Multimedia Systems (2023) 29:421–433
https://doi.org/10.1007/s00530-022-01011-8

REGULAR PAPER

Gender estimation based on deep learned and handcrafted features in an uncontrolled environment

Sahar Dammak[1] · Hazar Mliki[2] · Emna Fendri[3]

Received: 15 March 2022 / Accepted: 29 September 2022 / Published online: 8 October 2022

Abstract
Automatic gender estimation provides a valuable information in the face of analysis tasks and has been widely used in many fields of applications like human–computer interaction, biometrics, video surveillance, activity recognition. This paper introduces a new facial gender estimation method based on a hybrid architecture which combines deep learned features as global information and handcrafted features as local information. A Min Redundancy Max Relevance algorithm was applied to select the highest relevant and lowest redundant features. Such process provides a good compromise in terms of speed and accuracy rate. The experimental study was conducted on the Image Of Groups and FERET databases. The obtained results proved the efficiency of the proposed method in dealing with the facial gender estimation task in an uncontrolled environment.

Keywords Gender estimation · Deep learning · VGG-16 · MB-LBP · SVM · Mutual information

1 Introduction

Automatic facial gender estimation has been one of the interesting and challenging research topics recently. Therefore, it has been widely used in several fields of applications like human–computer interaction, biometrics, video surveillance, activity recognition. Robust gender estimation in an uncontrolled environment remains a thorny problem due to the appearance variation of faces affected by occlusion, pose variation, low resolution, scale variation, illumination variation, among others.

In the literature, most of the gender estimation methods used handcrafted features followed by a classifier and they are commonly known as handcraft-based methods Zhang et al. [50], Mohamed et al. [31], Sheetlani et al [44], Chen and Jeng [5], Fekri-Ershad [16], Irhebhude et al [24]. The recent success of the deep learning (CNNs) has made the research community apply it on gender estimation Aslam et al. [4], Serna et al [42]. Unlike the handcraft-based methods, where features are extracted by predefined descriptors, the CNN learns the features automatically from the input facial images.

This paper presented a new hybrid facial gender estimation method in an uncontrolled environment that combines deep learned and handcrafted features. The hybrid architecture takes benefit from their advantages: the deep learned features are based on the pre-trained VGG-16 model Simonyan and Zisserman [46] to extract the deep features which contain global information while the handcrafted features are based on the Local Binary Pattern (LBP) Ojala et al. [32] to encode local information from the input facial images. In fact, a feature-level fusion which consists in combining the deep learned and handcrafted features was proposed. Then, a features selection technique based on mutual information was applied to select the highest relevant and lowest redundant features. Finally, the selected features were used to train a Support Vector Machine (SVM) classifier Cristianini and Shawe-Taylor [7].

The main contributions of this work can be summarized as follows:

Communicated by Changsheng.

✉ Sahar Dammak
sahardammak@fsegs.u-sfax.tn

[1] MIRACL-FSEG, University of Sfax Faculty of Economics and Management of Sfax, Road Airport Km 4, 3018 Sfax, Tunisia

[2] MIRACL-ENET'COM, University of Sfax National School of Electronics and Telecommunications of Sfax, Road Tunis city El Ons, 3018 Sfax, Tunisia

[3] MIRACL-FS, University of Sfax Faculty of Sciences of Sfax, Road Sokra Km 3, 3018 Sfax, Tunisia

- Introducing a robust facial gender estimation method in an uncontrolled environment able to handle the face gender variations namely: pose variation, low resolution, occlusion, illumination variations, scale variations.
- Enhancing the gender estimation by combining global and local informations at the feature-level. The global features are provided from deep learned features using the VGG-16 model. As for the local features, they are extracted using the Mb-LBP descriptor. Such combination will afford fruitful feature vector able to encode the face appearance variability.
- Applying the Min Redundancy Max Relevance algorithm to select the highest relevant and lowest redundant features. Such process helps reduce the memory space and the complexity of the proposed method.

The remainder of this paper is organized as follows. Section 2 reviewed the related works. The proposed method was introduced and detailed in Sect. 3. The experiments and results were illustrated in Sect. 4. Finally, Sect. 5 provided the conclusion and some perspectives for future research works.

2 State of the art

In the literature, many methods have been proposed for gender estimation. The facial gender estimation methods consist of two main phases: features extraction and features classification. Relying on the way of the features are extracted, the existing methods were classified into three categories: handcrafted features based methods, deep learned features based methods and hybrid methods.

2.1 Handcrafted features-based methods

The handcrafted features based methods apply standard descriptors such as Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT) Lowe [29], Histogram of Oriented Gradients (HOG) Dalal and Triggs [9] to encode facial features. These features are then used to train different classifiers such as neural network Hecht-Nielsen [21], SVM, KNearest Neighbor (KNN) Dokmanic et al. [13], etc.

In Zhang et al. [50], the authors proposed an approach based on multi-scale facial fusion features (MS3F) extracted by combining the LBP and the Local Phase Quantization (LPQ) descriptors Ojansivu and Heikkilä [33]. Then, the authors generated the multi-scale features through Multi-block (MB) and Multilevel (ML) methods. Finally, an SVM classifier was applied to identify the gender class.

As for Mohamed et al. [31], the Discrete Cosine Transform (DCT) Chitprasert and Rao [6] and the Discrete Wavelet Transform (DWT) Sidney Burrus et al [45] were combined to extract discriminant features which encode the face appearance and geometry. Thereafter, the KNN, fuzzy of KNN Keller et al [25] and SVM are used to get the gender label.

In the same context, Sheetlani et al. [44] proposed a gender estimation method based on multi-resolution statistical descriptors derived from the histogram of the DWT. The authors applied the DWT to the image to extract a multi-resolution features. Finally, three classifiers, namely Nearest Neighbor, Support Vector Machine and Linear Discriminant Analysis were used to identify the face gender.

In Irhebhude et al. [24], authors applied the HOG and RILBP descriptors to extract features. Then, the discriminant ones was selected using the PCA method. Finally, the SVM classifier was applied to identify the gender class.

2.2 Deep learned features-based methods

The deep learned features methods are based on CNN architectures which are usually composed of two phases: the features extraction phase and the features classification phase. The features extraction phase includes independent processing layers (Convolution, Pooling, RELU) allowing to extract automatically relevant features. The obtained features are used as an input for the features classification phase Dwivedi and Singh [15], Afifi and Abdelhamed [1].

In Dwivedi and Singh [15], a review is proposed on the use of the CNN for gender estimation. The authors carried out a comparative study by changing some parameters (convolutional layer number, filter number, filter size, training image number, softmax layer number, etc.) in Alexnet Krizhevsky et al. [27] architecture to choose the most efficient architecture.

Afifi and Abdelhamed [1] introduced a gender estimation method based on the combination of isolated facial features and a holistic features (foggy face). The authors extracted five face patches, namely: foggy face, nose, mouth and left eye and right eye. Then, each patch is fed as an independent input to a pre-trained CNN dedicated to classify this face patch. Finally, a score fusion method based on AdaBoost Freund and Schapire [17] was proposed to combine the independent classification scores to infer the final decision.

As for Aslam et al. [3], a Cascaded Deformable Shape model was used to extract facial feature regions from a facial image namely: the eyes, the nose, the mouth and the foggy faces. Then, a four-dimensional (4-D) representation was constructed using these facial feature regions. Finally, the VGG-16 pre-trained model was fine-tuned using this 4-D array for a final gender decision.

In the same context, Ryu et al. [39] proposed a gender and race classification solution based on the FaceNet model Schroff et al. [40] followed by an avgpool layer. Then, fully

connected layers were added to identify the gender and race classes.

Sharma et al. [43] proposed a convolutional neural network (CNN) based method for face gender and age estimation. The proposed CNN architecture reduces the number of operations needed to process an image to improve the computational performance. Therefore, a pre-processing is proposed to resize the images and store the annotation (age and gender information). Then, the obtained images are used to train the proposed CNN model which consist of convolutional layers followed by a Relu layer. Each convolutional block is followed by a max-pooling layer. In addition, the authors used a fully connected layer output with a sigmoid function followed by a Softmax jayer that gives the probability for each class of gender and age.

In Serna et al. [42], the authors present a preliminary analysis of how biased data affect the learning process of deep neural network architectures in terms of activation level. In fact, a study of the ethnic effect on the learning process of gender classifiers was performed. To evaluate the bias effect on the learning process, two gender detection models based on VGG16 and Resnet are used.

In Amri et al. [2], the authors provide a comparative experimental study of the significance of each part of the face (eyes, mouth, nose) in the gender facial recognition via convolutional neural networks (CNN). The objective of the proposed method is to find the most crucial part of the face to determine the most important part in the gender recognition task. Then, a second study on the degree of importance of the eyes for both genders was performed by training the CNN model using only eyes.

2.3 Hybrid methods

The hybrid methods consist of combining handcrafted features-based methods with those based on deep learning methods to take advantage of their strengths. Different strategies combining handcraft with deep learning based methods have been studied in the literature such as: feeding the CNN with handcrafted information Hosseini et al [22], Aslam et al. [4] or classifying the deep features extracted from the CNN with a standard classifier Duan et al. [14].

Duan et al. [14] proposed a hybrid CNN–ELM method to achieve age and gender estimation. The authors combined the proposed Convolutional Neural Networks (CNN) with the Extreme Learning Machine (ELM) Huang et al. [23] in one network and integrated the synergy of two classifiers to estimate the age and gender. This network consists of two phases: features extraction and classification. The CNN convolutional layer is used for the features extraction. Then, the authors merged the ELM with the fully connected layers to obtain a hybrid age and gender model.

To improve the performance of the pre-trained Alexnet, Hosseini et al. [22] applied a Gabor filter Petkov [35] to the input images to extract the handcrafted features. Finally, the weighted sum of the input images and Gabor responses are used as an input to the CNN to generate a gender estimation model.

As for Aslam et al. [4], the authors performed an Inter-Component Transform (ICT) on the images to transform them into YCbCr images and preserve just the Y component (luminance), on which, the DWT was used, so as to obtain the low-resolution subband image. The acquired images are then used as input to the pre-trained Alexnet model to create a gender prediction model.

Rwigema et al. [38] proposed a hybrid method which combines the decisions obtained from two neural networks to increase the accuracy of age and gender estimation. The first neural network consists of a Gabor filters for the features extraction and an SVM for classification. As for the second neural network, it is a convolutional neural network (CNN). Then, a sum rule decision fusion was used to combine the decision obtained from the SVM with those obtained from the CNN model. Authors in Dammak et al [10] proposed an age classification method that consists in classifying human faces into different age groups. The proposed method aims to explore the correlation between age and gender information. The used gender estimation method is a hybrid method which consists in applying a score-level fusion of deep learned and handcrafted models.

2.4 Discussion

Through this brief review, the gender estimation methods were classified into three categories: handcrafted features based methods, deep learned features based methods and hybrid methods. The handcrafted features based methods apply discriminant descriptors to encode low level information such as texture, shape, curve or edge features. Nonetheless, their accuracy depends on the choice of descriptors and databases constraints. Regarding the deep learning methods, their results significantly outperform those reported by the handcrafted features based methods [52]. Indeed, the deep learning-based methods are known for their effectiveness in capturing complex visual variations by leveraging a large amount of training data [49]. In addition, they do not require any choice of a specific descriptor. Lots of studies have shown that features extracted from deep learning methods contain more global information than handcrafted features based methods thanks to the large depth of deep learning network [28]. The drawback of having too many network layer is the loss of lower layer information (*i.e.* color, texture, shape and edge). The hybrid methods, however, combine the strengths of handcraft-based methods and deep learning methods. Particularly, global and local informations

are combined. Previous studies in some classification systems [19, 20, 26] have shown that features fusion strategy improves significantly the classification performance. Such improvement helps to deal with various constraints such as: pose, illumination, and scale variations as well as low resolution and occlusion.

In the light of the above discussion, we proposed a new hybrid method for gender estimation based on the combination of global information with local information in an uncontrolled environment.

3 Proposed method for gender estimation

The proposed gender estimation method consists in developing a new method based on a hybrid architecture that combines information from both handcrafted and deep learned features. For the handcrafted features, they were extracted using the LBP. As for the deep learned features, the pre-trained VGG-16 model was adapted to the gender estimation task to automatically extract features. Then, a features selection technique based on mutual information was applied to the combined features. Finally, the selected features were used to train the SVM classifier.

As illustrated in Fig. 1, the proposed method consists of four main steps: (1) Face Preprocessing, (2) Face features extraction, (3) Face features selection and (4) Face features classification.

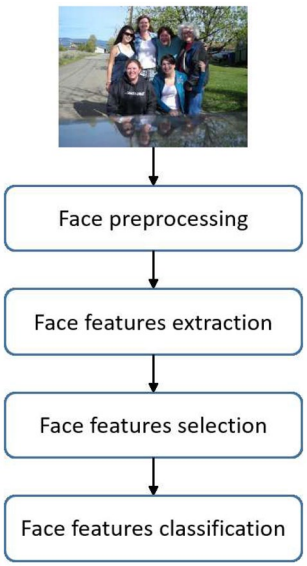

Fig. 1 Proposed gender estimation method

3.1 Face preprocessing

The face pre-processing step is illustrated in Fig. 2. In this step, a face detection [30] was performed.

To improve the quality of the detected face image, histogram equalization [11] was applied to stretch the contrast of the image and reduce illumination variation in the face images.

Then, the Gabor filter responses were extracted from the equalized image. The Gabor filter is a complex plane wave (a 2D Fourier basis function) multiplied by an origin-centered Gaussian. Indeed, the idea underlying the choice of the Gabor filter is the fact that this filter can tackle the face shape and orientation in multiview face images which are quite important features for the problem of gender estimation. In addition, the Gabor filters multi-resolution and multi-orientation was used to identify the local information with prominent features. A Kernel of 2D Gabor filter function [35] is expressed as follows:

$$G(x,y) = \exp\left(-\frac{x'^2 + \gamma y'^2}{2\sigma}\right)\cos\left(2\pi\frac{x'}{\lambda} + \phi\right), \tag{1}$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$, λ is the wavelength of the real part of Gabor filter kernel, θ is the orientation of the stripes of the function, ϕ is the phase offset, γ is the spatial ratio and σ is the standard deviation.

Next, the weighted sum of the Gabor filter responses (WG) and the equalized image (WI) was computed to generate a new enhanced image.

3.2 Face features extraction

For the face features extraction, the pre-trained VGG-16 was used to extract the deep learned features and the MB-LBP to encode the handcrafted features.

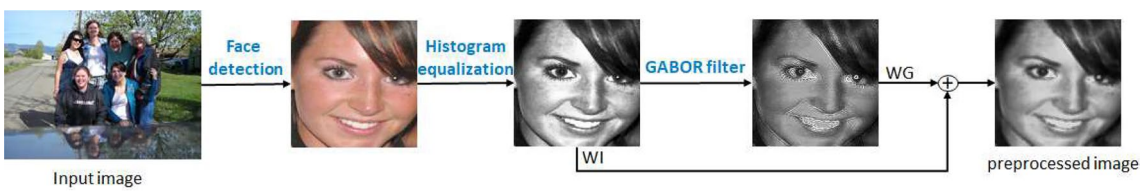

Fig. 2 Face preprocessing step

3.2.1 Deep learned features

To extract deep feature vector,a CNN model based on the pre-trained VGG-16 which has shown good results on the ImageNet challenge [37] was used. The VGG16 is considered as one of the most performant deep feature extractors [34] and it achieved good results to deal with the gender estimation [3]. To adapt the model to the gender estimation task, the pre-trained VGG-16 was fine-tuned and optimized with stochastic gradient descent. In addition, a Dropout with ratio 0.5 was applied after each Fully Connected (FC) layer to ovoid the problem of overfitting. To overcome the small training set limitation, the weights was frozen in the five first convolutional layers and train the rest of the layers. The data augmentation was also used to train the model, which is a frequent and crucial preprocessing step for CNN-based methods to reach high performance. In fact, more face images were created to find various situations based on flipping, color casting, rotation, noise, histogram, and sigmoid to improve the classification accuracy. The fine-tuned VGG-16 architecture is shown in Fig. 3.

To better understand the insights of the fine-tuned CNN gender estimation model, a Gradient-weighted Class Activation Mapping (Grad-CAM) [41] visual explanation technique was applied. Grad-CAM generates blue and red effects, i.e., heatmap around the informative face regions. In fact, it focuses on the specific facial features that are mainly responsible for indicating human gender. Figure 4 represents the Grad-CAM visualization of some input

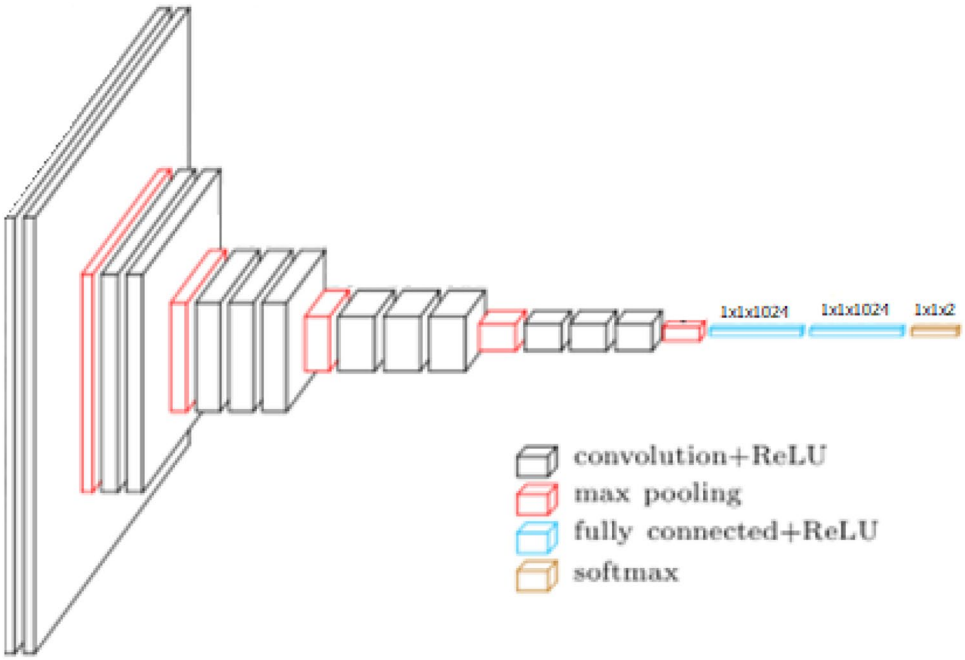

Fig. 3 Fine-tuned VGG-16 architecture

Fig. 4 Grad-CAM visualization of different gender classes

images where the areas around the mouth, nose, cheeks, and eyes that mainly affect the gender trait prediction, are highlighted for all the face images

3.2.2 Handcrafted features

For the handcrafted features, an extension of the LBP descriptor was used, called MB-LBP (multi block LBP) [51], to enhance the description ability of the LBP operator. Relying to previous study [48], this descriptor has shown improved results to face the gender estimation task. In fact, the LBP is an operator that encodes the local facial features in a multi-resolution spatial histogram and combines the distribution of local intensity with the spatial information. Thanks to its invariance to monotonic gray level variation and computational efficiency, this operator has shown high performance. This operator labels an image pixels by thresholding each pixel 3×3 neighborhood with the center value, and treating the result as a binary number. Then the histogram of the labels can be used as a texture descriptor. Equation 2 defines the LBP operator:

$$LBP_{R,P} = \sum_{i=0}^{P-1} s(g_i - g_c 2^i) \quad \text{where } s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x \leq 0 \end{cases}, \quad (2)$$

where R is the radius of the circle, P is the number of the surrounding pixels of the LBP operator, g_c is the gray level of the center pixel of the circle and g_i is the gray level of the surrounding pixels.

For the proposed MB-LBP, the facial regions' local information is retained. Therefore, the face is divided into 64 small block regions of size 16x16 pixels per block. Afterward, each region LBP histogram is extracted and concatenated into an enhanced single feature histogram. Figure 5 shows the process of feature vector extraction using the MB-LBP.

3.2.3 Feature-level fusion

To visualize the information process within CNN, the features maps extracted from the proposed CNN model were displayed. In the fine-tuned CNN model which is based on the VGG-16, there are five convolutional blocks. Figure 6 represents visualization of the different features maps extracted from the 5 blocks of the fine-tuned CNN model.

In fact, the feature maps highlight the regions sensitivity and show the detected features. In fact, the feature maps close to the input layers detect small and fine-grained details, and as far as we progress deeper into the model, the feature maps close to the output layers capture more general and global features. Thus, the model abstracts the features from the image into more global concepts that can be used for classification. The main drawback of integrating many network layers is the loss of the lower layer information (i.e. color, texture, shape and edge) illumination [28]. Therefore, the MB-LBP descriptor was applied to encode local information. Figure 7 represents the extracted facial texture feature information using the MB-LBP descriptor. These local features are robust against variations in pose and illumination [51].

The proposed feature-level fusion strategy is illustrated in Fig. 8. The deep features extracted by convolutional neural networks (CNN) and those extracted by the MB-LBP were combined.

To extract automatic features from the fine-tuned CNN models, the activation map of the last fully connected layer was considered as the deep feature vector. To extract the handcrafted feature vector, the MB-LBP descriptor was applied to encode the relevant face features. The two obtained feature vectors were normalized and concatenated into one discriminant hybrid feature vector.

3.3 Face features selection

The obtained hybrid feature vector is in 2048-dimensional space (1024 handcrafted features and 1024 deep features). Because the feature vector high dimensionality requires high processing power, the mutual information technique was applied to select the most discriminant features. As a matter of fact, the mutual information (also known as cross-entropy or gain-information) is a feature selection technique that is commonly used to evaluate the stochastic dependency of two discrete and random features [47]. The mutual information among two variables x and y is computed using their joint probabilistic distribution $p(x, y)$ and their respective marginal probabilities $p(x)$ and $p(y)$ as follows:

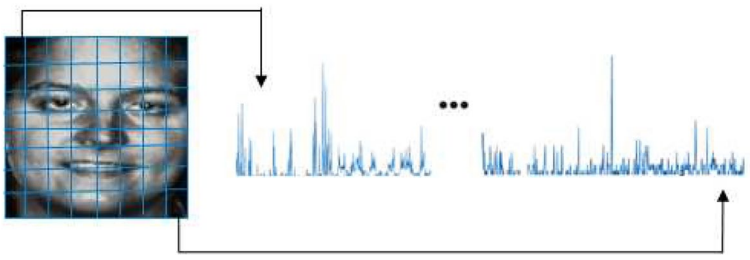

Fig. 5 Process of features extraction using the MB-LBP

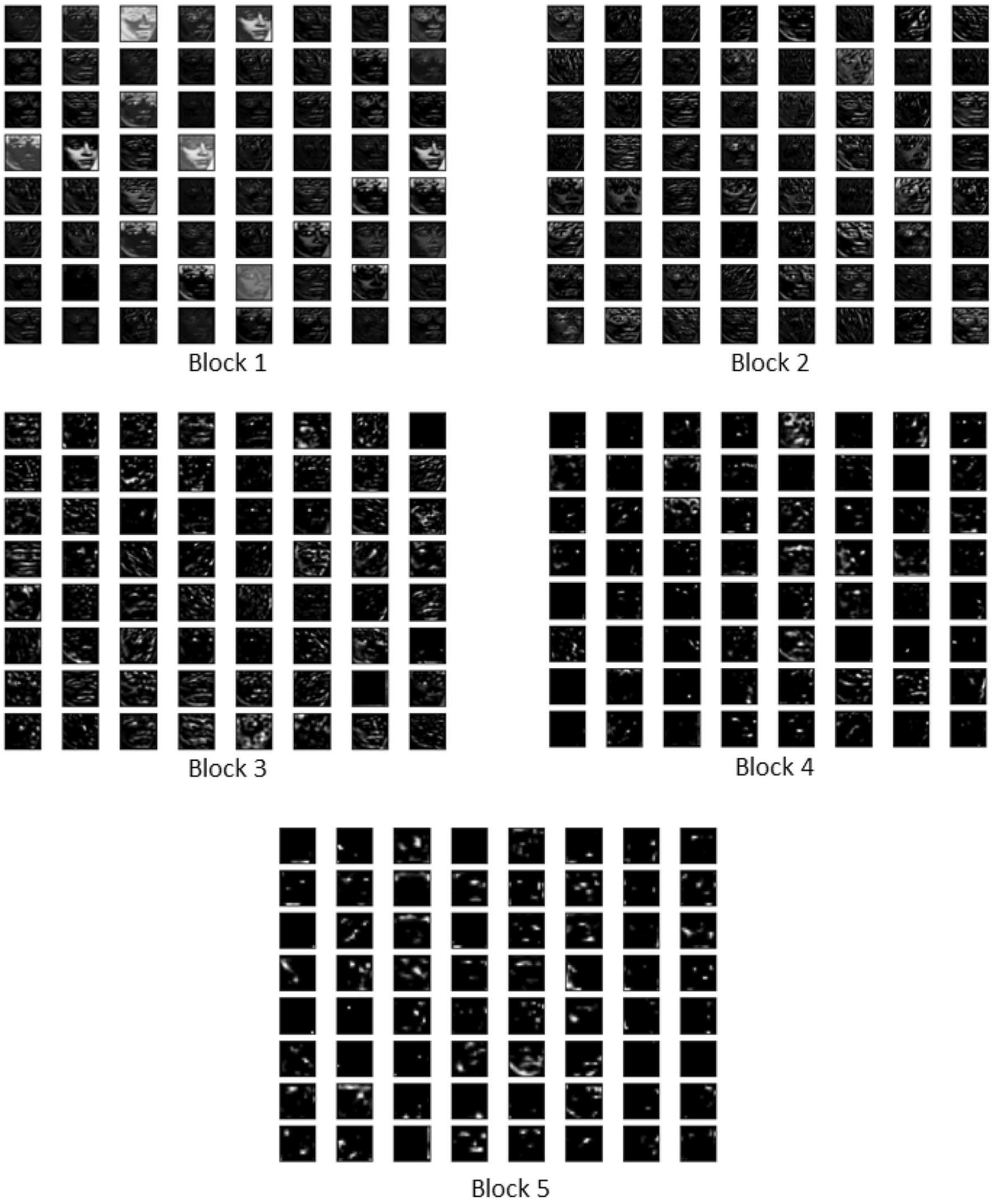

Fig. 6 Visualization of the feature maps extracted from the 5 blocks of the fine-tuned CNN model

$$I(X,Y) = \int_{\Omega y} \int_{\Omega x} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) d_x d_y, \tag{3}$$

where Ωx and Ωy are, respectively, the sample space of X and Y. Regarding $p(x)$, $p(y)$, and $p(x, y)$, they are respectively the probability density functions of X, Y, and (X, Y), respectively. The Min Redundancy Max Relevance algorithm founded on the classic mutual information statistical metrics was applied. The main aim is to reduce feature redundancy while increasing their relevance [12]. The features redundancy and relevance are calculated as follows:

$$\text{Redundancy}\,(k) = \frac{1}{|F|^2 \sum_{k,l \in F} I(k,l)}, \tag{4}$$

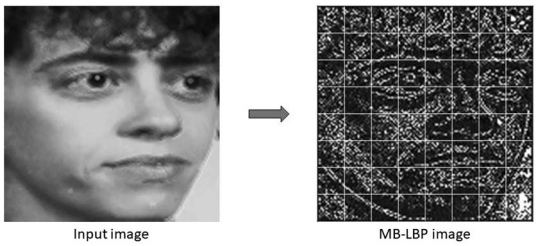

Fig. 7 Visualization of the MB-LBP image

$$\text{Redundancy} = \frac{1}{|F|^2 \sum_{k,l \in F} I(k, Y)}, \tag{5}$$

where $|F|$ is the size of features, $I(k, l)$ is the mutual information of the k and l features and $I(k, Y)$ is the mutual information between the feature k and the set of labels of Y class. Using the Min Redundancy Max Relevance algorithm (mRMR), the number of hybrid feature vectors which have the largest variation can be reduced. As a result, the feature vector with this small number of features can describe the original features. In the experiments, the number of the discriminant features is empirically fixed.

3.4 Face features classification

After the features selection, the SVM was used to classify the input facial images into male or female. In fact, the SVM method is used to find the best hyper-plane that can separate the samples of one class from those of other classes using various support vectors. For a nonlinear problem, the SVM method uses several kernel functions to map the input feature vectors to a higher-dimensional space where the problem can be separated linearly. In the proposed method, the Radial Basic Function (RBF) kernel was used, thanks to its effectiveness in the gender estimation task [8, 50].

4 Experimental study

To study the performance of the proposed method, three series of experiments were carried out on two famous datasets.

4.1 Datasets description

The proposed method was evaluated on the FERET [36] and Image Of Groups [18] datasets.

The FERET dataset is widely used in various face processing applications. It consists of 14,126 images for 1199 different persons with different captured expressions, poses, and lighting variations. In the experiments, for fair comparison with state of the art works, the color FERET dataset was used. It has 5,786 images (3,816 male images and 1,970 female images) of the frontal and the near-frontal face images (pose angle lies between −45° and +45°). Samples images from the FERET dataset are shown in Fig. 9.

The Image Of Groups (Groups) dataset contains 28,231 face images collected from Flickr images. The Groups dataset is considered, in the literature, as the most challenging and complex dataset for the gender estimation problem since it includes many constraints such as low resolution, pose variation, occlusion, luminosity variation, etc. Sample images from Groups dataset are shown in Fig. 10.

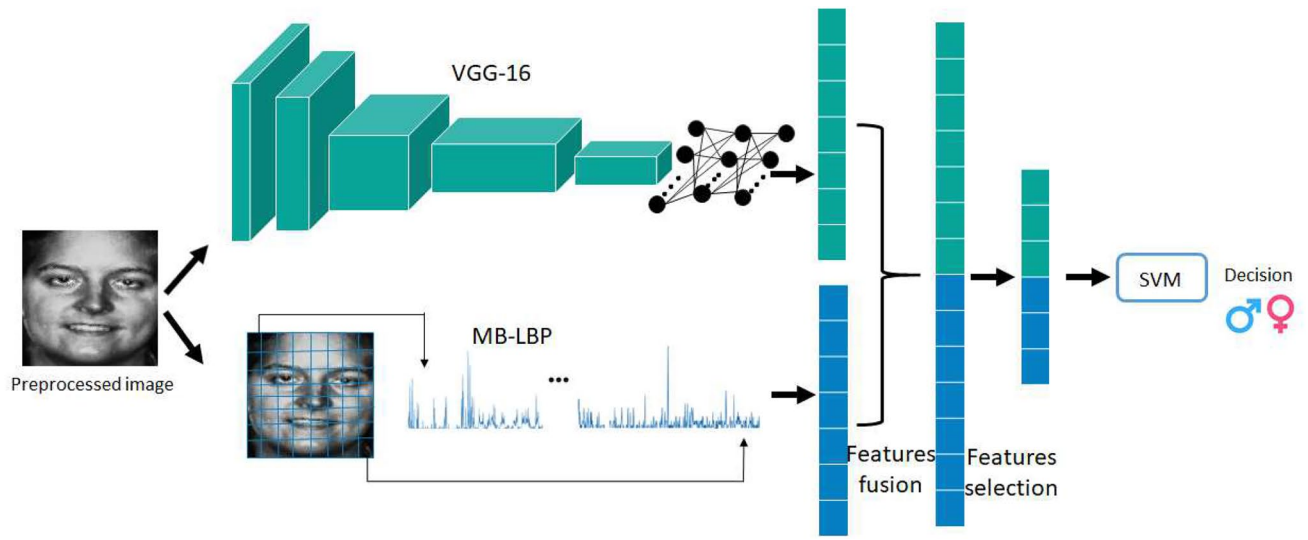

Fig. 8 Feature-level fusion proposed method

4.2 Experimental protocol

For fair comparison, we followed the settings used in the recent gender estimation works [3] [4].

As a validation measure, the standard Accuracy (Acc) metric [44] was used.

4.3 Experimental results

To evaluate the performance of the proposed method for facial gender estimation, three series of experiments was conducted. The first seeks to validate the use of the features selection method. As for the second series, it aimed to validate the feature-level fusion method. The third series of experiments compared the performance of the proposed method with the most recent gender estimation methods in the state of the art.

4.3.1 First series of experiments

To validate the contribution of selecting the discriminant features, a performance comparison of the proposed feature-level fusion method with and without features selection was conducted. This comparison deals with not only the classification accuracy rate, but also the size of the feature vector and the time execution. Table 1 shows this evaluation.

Based on the obtained results, two mains gains can be noted. The first is in terms of space memory: a gain of more than 10 times in the size of the feature vector on the Groups and more than 2.2 times on the FERET datasets. The second gain is in terms of speed: a gain in time execution more than

Fig. 9 Samples images from the FERET dataset

Fig. 10 Samples images from the Groups dataset

Table 1 Evaluation of features selection on the Groups and FERET datasets in terms of accuracy, space and time cost

Metrics	Datasets			
	Groups		FERET	
	Without selection	With selection	Without selection	With selection
Accuracy	96.39%	96.47%	99.14%	99.14%
Feature vector size	2048	200	2048	900
Time cost	1.60 ms	0.30ms	1.54 ms	0.75 ms

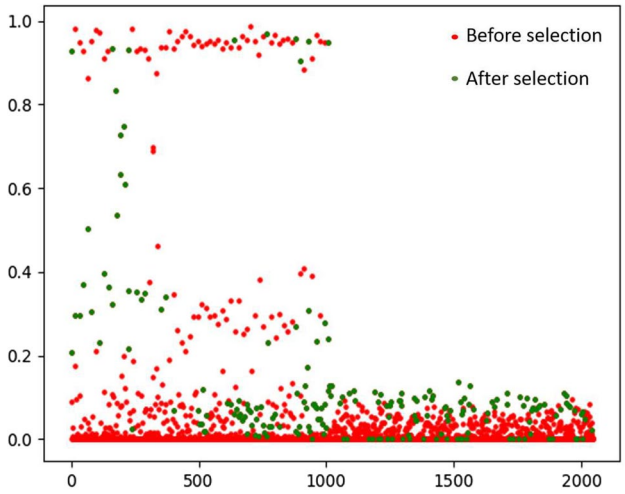

Fig. 11 Scatter plot visualizations of a feature vector before and after the features selection step

Table 2 Evaluation of the feature-level fusion strategy of the proposed CNN and handcraft models in terms of accuracy on the Groups and FERET datasets

Methods	Datasets	
	Groups	FERET
Proposed handcraft model (MB-LBP)	64.16%	79.9%
Proposed CNN model (Fine-tuned VGG16)	95.32%	99.04%
Proposed feature-level fusion method	96.47%	99.14%

5 times on the Groups dataset and more than 2 times on the FERET dataset, which is very important in the context of real-time applications.

Figure 11 illustrates a sample of features before and after the features selection in scatter plot. The combined features before the features selection step show redundancy (red dots). Therefore, the features selection technique was applied to select the most discriminant features. As shown in the scatter plot, the green dots represent the most relevant and discriminant features.

4.3.2 Second series of experiments

This series of experiments highlights the importance of the hybrid architecture for gender estimation. The fine-tuned VGG-16 and MB-LBP models were evaluated apart and then the performance of the feature-level fusion strategy was studied . Table 2 reports the accuracy rates for each of the generated models on the Groups and FERET datasets.

The obtained results reveal the remarkable gain while combining the deep learned and handcrafted features. Indeed, the obtained hybrid architecture generates more discriminant information. In fact, the proposed feature-level fusion method recorded 96.39% as accuracy rate, which is better than those achieved by the proposed CNN and handcraft models used apart on the Groups dataset. As for the FERET dataset, the proposed feature-level fusion method achieved 99.14% as accuracy rate, which outperforms those obtained by the proposed CNN and handcraft models. These observations confirm the effectiveness and the merit of the combination of information from deep learned and handcrafted features.

Figure 12 shows some facial image samples that are correctly classified with the proposed hybrid method but not correctly classified neither with the proposed CNN model nor with the proposed Mb-LBP-based model.

Fig. 12 Facial image samples correctly classified using the hybrid proposed method

Table 3 Comparison with the state of the art on the Groups and FERET dataset in terms of accuracy

Methods	Datasets	
	Groups	FERET
Handcrafted feature based methods		
[50]	86.11%	–
Deep learned feature based methods		
[15]	–	90.33%
[3]	95.00%	98.90%
Hybrid methods		
[4]	96.03%	98.84%
[10]	97.50%	99.71%
The proposed method	96.47%	99.14%

4.3.3 Third series of experiments

This series of comparisons considered different studies using different methods. In fact, [50] is based on the handcrafted features, whereas [3, 15] rely on the deep learned features; As for [4, 10] and the proposed method, they can be categorized as hybrid methods. The main purpose of the experiments was to compare the performance of the proposed method with the above-indicated studies in terms of accuracy on the Groups and FERET datasets. Table 3 displays the results of this comparative study.

As presented in Table 3, the proposed method has shown promising results thanks to the combination of the deep learned and handcrafted features. Indeed, compared to the handcrafted features method [50], a gain of more than 11% was recorded on the FERET dataset. In addition, compared to the deep learned features based methods, the achieved results outperform the methods of [3, 15] on the Groups and FERET datasets.

Referring to the hybrid methods, the proposed method achieved competitive results. In fact, a gain of 1.47% was reported on the Groups dataset and of 0.87% on the FERET dataset compared to [4] owing to the capability of the fine-tuned VGG-16 to extract the relevant deep features in addition to the local features in multi-resolution spatial histogram extracted by the MB-LBP descriptor. Although [10], based on the score-level fusion, slightly outperforms the proposed method, it is very expensive in computing time. Table 4 represents a computational cost comparison on the Groups and FERET datasets and highlights the important gain in terms of the classification process time. This gain is ensured, thanks to the features selection step which reduces the number of hybrid features. Such time cost gain is very important in the context of real-time applications.

Through this experimental study, the proposed method provides a good compromise in terms of speed and accuracy rate.

Table 4 Computational cost comparison on the Groups and FERET datasets

Datasets	Feature-level fusion method	Score-level fusion method [10]
Groups	0.37 ms	22.95 ms
FERET	0.75 ms	22.63 ms

5 Conclusion

In this paper, the facial gender estimation problem was addressed in an uncontrolled environment by introducing a new hybrid architecture that combines information from both deep learned and handcrafted features. For the deep learned features, a fine-tuned VGG-16 model to the gender estimation task was used to automatically extract relevant features. As for the handcrafted features, the MB-LBP descriptor was applied to encode local features.

The proposed method was evaluated on two widely used datasets (Image Of Groups and FERET) to demonstrate the generalization and discriminative power of the proposed model. This study has shown how the MB-LBP and CNN could complement each other in improving the accuracy results. Moreover, thanks to the features selection step, the proposed method provides a good compromise in terms of speed and accuracy rate.

This research provides new perspectives to evaluate the proposed method on additional databases also not only for face but also for other modalities. Moreover, a face recognition using facial gender information may be performed.

References

1. Afifi, M., Abdelhamed, A.: Afif4: deep gender classification based on adaboost-based fusion of isolated facial features and foggy faces. J. Vis. Commun. Image Represent. **62**, 77–86 (2019)
2. Amri, R., Gazdar, A., Barhoumi, W.: A comparative study on the importance of each face part in facial gender recognition via convolutional neural networks. In: 2021 IEEE/ACS 18th International Conference on Computer Systems and Applications (AICCSA), pp 1–86, IEEE (2021)
3. Aslam, A., Hussain, B., Cetin, A.E., Umar, A.I., Ansari, R.: Gender classification based on isolated facial features and foggy faces using jointly trained deep convolutional neural network. J. Electron. Imaging **27**(5), 053–023 (2018)
4. Aslam, A., Hayat, K., Umar, A.I., Zohuri, B., Zarkesh-Ha, P., Modissette, D., Khan, S.Z., Hussian, B.: Wavelet-based convolutional neural networks for gender classification. J. Electron. Imaging **28**(1), 013012 (2019)
5. Chen, W.S., Jeng, R.H.: A new patch-based lbp with adaptive weights for gender classification of human face. J. Chin. Inst. Eng. 1–7 (2020)
6. Chitprasert, B., Rao, K.: Discrete cosine transform filtering. Signal Process. **19**(3), 233–245 (1990)

7. Cristianini, N., Shawe-Taylor, J.: An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge (2000)
8. Dagher, I., Azar, F.: Improving the svm gender classification accuracy using clustering and incremental learning. Expert. Syst. **36**(3), e12372 (2019)
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol 1, pp 886–893. IEEE (2005)
10. Dammak, S., Mliki, H., Fendri, E.: Gender effect on age classification in an unconstrained environment. Multimed. Tools Appl. **80**(18), 28001–28014 (2021)
11. Devries, T., Biswaranjan, K., Taylor, G.W.: Multi-task learning of facial landmarks and expression. In: 2014 Canadian Conference on Computer and Robot Vision, IEEE, pp 98–103 (2014)
12. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. **3**(02), 185–205 (2005)
13. Dokmanic, I., Parhizkar, R., Ranieri, J., Vetterli, M.: Euclidean distance matrices: essential theory, algorithms, and applications. IEEE Signal Process. Mag. **32**(6), 12–30 (2015)
14. Duan, M., Li, K., Yang, C., Li, K.: A hybrid deep learning cnn-elm for age and gender classification. Neurocomputing **275**, 448–461 (2018)
15. Dwivedi, N., Singh, D.K.: Review of deep learning techniques for gender classification in images. In: Harmony Search and Nature Inspired Optimization Algorithms, pp. 1089–1099. Springer, New York (2019)
16. Fekri-Ershad, S.: Developing a gender classification approach in human face images using modified local binary patterns and tani-moto based nearest neighbor algorithm (2020). arXiv preprint arXiv:2001.10966
17. Freund, Y., Schapire, R.E.: A desicion-theoretic generalization of on-line learning and an application to boosting. In: European Conference on Computational Learning Theory, pp. 23–37. Springer, New York (1995)
18. Gallagher, A.C., Chen, T.: Understanding images of groups of people. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp 256–263 (2009)
19. Georgescu, M.I., Ionescu, R.T., Popescu, M.: Local learning with deep and handcrafted features for facial expression recognition. IEEE Access **7**, 64827–64836 (2019)
20. Golrizkhatami, Z., Acan, A.: Ecg classification using three-level fusion of different feature descriptors. Expert Syst. Appl. **114**, 54–64 (2018)
21. Hecht-Nielsen, R.: Kolmogorov's mapping neural network existence theorem. In: Proceedings of the International Conference on Neural Networks, Vol. 3, pp 11–14. IEEE Press, New York (1987)
22. Hosseini, S., Lee, S.H., Kwon, H.J., Koo, H.I., Cho, N.I.: Age and gender classification using wide convolutional neural network and gabor filter. In: 2018 International Workshop on Advanced Image Technology (IWAIT), IEEE, pp 1–3 (2018)
23. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. Neurocomputing **70**(1–3), 489–501 (2006)
24. Irhebhude, M.E., Kolawole, A.O., Goma, H.K.: A gender recognition system using facial images with high dimensional data. Malays. J. Appl. Sci. **6**(1), 27–45 (2021)
25. Keller, J.M., Gray, M.R., Givens, J.A.: A fuzzy k-nearest neighbor algorithm. IEEE Trans. Syst. Man Cybern. **4**, 580–585 (1985)
26. Khan, A., Chefranov, A., Demirel, H.: Image scene geometry recognition using low-level features fusion at multi-layer deep cnn. Neurocomputing **440**, 111–126 (2021)
27. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp 1097–1105 (2012)
28. Li, X., Ma, X., Song, P.: Fusion of deep feature and hand-crafted features for terrain recognition. In: IOP Conference Series: Materials Science and Engineering, IOP Publishing, vol 646, p 012052 (2019)
29. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision (2004)
30. Mliki, H., Dammak, S., Fendri, E.: An improved multi-scale face detection using convolutional neural network. In: Sps, S. (ed.) Signal, Image and Video Processing. Springer, New York (2020). https://doi.org/10.1007/s11760-020-01680-w
31. Mohamed, S., Nour, N., Viriri, S.: Gender identification from facial images using global features. In: 2018 Conference on Information Communications Technology and Society (ICTAS), IEEE, pp 1–6 (2018)
32. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recogn. **29**(1), 51–59 (1996)
33. Ojansivu, V., Heikkilä, J.: Blur insensitive texture classification using local phase quantization. In: International conference on image and signal processing, pp 236–243. Springer (2008)
34. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition (2015)
35. Petkov, N.: Biologically motivated computationally intensive approaches to image pattern recognition. Futur. Gener. Comput. Syst. **11**(4–5), 451–465 (1995)
36. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The feret database and evaluation procedure for face-recognition algorithms. Image Vis. Comput. **16**(5), 295–306 (1998)
37. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015)
38. Rwigema, J., Mfitumukiza, J., Tae-Yong, K.: A hybrid approach of neural networks for age and gender classification through decision fusion. Biomed. Signal Process. Control **66**, 102459 (2021)
39. Ryu, H.J., Adam, H., Mitchell, M.: Inclusivefacenet: Improving face attribute detection with race and gender diversity. In: arXiv preprint (2017) arXiv:1712.00193
40. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 815–823 (2015)
41. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp 618–626 (2017)
42. Serna, I., Pena, A., Morales, A., Fierrez, J.: Insidebias: measuring bias in deep networks and application to face gender biometrics. In: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, pp 3720–3727 (2021)
43. Sharma, N., Sharma, R., Jindal, N.: Face-based age and gender estimation using improved convolutional neural network approach. In: Wireless Personal Communications, pp 1–20 (2022)
44. Sheetlani, J., Dhawale, C., Pardeshi, R.: Gender identification from frontal facial images using multiresolution statistical descriptors. In: Computing, Communication and Signal Processing, pp. 977–986. Springer, New York (2019)
45. Sidney Burrus, C., Gopinath, R.A., Guo, H.: Introduction to wavelets and wavelet transforms. In: A Primer. Prentice Hall, New York (1998)
46. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: arXiv preprint (2014) arXiv:1409.1556

47. Soofi, E.S.: Principal information theoretic approaches. J. Am. Stat. Assoc. **95**(452), 1349–1353 (2000)
48. Tianyu, L., Fei, L., Rui, W.: Human face gender identification system based on mb-lbp. In: 2018 Chinese Control And Decision Conference (CCDC), IEEE, pp 1721–1725 (2018)
49. Xizhao, W., Yanxia, Z., Farhad, P.: Recent advances in deep learning. Int. J. Mach. Learn. Cybern. **11**, 747–750 (2020)
50. Zhang, C., Ding, H., Shang, Y., Shao, Z., Fu, X.: Gender classification based on multiscale facial fusion feature. Mathematical Problems in Engineering 2018 (2018)
51. Zhang, L., Chu, R., Xiang, S., Liao, S., Li, S.Z.: Face detection based on multi-block lbp representation. In: International Conference on Biometrics, pp 11–18. Springer (2007)
52. Zheng, Y., Zhu, C., Luu, K., Bhagavatula, C., Le, T.H.N., Savvides, M.: Towards a deep learning framework for unconstrained face detection. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–8, IEEE (2016)