cropping, checking for occlusion face, and finally classification stage of the image according to gender (male or female), as illustrated in Fig. 4. The steps of classification of the gender are summarized in Algorithm 1.

Remove the Background

In this step, the background is removed from the image using the Deep Labelling Version 3 plus (DeepLabV3+) network, which was pre-trained on human image segmentation. Note that the input to this network is a colored image.

Algorithm 1: Gender classification

Input: Face image
Output: Gender classification
Step 1: Start
Step 2: Read the colored image.
Step 3: Remove the background using (DeepLabV3+).
Step 4: Detect the skin area using (HSV & YCbCr).
Step 5: Detect the face in the image using (Haar-Cascades).
Step 6: Align the face and crop it from the image.
Step 7: Check if the face is occlusion using
(Mouth detection by Haar-Cascades): **Go to step 8**
A. Image pre-processing, resizing image into (28 × 28) and converting into a grayscale image.
B. Input the face image into the proposed deep wavelet model to classify the gender.
C. Go to step 9.
Step 8: Occlusion face
A. Crop the eyes pair region from the face image (Haar-Cascades).
B. Image pre-processing, resizing image into (88 × 45) and converting into a grayscale image.
C. Input the eyes image into the proposed CNN model to classify the gender.
D. Segment the eyes region into right and left eyes.
E. Image pre-processing, resizing image into (50 × 50) and converting into a grayscale image.
F. Input the eyes image into the proposed CNN model to classify the gender.
Step 9: End

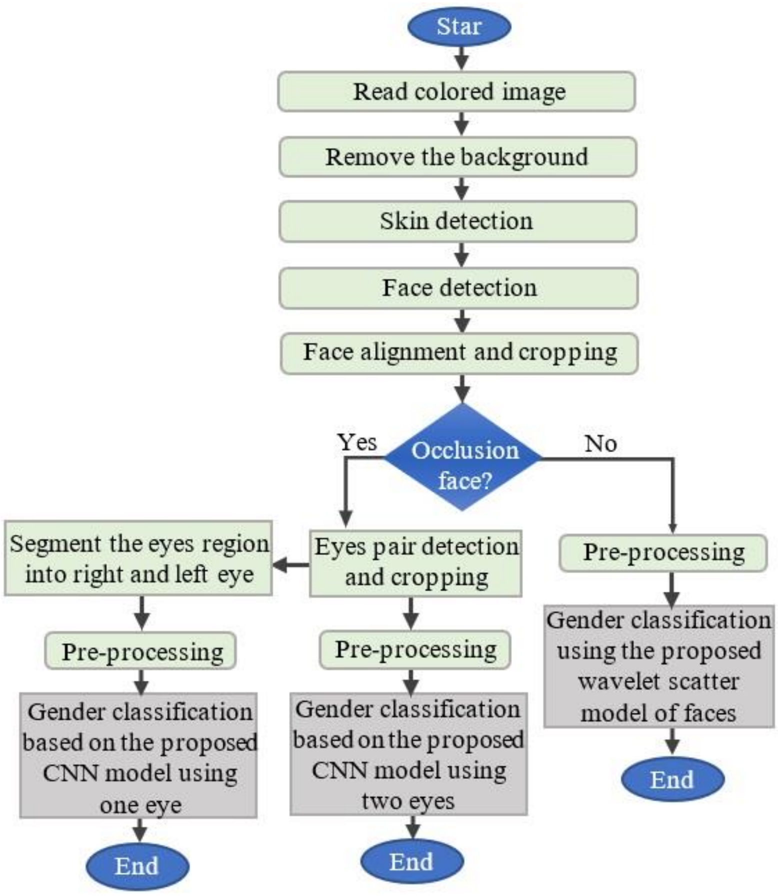


Fig. 4: Flowchart of the proposed algorithm

Skin Detection

The proposed method for skin detection is presented in Fig. 5. We proposed to convert the input color image into two primary color space models, which are Hue, Saturation, Value (HSV) and luminance, Chrominance blue and Chrominance red (YCbCr), to produce two images from the two color spaces (im 1 and 2). For each channel out of the three channels of each image (im 1 and 2), we determined a range of color values experimentally, where these values remain as it is in this channel, while the other values out of this range are converted into zero. These two images (im 1 and 2) are combined (each channel combined with the corresponding channel). The two images (im 1 and 2) are combined by comparing each pixel in each channel in (im1) with the corresponding pixel in the corresponding channel in (im2) according to relation 1. This is done by selecting the same pixel value for the new image when the two pixels have the same value and we select the larger value when the dissimilar pixels value and finally, the three channels concatenating:

$$PNi = \begin{cases} If\ P1i = P2i & Then\ PNi = P1i \\ Else & PNi = \max\left(P1i, P2i\right) \end{cases} \tag{1}$$

where:

- P is the pixel value
- Refer to image 1
- Refer to image 2
- N is the new image

Face Detection

After extracting the skin region, we tried to detect and crop the face using the haar cascades classification method, which depends on Haar-like features to identify the face regions (eyes, nose, mouth). The Haar-like feature is applied to the image with different sizes to find the feature value by subtracting the black average from the white average of the pixels. When the value gets closer to 255, the more likely eyes existed in this region. After finding the parts of the face, the possible area of the face will be surrounded by a box and at the end, all the boxes that are close to each other will be merged into one box that contains all the parts of the face in the image.

Face Alignment and Cropping

To align the face, we proposed a method that depends on detecting the eyes. Note that the first step is to detect the eyes by Haar-Cascades and find the center of each eye. Consequently, the second step is to find the length of the distance between the eyes centered on the Y-axis (*dx*) and the Y-axis (*dy*). The third step is to find the angle, where the angle is calculated by Eq. 2. To find the rotation direction we check if the center value of the left eye is greater than the center value of the right eye according to the y-axis, then the image will be rotated clockwise using relation (3) and vice versa with relation (4). Figure 6 illustrates all the alignment steps. Finally, the image with the removed background is rotated by the final angle and the face is cropped from it:

$$\theta = \tan^{-1}\frac{dy}{dx} \tag{2}$$

where:

θ = The rotation angle
dx and *dy* = The distances

Rotate clockwise

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\tan\left(\theta/2\right) & 1 \end{bmatrix} \begin{bmatrix} 1 & \sin\theta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\tan\left(\theta/2\right) & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{3}$$

Rotate counter clockwise

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} 1 & -\tan\left(\theta/2\right) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \sin\theta & 1 \end{bmatrix} \begin{bmatrix} 1 & -\tan\left(\theta/2\right) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{4}$$

Rotate counter clockwise

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{5}$$

where:

*x** and *y** = The new pixel coordinates
x and *y* = The old pixel coordinates

Using the classic rotation matrix (relation 5) will degrade the image, so we suggested using relations (3 and 4).

Check Occlusion Face

This stage is designed to distinguish the faces of a large part of its occlusion by something such as a mask or anything else. In this case, it is difficult to detect gender without a face. Hence, we focus on detecting gender based on the eyes. The detection of occlusion faces (for example, masked faces) is done using the haar-cascades classification method for mouth detection. This is because the face has no occlusion when the mouth is found in the image, therefore, the gender classification process is based on the face image.

Classify the Gender from the Face Images

In this case, the wavelet scatters method is proposed to classify the gender from the face image. It is a new method and efficient as a deep learning method. We suggested using the architecture of the wavelet scatter by using eight levels of wavelet decomposition. The image size was 28 × 28, the number of epochs was 70 and the batch size was 128. The parameters used in this proposal are listed in Table 1.

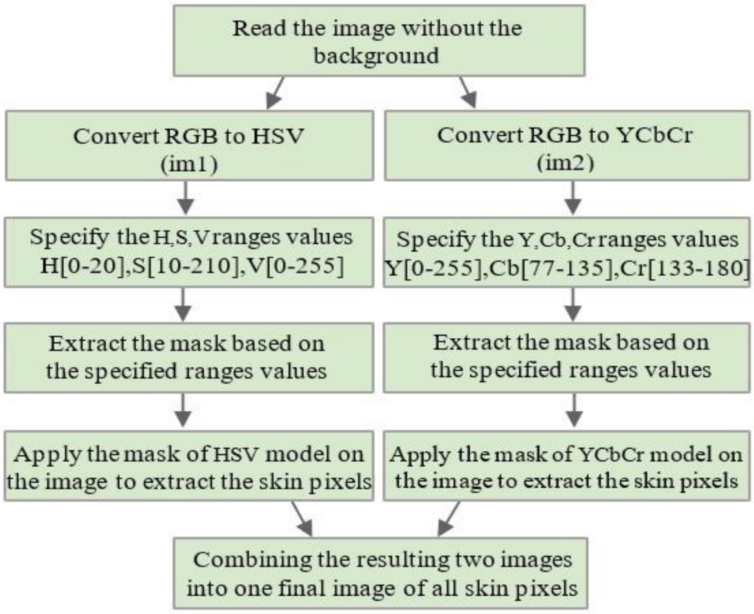


Fig. 5: Flowchart of the proposed method for skin detection

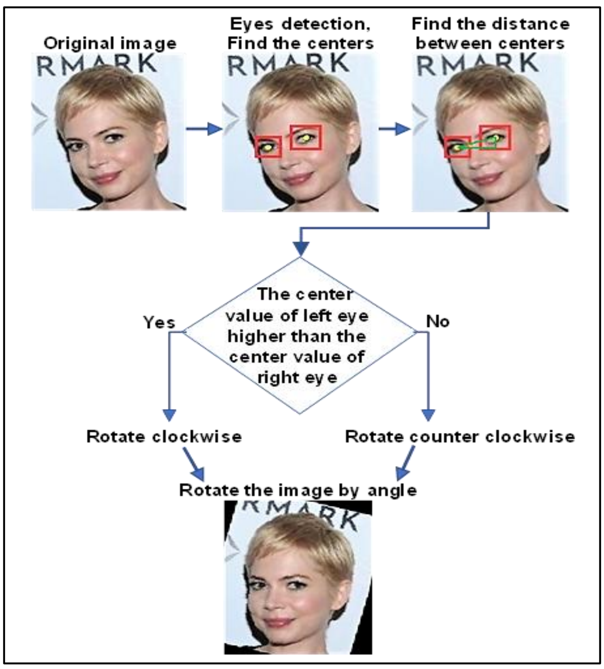


Fig. 6: Flowchart of the proposed method for face alignment

Table 1: The parameters used in the proposed wavelet scattering model for gender classification for faces

Layer (type)	Output shape	# Parameters
The input layer	(28,28,1)	0
Scattering 2D	(81,7,7)	0
Flatten	(3969)	0
Dense	(512)	2032640
Dense	(1)	513

Total params: 2,033,153
Trainable params: 2,033,153
Non-trainable params: 0

Table 2: The parameters used in the proposed wavelet scattering network for gender classification for faces

Layer (type)		Output shape	# Parameters
	The input layer	(88,45,1)	0
Block 1	Conv2D (16 filters)	(88,45,16)	160
	Conv2D (16 filters)	(88,45,16)	2320
	MaxPooling2D	(44,23,16)	0
Block 2	Conv2D (32 filters)	(44,23,32)	4640
	Conv2D (32 filters)	(44,23,32)	9248
	MaxPooling2D	(22,12,32)	0
Block 3	Conv2D (64 filters)	(22,12,64)	18496
	Conv2D (64 filters)	(22,12,64)	36928
	MaxPooling2D	(11,6,64)	0
	Flatten	(4224)	0
	Dense	(512)	2163200
	Dropout	(512)	0
	Dense	(1)	513

Total params: 2,235,505
Trainable params: 2,235,505
Non-trainable params: 0

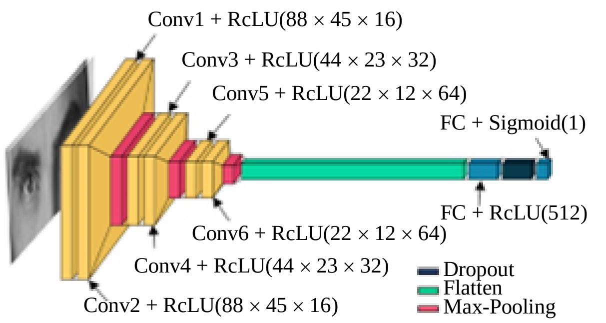


Fig. 7: The architecture of the proposed CNN model for gender classification for eyes

Classify the Gender from the Eyes Images

We proposed employing the haar cascades classification method to detect the eye. This method detects and crops only one eye. Thus, to crop the region of both eyes, we proposed first measuring the detected eye's size and then doubling the region's size in the direction of the x-coordinate. If the right eye is detected, the cropping direction will be to the left and vice versa. After cropping the eyes pair image, we segmented it into left and right eye images because we proposed to detect the gender using a pair of eyes and one eye.

Before submitting the eye images to the classification model, the images will be resized to (88 × 45) for pair eye images and (50 × 50) for one-eye images and converted into grayscale images.

A Convolution Neural Network (CNN) model was proposed for classifying the eye as belonging to male or female. The architecture of the proposed CNN used when the input image contains a pair of eyes or one eye is demonstrated in Fig. 7. The difference in applying this architecture with one or a pair of eyes is the size of the input image, where the size of one eye image is (50 × 50) and the image size for the pair of eyes is (88 × 45). The proposed CNN parameters are listed in Table 2.

Results and Discussion

The Dataset Used

In this study, the celeb faces attributes dataset (CelebA) was used. It is a large-scale face dataset with more than 200 K celebrity images. Other than that, the images in this dataset cover large pose variations and background clutter. In addition, CelebA has large diversities, large quantities, and rich annotations (Liu *et al*., 2018). The number of images used in this study was 35,928, balanced between two classes (male and female). For the training network, we used 70% of the images and 20% for validation. Note that the remaining 10% was utilized for testing. Figure 8 illustrates a sample of face images from the dataset. For the facial model, only 8,888 of the images were used.

For the network's training on detecting the gender from the eye, we used the same CelebA dataset and cropped the eye, Fig. 9. Similarly, we segment these images into left and right eyes to be entered into the model of one-eye images. While testing, a new dataset created for masked faces is used. The region of the eye is cropped from these images, as presented in Fig. 10.

Remove the Background

Deep Labelling Version 3 plus (DeepLabV3+) is employed for background removal. This network is trained on human image segmentation. DeepLabV3+ works on changing each pixel into specific colors, such as black, white, etc. Table 3 displays a sample of the result when using DeepLabV3+.

Fig. 8: Sample images from the CelebA dataset

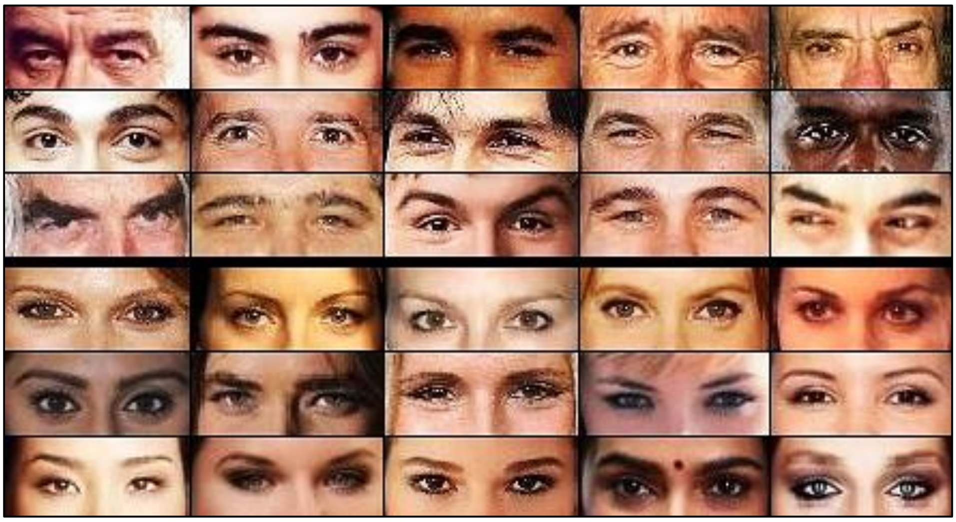


Fig. 9: Sample of images from the eyes pair dataset

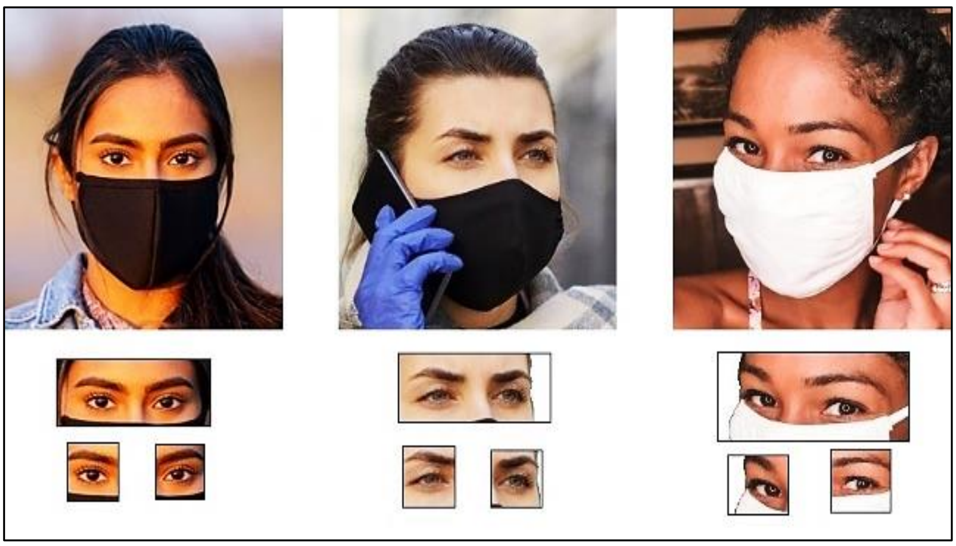
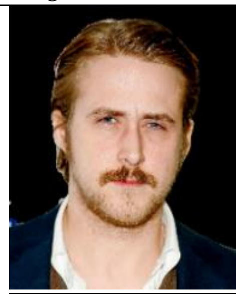


Fig. 10: Eyes segmentation for masked faces

Table 3: Removing the background by DeepLabV3+

Image with background	Extract the human masks	Image without the background
		
		
		

Skin Detection

As we mentioned, skin detection is achieved using two color spaces (HSV) and (YCbCr). Hence, the image result from the previous step will be transformed into HSV and YCbCr color space. Correspondingly, the skin is detected from each color space and the results of the two images are combined to get the final skin region of the input image. A sample of the results is shown in Table 4. Two color spaces are used to eliminate the skin detection drawback of each one by the other color space.

Face Detection

The process of face detection is done using the Haar cascade method. Nevertheless, the Haar cascade method cannot always detect the faces in the images. Moreover, this method contains some drawbacks. For example, it detects only the frontal, fully visible faces and gives false detection of different areas as a face, as demonstrated in Fig. 11. Thus, we proposed the skin detection method stage before face detection to improve the face detection process.

Face Alignment and Cropping

Aligning the face depends on using the Haar cascade method for detecting the eyes. Samples of the results are presented in Table 5. The difficulties of aligning the face arise when the face is wearing sunglasses because it hides the eyes, unidentifiable of the center. Other than that, the cropping of the face will be done to include all the face pixels.

The reason for using relation (4) Instead of relation (5) is shown in Fig. 12, where some pixels are turned off (with black color) because when multiplying with sine and cosine, the value of x and y (the new pixel location) must be an integer. So, sometimes repeated values will result, or some locations will never result, which leads to the black dots in the image.

Choosing the Optimum Architecture for the Proposed Wavelet Scattering Model

A test is done for the proposed wavelet scattering model to classify the gender from face images using 888 images. The accuracy achieved was 98.09% for both male and female classes.

The facial model's training and test accuracy value and loss function are shown in Fig. 13-14, where the training accuracy was 99.97% and the testing accuracy (validation) was 97.87%.

DOI: 10.3844/jcssp.2023.345.362

Fig. 11: Samples of not detected faces by the Haar-Cascade method

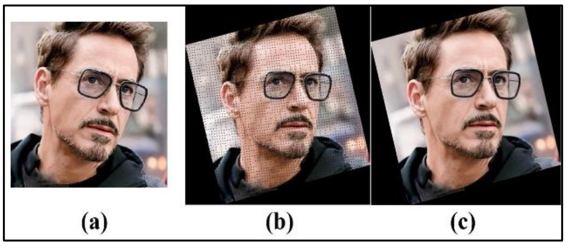


Fig. 12: (a) Original image, (b) The rotated image with relation (5), (c) The rotated image with the three matrices, relation (4)

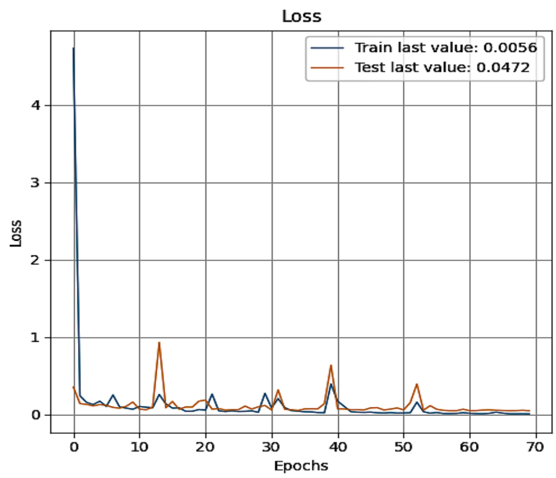


Fig. 13: The loss function for the facial model

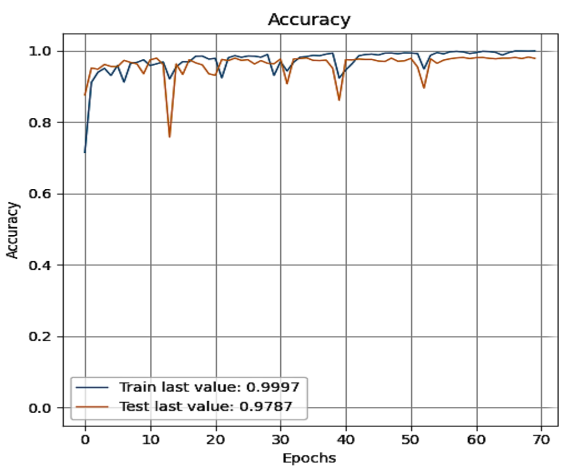


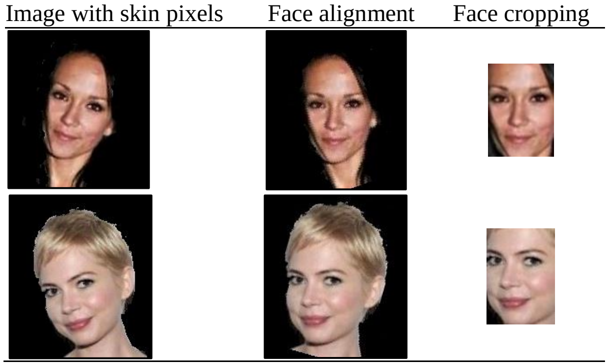
Fig. 14: The facial model accuracy while the training process

Table 4: Skin detection by the proposed method

Image without background	Skin detection by HSV	Skin detection by YCbCr	Combination HSV&YCbCr

Table 5: Face alignment and cropping

Image with skin pixels	Face alignment	Face cropping



Choosing the Optimum Architecture for the Proposed CNN Model

Several experiments were conducted to choose the optimal network architecture. This is to obtain the best accuracy and results in the gender classification process. Apart from that, these experiments include testing the best input image size, type of optimizer, number of epochs, number of steps-per-epoch, and batch size.

The first test was applied to measure the best image size for input into the proposed Convolution Neural Network (CNN). The result is illustrated in Fig. 15, where the image size (88 × 45) was the best for eye images.

In the second test, we suggest three types of optimizers to test on the proposed models (Adam, RMSprop, and Stochastic Gradient Descent (SGD)). The accuracy achieved by each optimizer is shown in Fig. 16. Note that the best accuracy was obtained using the Adam optimizer.

The number of epochs has a significant impact on how well CNN performs. As the epoch number rises, weight changes become more frequent. Thus, finding the best number of epochs in any CNN model is significant. Moreover, various epochs are utilized to measure CNN performance, as presented in Fig. 17. For example, the best accuracy is obtained using 170 epochs for the eyes model.

Subsequently, the steps-per-epoch has a significant effect on the training of the model. A test is applied with different steps to find the best one with higher accuracy. Figure 18 indicates the best number was 20.

The last test was to determine the batch size (Fig. 19), many batch sizes are tested and the accuracy is measured for each size. This resulted in the best batch size being 128 for the eyes model.

The training and test accuracy value and loss function for the eye model is shown in Fig. 20-21, where the training accuracy was 99.8% and the testing accuracy (validation) was 98.12%.

Note that the test is done for the proposed CNN to gender classification from the eye using 3,592 face images. The proposed CNN for detecting gender from the eyes is shown in the confusion matrix in Fig. 22. Meanwhile, the testing measurements are presented in Table 6. The network classified the gender correctly for most of the input images.

Testing the Proposed Models on Challenging Faces

Train the Wavelet Scattering Model on Faces from Different Datasets

The proposed facial model was trained on five different datasets of faces (CelebA, Labeled Faces in the Wild (LFW), Adience, WIKI, and Internet Movie Database (IMDB)), which have challenging face images such as some of the faces not fully visible, not completely frontal, wearing glasses with different styles, wearing a hat, have low quality, image lighting variation, children's faces, with and without makeup, open and closed mouth and closed eyes Fig. 23. The accuracy is measured and tabulated in Table 7, with the best accuracy achieved on the IMDB dataset at 95.84%.

Another test is done for testing the proposed wavelet scattering network to gender classification of infants' faces. 400 images selected from the UTK-face dataset are used in this test (200 for each female and male class). The result of this test (confusion matrix) is shown in Fig. 24. The accuracy was 71% (Table 8). The result can regard as a promised result because the infant's faces are very challenging due to the high similarity between the infant's male and female faces, a sample of the infant's faces is shown in Fig. 25.

We also did another test to detect the face gender from the facial images that were contaminated with noise, Fig. 26 shows a sample of these images. We found that noise does not affect gender classification.

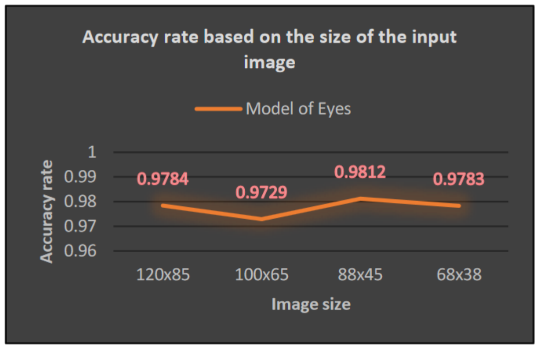


Fig. 15: The accuracy results when the proposed CNN was tested with different eye image sizes

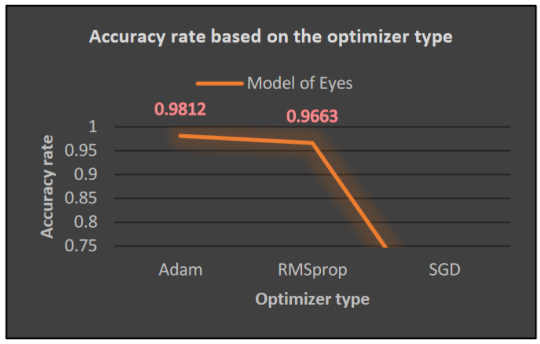


Fig. 16: The accuracy results when the proposed CNN was tested with different types of optimizers

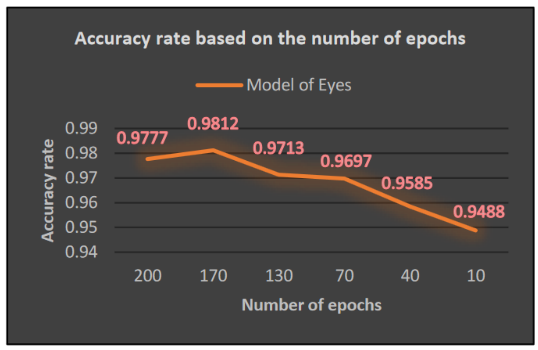


Fig. 17: The accuracy results when the proposed CNN was tested with different numbers of epochs

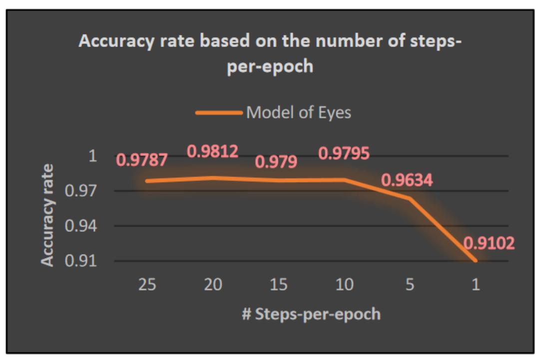


Fig. 18: The accuracy results when the proposed CNN was tested with different numbers of steps per epoch

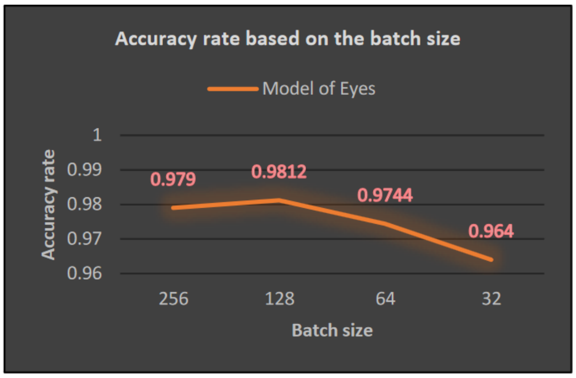


Fig. 19: The accuracy results when the proposed CNN was tested with different numbers of batch sizes

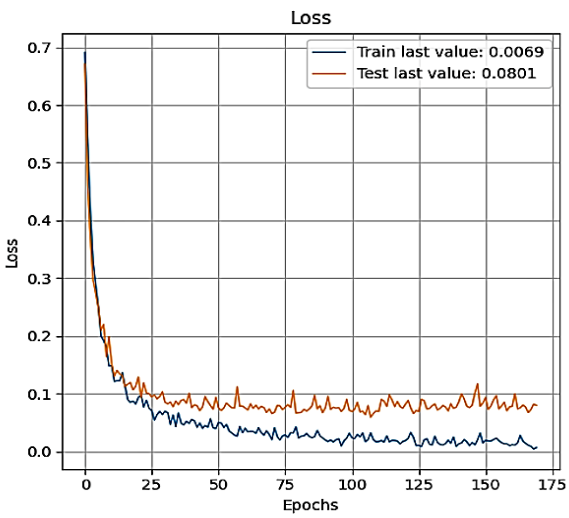


Fig. 20: The eyes model loss function

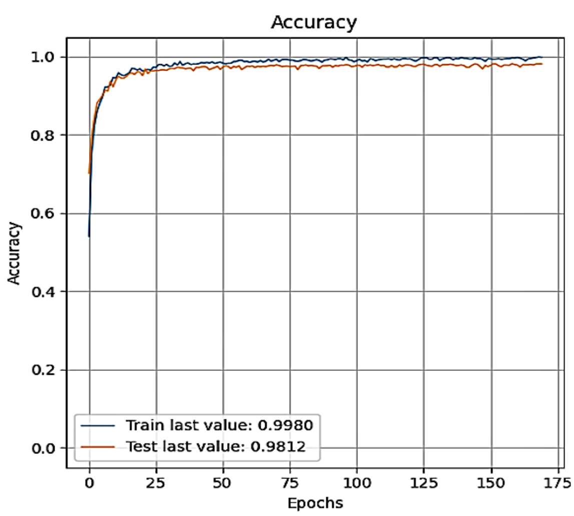


Fig. 21: The eyes model accuracy while the training process

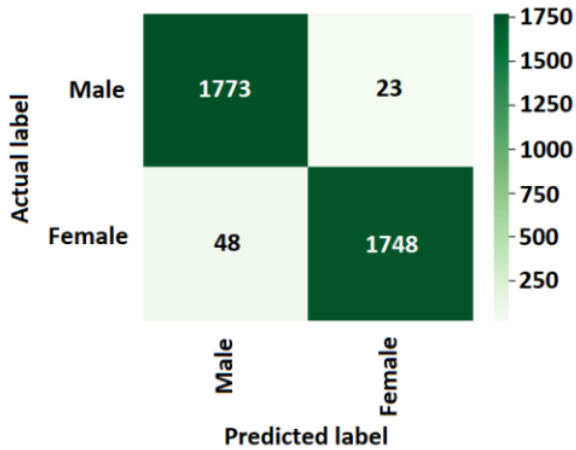


Fig. 22: Confusion matrix of the gender classification model from eyes

Fig. 23: Sample images of faces from the CelebA, LFW, Adience, WIKI, and IMDB datasets with challenges

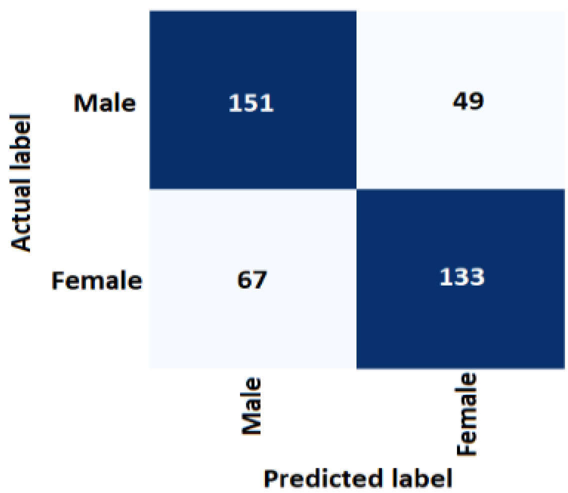


Fig. 24: Confusion matrix of the gender classification model from infant's faces

Train the CNN Model on Eyes from Different Datasets

We presented the proposed CNN model for gender classification from eyes (both eyes and one eye) trained in the four datasets (CelebA, Female and Male, LFW, Adience). The images have almost the same challenges, except for the eyes wearing sunglasses. As illustrated in Table 9, the best accuracy was achieved on the CelebA dataset at 97.14% for one eye and 96.1% for both eyes.

Other than that, in the context of testing the robustness of the suggested method, many experiments have been conducted on the proposed method for detecting the skin. The results have indicated that it is suitable for most nationalities and races. Table 10, the method detected the skin and face area of the Indian, Arabian, American, Chinese, and Nigerian images. We were able to make this algorithm suitable for most skin types and tones by experimenting with different ranges of tones employing the two-color models, HSV and YCbCr.

Another test was conducted to detect faces from images with more than one person and the results were very promising, as displayed in Table 11.

Comparing the proposed method with some of the previous techniques for gender classification is listed in Table 12 for face images and Table 13 for eyes images.

Fig. 25: Sample images of infant faces from the UTK-face dataset (F-female, M-male)

Fig. 26: Sample images of faces with noise

Table 6: The results of the gender classification models from eyes

Class	Precision %	Recall %	F1-score %	Support
Male	97.36	98.71	98.03	1,796
Female	98.70	97.32	98.00	1,796
Accuracy		98.02		3,592
Marco avg	98.30	98.01	98.01	3,592
Weighted avg	98.03	98.01	98.01	3,592

Table 7: The training results of the gender classification using the proposed wavelet scattering model with challenging face images from different datasets

Dataset	CelebA	LFW	Adience	WIKI	IMDB
# Images	8,888	8,888	8,888	8,888	8,888
Validation accuracy	94.63%	93.34%	88.81%	88.50%	95.5%
Testing accuracy	95%	93.62%	91%	90.2%	95.84%

Table 8: The results of the gender classification from infant's faces

Class	Precision %	Recall %	F1-score %	Support
Male	69.26	75.5	72.25	200
Female	73.08	66.5	69.63	200
Accuracy		71.0		400
Marco avg	71.17	71.0	70.94	400
Weighted avg	71.17	71.0	70.94	400

Table 9: The training results of the gender classification model on challenging eye images from different datasets

Dataset	CelebA	Female and male	LFW	Adience
		One eye		
# Images	71,856	11,525	22,824	36.366
Validation accuracy	97.49%	95.01%	92.46%	88.01%
Testing accuracy	97.14%	93.54%	89.65%	72.13%
		Both eyes		
# Images	34,525	-	11,416	18,183
Validation accuracy	96.32%	-	93.19%	88.42%
Testing accuracy	96.1%	-	93.1%	83.34%

DOI: 10.3844/jcssp.2023.345.362

Table 10: Face and eyes segmentation steps. (a) The input image, (b) The image without the background, (c) Skin detection by HSV, (d) Skin detection by YCbCr, (e) Combining the resulting image from a and b, (f) Crop the face, (g) and (h) Crop the eyes

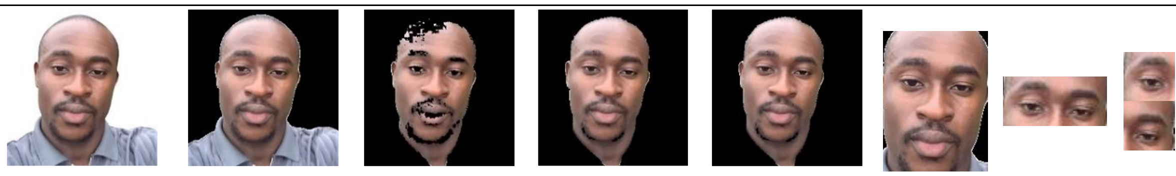
Race	a – h
Indian	
Arabian	
American	
Chinese	
Nigerian	

Table 11: The application of the proposed method to an image containing more than one person

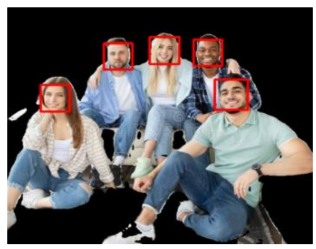
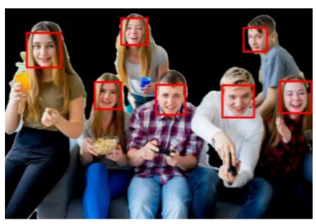
Original image	Remove background	Skin detection	Face detection
			
			

Table 12: Comparison of the gender classification accuracy from face images

Method	Dataset	Accuracy (%)
Pre-trained CNN, Pre-trained AlexNet (Tilki *et al*., 2021)	Adience	92.40,90.50
Deep CNN (Afifi and Abdelhamed, 2019)	LFW	95.98
Deep CNN (Sumi *et al*., 2021)	Kaggle dataset, Nottingham scan	97.44,90
Deep CNN (Yildiz *et al*., 2021)	Adience, VGGFace2	85.52,93.71
Pareto frontier CNN (Islam *et al*., 2020)	WIKI-cleaned	90
Pre-trained CNN (Zhou *et al*., 2019)	Adience	93.22
VGGNet arch (Dhomne *et al*., 2018)	Celebrity faces	95
Hyper face (Ranjan *et al*., 2017)	CelebA, LFWA	98,94
Face tracer (Kumar *et al*., 2008)	CelebA, LFWA	84,91
Deep CNN (Kamaru, 2020)	CelebA, LFWA	96,95
Deep CNN (Benkaddour *et al*., 2021)	WIKI, IMDB	93.56,94.49
CNN + ELM (Extreme Learning Machine) (Micheala and Shankar, 2021)	Adience	90.2
LMTCNN (Lightweight Multi-task CNN) (Lee *et al*., 2018)	Adience	85
Wide CNN + Gabor Filter (Hosseini *et al*., 2018)	Adience	88.9
2DPCA on real Gabor space + SVM (Rai and Khanna, 2015)	LFW	89.1
Our proposed deep wavelet model	CelebA-cleaned, CelebA	98.09,95
	LFW	93.62,91
	Adience	91
	WIKI	90.2
	IMDB	95.84

Table 13: Comparison of the accuracy of gender classification algorithms for eye images with our results on different datasets

Method	Dataset	Accuracy one eye (%)	Accuracy both eyes (%)
Pre-trained NASNetLarge, Pre-trained Xception (Cimtay and Yilmaz, 2021)	Female and male	95.85,96.98	-
CNN model (Alonso-Fernandez *et al*., 2021)	Adience	76.6	78.9
SRCNNs (Tapia *et al*., 2019)	Selfies images	90.15 R, 87.15 L	-
Pre-trained CNN + SVM (Alonso-Fernandez *et al*., 2020)	LFW	92.6	93.4
CNN model (Rattani *et al*., 2018)	Ocular images	89.01	-
Our proposed CNN model	CelebA-cleaned, CelebA	97,97.14	98.02, 96.1
	Female and male	93.54	-
	LFW	89.65	93.1
	Adience	72.13	83.34

Conclusion

In this study, we classified gender from facial and eye images by proposing a method consisting of six basic stages. In the second stage, we proposed a new method that detects skin regions for most human races (Indian, Arabian, American, Chinese, Nigerian) with different skin colors and various numbers of humans in the image. The result was excellent, as shown in the result section. Correspondingly, the skin detection method has improved the work of the haar cascade method for face detection by decreasing the challenges in the image. Using two color spaces has a high impact on improving skin detection methods. In this study, we detect gender even with occlusion faces. For the gender classification from the eye (masked face), we suggested a Convolution Neural Network (CNN) model. In addition, we proposed a deep wavelet network to classify the faces, which is one of this study's contributions. We also showed that our models work well with images that have challenges (some of the faces are not fully visible, not completely frontal, wearing glasses with different styles, have low quality and have different lighting, children’s faces, with and without makeup, open and closed mouth, closed eyes, noisy image, and infants’ face). Finally, a comparative analysis has been done on several studies with our proposal that demonstrated the robustness and efficiency of the proposed method. In the future, it is planned to develop a group recommendation system for a group of users in public places.

Acknowledgment

There is no conflict of interest and no financial support from any other party.

Funding Information

The authors have not received any financial support or funding to report.

Author's Contributions

Zahraa Salam Abu Ghrban: Software, formal analysis, investigation, resources, data curation, written original draft preparation.

Nidhal Khdhair EL Abbadi: Conceptualization, methodology, visualization, supervision, project administration, review, and edited of the final paper.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and that no ethical issues are involved.

References

Afifi, M., & Abdelhamed, A. (2019). Afif4: Deep gender classification based on an adaboost-based fusion of isolated facial features and foggy faces. *Journal of Visual Communication and Image Representation*, *62*, 77-86. https://doi.org/10.1016/j.jvcir.2019.05.001

Alonso-Fernandez, F., Diaz, K. H., Ramis, S., Perales, F. J., & Bigun, J. (2020, September). Soft-biometrics estimation in the era of facial masks. In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)* (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/9211002

Alonso-Fernandez, F., Hernandez-Diaz, K., Ramis, S., Perales, F. J., & Bigun, J. (2021). Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images. *IET Biometrics*, *10*(5), 562-580. https://doi.org/10.1049/bme2.12046

Benkaddour, M. K., Lahlali, S., & Trabelsi, M. (2021, February). Human age and gender classification using convolutional neural network. In *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)* (pp. 215-220). IEEE. https://doi.org/10.1109/IHSH51661.2021.9378708

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).

Cimtay, Y., & Yilmaz, G. N. (2021). Gender classification from eye images by using pre-trained convolutional neural networks. *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, *14*, 39-44. https://dergipark.org.tr/en/download/article-file/2163832

Dhomne, A., Kumar, R., & Bhan, V. (2018). Gender recognition through face using deep learning. *Procedia Computer Science*, *132*, 2-10. https://doi.org/10.1016/j.procs.2018.05.053

Hosseini, S., Lee, S. H., Kwon, H. J., Koo, H. I., & Cho, N. I. (2018, January). Age and gender classification using wide convolutional neural network and Gabor filter. In *2018 International Workshop on Advanced Image Technology (IWAIT)* (pp. 1-3). IEEE. https://doi.org/10.1109/IWAIT.2018.8369721

Islam, M. M., Tasnim, N., & Baek, J. H. (2020). Human gender classification using transfer learning via Pareto frontier CNN networks. *Inventions*, *5*(2), 16. https://doi.org/10.3390/inventions5020016

Kamaru, Z. F. H. (2020). Gender classification using custom convolutional neural networks architecture. *International Journal of Electrical & Computer Engineering (2088-8708)*, *10*(6).

Kumar, N., Belhumeur, P., & Nayar, S. (2008). Facetracer: A search engine for large collections of images with faces. In *Computer Vision-ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV 10* (pp. 340-353). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88693-8_25

Lee, J. H., Chan, Y. M., Chen, T. Y., & Chen, C. S. (2018, April). Joint estimation of age and gender from unconstrained face images using lightweight multi-task CNN for mobile applications. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 162-165). IEEE. https://doi.org/10.1109/MIPR.2018.00036

Liu, Z., Luo, P., Wang, X., & Tang, X. (2018). Large-scale celeb faces attributes (celeba) dataset. *Retrieved August*, *15*(2018), 11.

Mallat, S. (2012). Group invariant scattering. *Communications on Pure and Applied Mathematics*, *65*(10), 1331-1398. https://doi.org/10.1002/cpa.21413

Micheala, A. A., & Shankar, R. (2021). Automatic Age and Gender Estimation using Deep Learning and Extreme Learning Machines. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(14), 63-73.

Mohammed, S. (2020). A Study on Gender and Age Classification as the Two Most Vital Tools in the Identification and Verification System. *The Eurasia Proceedings of Educational and Social Sciences*, *17*, 57-66.

Rai, P., & Khanna, P. (2015). An illumination, expression, and noise invariant gender classifier using two-directional 2DPCA on real Gabor space. *Journal of Visual Languages & Computing*, *26*, 15-28. https://doi.org/10.1016/j.jvlc.2014.10.016

Ranjan, R., Patel, V. M., & Chellappa, R. (2017). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation and gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(1), 121-135. https://doi.org/10.1109/TPAMI.2017.2781233

DOI: 10.3844/jcssp.2023.345.362

Rattani, A., Reddy, N., & Derakhshani, R. (2018). Convolutional neural networks for gender prediction from smartphone-based ocular images. *IET Biometrics*, *7*(5), 423-430. https://doi.org/10.1049/ietbmt.2017.0171

Singh, M., Nagpal, S., Vatsa, M., Singh, R., Noore, A., & Majumdar, A. (2017, October). Gender and ethnicity classification of iris images using deep class-encoder. In *2017 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 666-673). IEEE. https://doi.org/10.1109/BTAS.2017.8272755

Sumi, T. A., Hossain, M. S., Islam, R. U., & Andersson, K. (2021). Human gender detection from facial images using convolution neural network. In *Applied Intelligence and Informatics: First International Conference, AII 2021, Nottingham, UK, July 30-31, 2021, Proceedings 1* (pp. 188-203). Springer International Publishing. https://doi.org/10.1007/978-3030-82269-9_15

Tapia, J., Arellano, C., & Viedma, I. (2019). Sex-classification from cellphones periocular iris images. In *Selfie Biometrics: Advances and Challenges* (pp. 227-242). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-26972-2_11

Tilki, S., Dogru, H. B., Hameed, A. A., Jamil, A., Rasheed, J., & Alimovski, E. (2021). Gender classification using deep learning techniques. *Manchester J. Artif. Intell. Appl. Sci*., *2*(1), 126-131.

Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (Vol. 1, pp. I-I). IEEE. https://doi.org/10.1109/CVPR.2001.990517

Yildiz, K., Güneş, E., & Anil, B. A. S. (2021). CNN-based Gender Prediction in Uncontrolled Environments. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *9*(2), 890-898. https://doi.org/10.29130/dubited.763427

Zhou, Y., Ni, H., Ren, F., & Kang, X. (2019, August). Face and gender recognition system based on convolutional neural networks. In *2019 IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 1091-1095). IEEE. https://doi.org/10.1109/ICMA.2019.8816192