APPLICATION OF SOFT COMPUTING

An Intelligent handcrafted feature selection using Archimedes optimization algorithm for facial analysis

Imène Neggaz[1] · Hadria Fizazi[1]

Abstract

Human facial analysis (HFA) has recently become an attractive topic for computer vision research due to technological progress and mobile applications. HFA explores several issues as gender recognition (GR), facial expression, age, and race recognition for automatically understanding social life. This study explores HFA from the angle of recognizing a person's gender from their face. Several hard challenges are provoked, such as illumination, occlusion, facial emotions, quality, and angle of capture by cameras, making gender recognition more difficult for machines. The Archimedes optimization algorithm (AOA) was recently designed as a metaheuristic-based population optimization method, inspired by the Archimedes theory's physical notion. Compared to other swarm algorithms in the realm of optimization, this method promotes a good balance between exploration and exploitation. The convergence area is increased By incorporating extra data into the solution, such as volume and density. Because of the preceding benefits of AOA and the fact that it has not been used to choose the best area of the face, we propose utilizing a wrapper feature selection technique, which is a real motivation in the field of computer vision and machine learning. The paper's primary purpose is to automatically determine the optimal face area using AOA to recognize the gender of a human person categorized by two classes (Men and women). In this paper, the facial image is divided into several subregions (blocks), where each area provides a vector of characteristics using one method from handcrafted techniques as the local binary pattern (LBP), histogram-oriented gradient (HOG), or gray-level co-occurrence matrix (GLCM). Two experiments assess the proposed method (AOA): The first employs two benchmarking datasets: the Georgia Tech Face dataset (GT) and the Brazilian FEI dataset. The second experiment represents a more challenging large dataset that uses Gallagher's uncontrolled dataset. The experimental results show the good performance of AOA compared to other recent and competitive optimizers for all datasets. In terms of accuracy, the AOA-based LBP outperforms the state-of-the-art deep convolutional neural network (CNN) with 96.08% for the Gallagher's dataset.

Keywords Archimedes optimization algorithm (AOA) · Human facial analysis (HFA) · Wrapper feature selection (FS) · Handcrafted methods · Automatic selection

1 Introduction

Human vision allows performing several tasks in parallel and rapidly, particularly facial detection, gender recognition, and recognizing the state of mind, which differentiates the human being from others (Khan et al. 2020).

The automation of gender recognition represents a real challenge for scientific researchers , and it has a significant impact on the commercial field and video surveillance (Greco et al. 2021). For example, shopping centers are interested in knowing the sales rate and the category of people who buy their products, particularly the gender, age, and origins, to increase the sales rate. Also, another area requires the application of gender recognition to detect suspected people captured by surveillance cameras in large spaces such as airports, shopping malls, and gas stations. To reduce the time of searching for the target suspected person, the gender recognition application can contribute profoundly to solving this

✉ Imène Neggaz
imene.neggaz@univ-usto.dz

Hadria Fizazi
hadriafizazi@yahoo.fr

[1] Laboratoire Signal Image PArole (SIMPA), Département d'informatique, Faculté des Mathématiques et Informatique, , Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, USTO-MB, EL M'naouer, BP 1505, 31000 Oran, Algérie

issue, especially for critical situations as suicide bombing or airport attack. Also, the current situation of the Covid-19 pandemic obliged people to wear the mask that the automation of gender recognition plays a vital role in our life (Fitousi et al. 2021).

Automatic gender recognition has grown critical in recent years, particularly in crime detection. Numerous modalities have been employed to identify female and male subjects, including the face (Hsu et al. 2021), voice (Livieris et al. 2019), gait (Lee et al. 2022), hand (Afifi 2019), fingerprint (Jalali et al. 2021), and electrocardiogram (ECG) signals (Uçar et al. 2021).

For this study, we selected visual information that provides direct and more intuitive information instead of more abstract signals such as biosignals (Comas et al. 2020). According to the literature, the human detection rate of GR is lower than 95%, which increases the difficulties of automatic gender recognition (Boon Ng 2012). In this field, several challenges are considered as rotated, occluded faces, and the human who is similar to women with long hair, which strongly affects the performance of gender recognition.

In general, the task of gender recognition requires three crucial steps (Greco et al. 2020). The first step detects the facial area, and the second aims to extract the features from faces. In contrast, the last step is reserved for realizing the task of classification into binary classes using machine learning approaches such as support vector machines (SVM), extreme learning machines (ELM), and multilayer perceptron (MLP). As a case study, we find the work of Goel and Vishwakarma. (Goel and Vishwakarma 2016) which applied kernel principal component analysis (KPCA) to reduce the size of intensity vectors and provide 64 principal components for gender classification. This vector is considered as an input of SVM that used 2-fold cross-validation on three datasets. The quantitative study has shown a performance of accuracy 97.375%, 99.7%, and 96.67% for the three datasets: GT, AT@T, and Faces94 datasets. Several descriptors based on handcrafted are introduced for face recognition, gender, and age identification during the last decades. This type includes mainly four categories: texture features, facial shape features, intensity pixels, and geometric features. The first type includes predominately local binary pattern (LBP) (Sajja and Kalluri 2019), local phase quantization (LPQ) (Nguyen and Huong 2017), and local ternary patterns (LTP) (Osman and Viriri 2020), while the second type employed basically histogram-oriented gradient (HOG) (Khalifa and Şengül 2018), pyramid HOG (PHOG), and multi-level HOG (ML-HOG) (Bekhouche et al. 2017). The third type used mostly gray-level co-occurrence matrix (GLCM) and rotated GLCM. Finally, the features are extracted by scale-invariant features transform (SIFT). It is important to highlight that many algorithms are designed to enhance the basic handcrafted methods by fusing textural

information with facial shape features for gender recognition (Micheal and Geetha 2019). This work consists of extracting two textural descriptors (dominated RLBP and rotation-invariant LPQ), combined with pyramid HOG, to determine the gender of persons automatically. In addition, the authors employed an SVM classifier based on three kernel functions (linear, polynomial, and RBF). The experimental study has been validated by three datasets: FEI, LFW, and Adience. The obtained results proved that the SVM-based RBF kernel achieved higher performance in terms of accuracy for FEI, LFW, and Adience with 95.3, 98.7, and 96%, respectively. In the same context, Geetha et al.(Geetha et al. 2019) designed a new gender descriptor which is based on edge feature, texture feature, and intensity characteristics. The first part is encoded by 8-local directional pattern (8-LDP), whereas LBP implements the second part and, the final part represents the pixels values of the image. The authors used FEI and a self-designed dataset to treat gender classification tasks. The obtained results attained 99% as accuracy for FEI and 94% for the self-designed dataset using an SVM classifier based on multi-blocks combined descriptor.

Nowadays, deep-learned features are used exponentially in machine learning, especially in computer vision for facial analysis (Singh et al. 2021), including gender, age, emotion and alignment (Aspandi et al. 2021), affect estimation (Aspandi et al. 2020), biosignal analysis (Comas et al. 2020), biomedical application (Peimankar and Puthusserypady 2021; Yu et al. 2021), and remote sensing (Peker 2021; Alhichri et al. 2021). Recently, several architecture are created by pretrained CNN such as VGG16, ResNet, GoogleNet, CaffeNet, AlexNet, and Inception for gender recognition (Huynh and Nguyen 2020; Savchenko 2019; Lapuschkin et al. 2017; Silva 2019; Abirami et al. 2020; Lin et al. 2020[?]). Furthermore, a comparative study between several pretrained CNN such as MobileNet, DensNet, Xception, and SqueezeNet is realized by (Greco et al. 2020) for gender recognition. Hence, a great competition between handcrafted features and deep-learned (DL) features have been highlighted. For instance, we find the work of (Lapuschkin et al. 2017), which used three pretrained CNN called CaffeNet, VGG16, and GoogleNet for estimating age and gender information.

The great number of attributes generated by both methods (Handcrafted and deep features) prompts researchers to develop new methods called wrapper feature selection based on metaheuristics (MHs). MHs are derived from different subjects, allowing the development of many optimization algorithms that can be merged with machine learning techniques.

Despite the significant development of metaheuristics generated by the imitation of animal behavior, mathematical and physical laws in the field of feature selection as manta-ray foraging optimizer (MRFO) (Ghosh et al. 2021),

emperor penguin optimizer (EPO) (Dhiman et al. 2021), Harris hawks optimizer (HHO) (Thaher et al. 2020), gray wolf optimizer (GWO) (Al-Tashi et al. 2020), whale optimization algorithm (WOA) (Mafarja and Mirjalili 2018), Henry gas solubility optimization (HGSO) (Neggaz et al. 2020), multiverse optimizer (MVO), equilibrium optimizer (EO) (Gao et al. 2020), and sine cosine algorithm (SCA) (Taghian and Nadimi-Shahraki 2019), the integration of these methods in complex problems such as gender recognition remains limited.

The theory of "No-free lunch" (Wolpert and Macready 1997) allows us to introduce a novel issue in the field of optimization. Furthermore, AOA is a physical-based population metaheuristic published in 2020. Also, AOA is a particular algorithm due to its encoding solutions. It includes the search space representation in D dimension with complementary information such as volume and density, which improves the convergence speed compared to other metaheuristics. This algorithm is more suitable for applying to real applications, and it contains fewer control parameters than other physical algorithms such as HGSO (Hashim et al. 2019), MVO (Mirjalili et al. 2016), and EPO (Dhiman and Kumar 2018). Despite its recent creation, this algorithm has grown significantly by affecting several application areas, such as optimization and engineering problems (Hashim et al. 2020), heart disease diagnosis (Anand 2021), industry design (Yıldız et al. 2021), WSN performance prediction (Preeti 2021), wind energy generation systems (Fathy et al. 2021), plant disease identification (Annrose et al. 2021), and PEM fuel estimation (Sun et al. 2021; Yao and Hayati 2021). Due to the prior merits of AOA and the fact that it has not been utilized to select the optimal area of interest on the face, this issue can be handled by utilizing a wrapper feature selection technique, which is a significant motivation in the field of computer vision and machine learning. Thus, gender recognition from face images is still a big challenge to this day, i.e., how to determine the most significant areas from face images characterized by local binary pattern (LBP), histogram of oriented gradient (HOG) or gray-level co-occurrence matrix (GLCM) descriptors intelligently and automatically?

This paper automatically determines the significant areas based on handcrafted features (LBP, HOG, or GLCM) from the face using the Archimedes optimization algorithm (AOA) to solve gender recognition problems using an optimal number of faces extracted areas.

The major contributions of this paper are as follows:

- Designing a novel wrapper physical algorithm AOA for predicting gender identification using an automatic selection of the optimal and significant areas of face images.
- Comparing the performance of AOA with several recent and robust optimizers as for facial analysis based on feature selection (FS).

- Evaluating the impact of three handcrafted features based on LBP, HOG, and GLCM.
- Testing the efficiency of AOA for gender recognition over smallest datasets (FEI (Thomaz and Giraldi 2010) and GT (Nefian 2013)) and largest dataset (Gallagher's database (Gallagher and Chen 2009)).
- Comparing the efficiency of handcrafted methods based on AOA to deep convolutional neural network (CNN) for largest dataset (Gallagher's database)

The following structure of our paper contains six sections. Section 2 explains some works which treat gender recognition based on handcrafted features, deep features, hybrid descriptors, and automatic feature extraction methods. In Sect. 3, three handcrafted descriptors are explained, including local binary pattern (LBP), histogram-oriented gradient (HOG), and gray-level co-occurrence matrix (GLCM). Section 4 gives the concept of the Archimedes optimizer algorithm in detail. After, we propose our architecture of AOA wrapper feature selection for gender recognition by defining the structure of the encoding solution of an immersed object, the score function, and the designed framework. The section represents the kernel of our paper which includes datasets description, parameters of algorithms, quantitative, and graphical results proved by statistical analysis using rank-sum Wilcoxon's (Neggaz et al. 2020). Finally, Sect. 7 shows our conclusion with some future horizons.

2 Related work

This part summarizes the potential work of literature related to facial analysis.

Firstly, we give a recap of handcrafted features for gender recognition. Secondly, a short description of deep-learned features is shown for human facial analysis. Thirdly, hybrid features are described. Finally, an overview of wrapper feature selection-based gender recognition is presented.

2.1 Handcrafted features

Dago-Casas et al. (2011) proposed a combination of Gabor jets, pixels, and LBP, tested on the largest datasets captured in a real-world environment, including LFW and Gallagher's datasets. The authors used two classifiers (SVM and LDA) to recognize gender as male and female. We noticed that the dimensionality is reduced using PCA. The obtained results show that Gabor+PCA+SVM is better than Gabor+PCA+LDA in terms of accuracy. The numerical results indicate that the accuracies are 86.61% for the Gallagher's dataset and 94.01% for the LFW dataset.

Castrillón-Santana et al. (2013) proposed a combination of global and local features for recognizing gender in

the FERET dataset. LBP and 2D-DCT were used as local features, while the authors used PCA as a global feature. The classification task is realized by k-NN, which reached 98.16% in terms of accuracy.

Surinta and Khamket. (2013) used a combination of LBP and HOG features taken from multiple resolutions of the head and shoulders pattern. The experimental study is conducted following Dago's procedure on Gallagher's dataset (distance inter-occular greater than 20 pixels with 5-fold). Two classifiers are used to accomplish the gender identification task: SVM and Bagging. In terms of accuracy, Bagging+LBP+HOG exceeds SVM+LBP+HOG, with the best model reaching 88.1%.

Ng et al. (2015) summarized a review of gender recognition in facial images They discussed several methods for extracting features, including LBP, SIFT, Gabor Wavelet, and external cues. Additionally, they mentioned numerous datasets that pose difficulties for GR, including AR, FERET, BioID, CMU-PIE, FRGC, MORPH, LFW, and Gallagher's. Due to the increased quantity of facial photographs, 28231 were collected in uncontrolled situations, and the last dataset poses a significant challenge in the field of gender identification.

Castrillón-Santana et al. (2016) extracted the periocular area firstly to determine the set of features using handcrafted methods including HOG, LBP, local ternary patterns (LTP), Weber local descriptor (WLD), and local oriented statistics information booster (LOSIB). Their idea consists of creating a grid with 6×7 cells. Then, they applied handcrafted methods to each cell. The experimental study is validated by Gallagher's dataset under Dago's protocol (Dago-Casas et al. 2011). The best descriptor is a handcrafted combination of HOG, uniform LBP (LBP^{u2}), LTP, and WLD, using SVM as the classifier. This combination achieved an accuracy of 82.91%.

Recently, several methods have been developed in the literature called handcrafted techniques. The extracted features are determined from the whole face or some regions by computing the local gradient parameters as a histogram of oriented gradients (HOGs) and scale-invariant feature transform (SIFT) (Surinta et al. 2019). The authors used a support vector machine as a classifier for identifying the two classes: female and male from Color FERET datasets. The experimental study has shown that HOG outperformed the SIFT descriptor-based SVM when the size of training data is reduced.

Zhang et al. (2018) developed a novel fusion of facial features for gender recognition. The vector of characteristics is obtained by combining local binary pattern (LBP), local phase quantization (LPQ), and a multi-block. The classification task is realized using a support vector machine (SVM) and tested on Gallagher's dataset. The experimental results

show that the proposed method outperformed another basic versions (LBP and LPQ).

Ghojogh et al. (2018) designed four frameworks for gender identification using facial images. The first framework consists of extracting features using the texture method based on LBP and reducing the dimension of vector features using PCA, which will be served as input for multilayer perceptron (MLP).The second framework used Gabor filters to provide the vector of features reduced by PCA and served as input for the kernel SVM classifier. The third framework extracts the lower part of face with the size of sub-image (30×30), which will be reshaped to column vector with the size of (900×1) and served as input for kernel SVM classifier. The last framework extracts 34 landmarks from the face, classified by linear discriminant algorithm (LDA). All proposed frameworks are assessed on FEI datasets, and the experimental results shows that the third framework outperforms others by 90% in terms of accuracy. However, the accuracy increases to 94% when the weighting vote takes the decision. Also, gender identification is solved by texture and geometric features, which can be determined by local binary pattern (LBP) and gray-level co-occurrence matrices (GLCM) (Omer et al. 2019). More recently, several enhanced versions of LBP have been developed for face and gender recognition as local directional pattern (LDP) and local phase quantization (LPQ). In the same context, a novel variant of LBP was proposed by Chen and Jeng. (2020) named adaptive patch-weight LBP (APWLBP). Their method used a pyramid structure to compute the gradient using weight parameters determined by Eigen theory. The main objective of (APWLBP) is to determine the optimal projection on the hyperplane with a high value of variance for gender recognition. The performance of APWLBP-based SVM is very competitive against CNN on three-dimensional Adience [1] and LFW datasets (Gary et al. 2007). In the same context, scale-invariant feature transform (SIFT) is combined with trainable features (CROSSFIRE) (Pai and Shettigar 2021).

2.2 Deep-learned features

In recent years, deep learning CNNs have dominated many computer vision applications.

Mansanet et al. (2016) developed a local-deep textcolor-brownneural network (LDNN) model for gender recognition; this model was constructed using a feed-forward NN without dropout to extract features from the input photographs. After detecting the edges of the facial picture, small image patches were selected around these edges. All the image patches were used to train the neural networks. The final output was calculated by averaging the expectations for all the patches in the

---

[1] https://talhassner.github.io/home/projects/Adience/Adience-data.html.

input test picture. The LDNN model achieves an accuracy of 96.25% and 90.58% on the largest datasets of LFW and Gallagher's datasets, respectively.

Orozco et al. (2017) proposed a deep CNN architecture to recognize the gender of the two largest datasets, named LFW (13233 pictures) and Gallegher's dataset (28231 images), captured in real-world conditions. The authors applied the first face detection step using Haar features embedded in the AdaBoost detector. Then, the classification task is realized by Ubunsa CNN. [2] The obtained results show that the accuracies of LFW and Gallagher's datasets are 95.42 and 91.48%.

A novel synergy between CNN and ELM is illustrated by the work of Duan et al. for age and gender identification (Duan et al. 2018). The CNN is used as an extractor of features. In contrast, ELM is used as a classifier to simultaneously determine the person's age and distinguish between male and female from face images. Acien et al. (2018) employed two architectures of pretrained CNN named VGG16 and ResNet for measuring ethnicity and gender information. Ito et al. (2018) designed four architecture of pretrained CNN: AlexNet, VGG-16, ResNet-152, and Wide-ResNet-16-8 for predicting the age and gender over IMDB-WIKI datasets. The experiment study showed that Wide-ResNet presented a textcolorbrownhigh accuracy compared to others pretrained CNN. Mane and Shah (2019) explored human facial analysis using three points: face recognition, gender, and expression recognition. This study focuses only on gender recognition; hence, the authors used CNN for gender recognition. Agrawal and Dixit (2019) used CNN as extractor for predicting age and gender from facial images. Then, the dimensionality reduction is applied using principal component analysis (PCA), and the task of classification is implemented using feed-forward neural network.

Haider et al. (2019) proposed a specific deep CNN architecture for real-time gender identification using smartphones. The architecture comprises 4 convolutional layers, 3 max-pooling layers, 2 fully connected layers, and a single layer for regression. The training is realized by fusing two datasets, FEI and CAS-PEAL, including 200 persons with 2800 faces and 1040 persons with 30,871. For the FEI dataset, the deep-gender registered 98.75% as accuracy by considering a specific process, including alignment, before reducing the size of the facial image. However, the proposed method reached 97.73% in terms of accuracy for CAS-PEAL-R1 datasets. It is important to indicate that the authors split their dataset to 5-fold as cross-validation.

Abdalrady and Aly (2020) used a straightforward feature fusion technique based on two simple CNN architectures. The authors created a simple CNN called PCANet that replaces the complex convolutional layer filters with a bank of PCA filters in each stage. Based on Gallagher's dataset, the

suggested model "2-stage PCANet" performance was analyzed and found to be 89.65%.

2.3 Deep-learned assisted by handcrafted features

A significant number of studies have been investigated deep-learned features and their impact compared to handcrafted features and fused features (handcrafted with deep features) for gender identification (Simanjuntak and Azzopardi 2020).

Dwivedi and Singh (2019) summarized firstly, some methods based on handcrafted as LBP, HOG, SIFT, weighted HOG, and CROSSFIRE Filter, and secondly, the authors explained the role of CNN, which can be used for double tasks, i.e., used as an extractor of features and classifier to recognize age and gender from the face.

Althnian et al. (2021) have employed three methods, including LBP, HOG, and PCA as handcrafted features, deep CNN features, and fused features based on three combinations named LBP-DL, HOG-DL, and PCA-DL. Furthermore, gender identification task is realized by two classifiers, SVM and CNN. The experimental results showed a high average accuracy of 88.1% obtained by fused features (LBP-DL) and SVM classifier, tested on two datasets, LFW and Adience. Additionally, Hsu et al. (2021) designed three occlusion methods assisted by AdienceNet and VGG16 for recognizing age and gender tasks.

2.4 Automatic feature extraction methods using evolutionary algorithms

A few works have been published in this area. For example, Zhou and Li (2019) used a genetic algorithm (GA) to classify gender-based faces automatically. Their idea involves applying GA in order to determine the optimal set of Eigen-features extracted from faces by PCA and classified by neural network. The experimental study is validated using two datasets, including FEI and FERET. The obtained results achieved 96% and 94% as accuracy rates, respectively.

Ghazouani (2021) applied genetic programming (GP) for HFA based on feature selection, specifically facial expression recognition. The author applied several steps as face detection using HOG as descriptor and SVM as binary classifier. Then, the step of features extraction is realized using the texture method (LBP) and geometric methods (linear and eccentricity features). Finally, the author applied GP to select the relevant features of the fused set to recognize the emotions.

3 Features extraction

Several methods have been proposed to describe the texture characteristic In this section. We present below a brief

---

[2] https://github.com/ciorozco/CNN_Ubunsa.

study of some existing techniques for extracting of texture features, which are applied to the analysis of facial images. The purpose of extracting descriptors (characteristic) in pattern recognition is to express primitives in a numerical or symbolic form called encoding. This part will introduce the descriptors used in the experiments and results part. These are first of all the local binary patterns (LBP), then the descriptors of histogram based on oriented gradient (HOG), and finally the gray-level co-occurrence matrix (GLCM).

3.1 Local binary patterns (LBP)

Texture descriptors based on local binary patterns were initially proposed by Ojala et al. (2002). The computation of the resulting image from the LBP application is akin to a correlation operation while applying a filter to a digital image. It suffices to process each image pixel by considering the eight pixels of its immediate neighborhood. The neighborhood of a pixel forms a matrix of 3×3 pixels where the pixel to be processed is in the center, and its neighborhood is around. Figure 1 shows an example of the execution of the LBP algorithm relating to the steps described below.

Step 1 Extraction of the neighborhood of the pixel to be processed. The eight intensity values of pixel's neighborhood to be processed are extracted from a matrix of 3×3 pixels. In this example, each pixel has a different gray intensity value. The pixel being processed has the value of intensity 40.

Step 2 A thresholding is performed on the intensity value of the neighboring pixels. Any pixel with an intensity value greater than or equal to the intensity value of the processed pixel is assigned the value 1. The value 0 is assigned to any intensity value lower than that of the processed pixel.

Step 3 A multiplier matrix is stored. This matrix will describe the resulting local binary form uniquely in the next step of the algorithm.

Step 4 Element-by-element multiplication. This operation is carried out between the matrix resulting from the thresholding of step 2 and the multiplying matrix of step 3.

Step 5 The summation of the values of the resulting matrix from step 4 is performed. This sum is related in the output image to the corresponding coordinates of the pixel to be processed in the input image. The algorithm re-executes steps 1 to 5 until all the input image pixels are processed. According to the procedure for identifying LBP, a histogram is calculated to characterize the frequency of appearance of various patterns. The computed number for each pixel in step 5 uniquely identifies a gray intensity pattern among the possible patterns. The shape of the resulting histogram is characteristic of the texture studied by the LBP algorithm.

In general, extracting features from facial images using LBP starts by dividing the input image into several blocks (7×7). Then, we extract the histogram for each block based on LBP. The final step consists of concatenating all histograms to realize gender recognition. To calculate the LBP code in a neighborhood of P pixels with a radius R, we simply count the occurrences of gray levels g_p greater than or equal to the central value using Eq. (1).

$$LBP_{P,R}(x_c, y_c) = \sum_{p=1}^{P} s(g_p - g_c)2^{p-1} \tag{1}$$

where g_p and g_c are the gray levels of a neighboring pixel and the central pixel, respectively. S indicates the Heaviside function defined by Eq. 2:

$$S(x) = \begin{cases} +1 \ if \ x \geq 0 \\ 0 \ if \ x < 0 \end{cases} \tag{2}$$

According to the work of Ghazouani (2021), $LBP_{8,1}$ provides a total of 58 uniform pattern and a total of 59 bins. As depicted in Fig. 2, the cropped face is divided into (7×7) blocks. The 59 bins have been computed for each block and concatenated to construct a global histogram for the entire face. Therefore, each face is represented by a histogram of (59×49) bins

3.2 Histogram of oriented gradient (HOG)

HOG is a powerful descriptor proposed by Dalal and Triggs in 2005, initially developed for human detection Dalal et al. (2005). However, later it is extended and applied to other computer vision problems, including facial recognition Hung (2021), gender and age estimation Patil (2021), detection of plant pathology's Pattnaik and Parvathi (2021), and facial expressions recognition (Lakshmi and Ponnusamy 2021). HOG describes the appearance and local shape of the object in an image using the distribution of gradients. The characteristic vector of an image $I(x, y)$ using the HOG technique is obtained by the following procedure:

Step 1—Divide the image $I(x, y)$ into equal blocks $(N_b \times N_b)$, where each block contains $(M \times M)$ regular cells of size (8×8) pixels.

Gradient values (G_h, G_y) are computed for each pixel using a centered $1 - D$ derivative filter in the horizontal and vertical directions. For this, the following masks (S_h, S_v) are used and defined by Eq.(3) and Eq. (4):
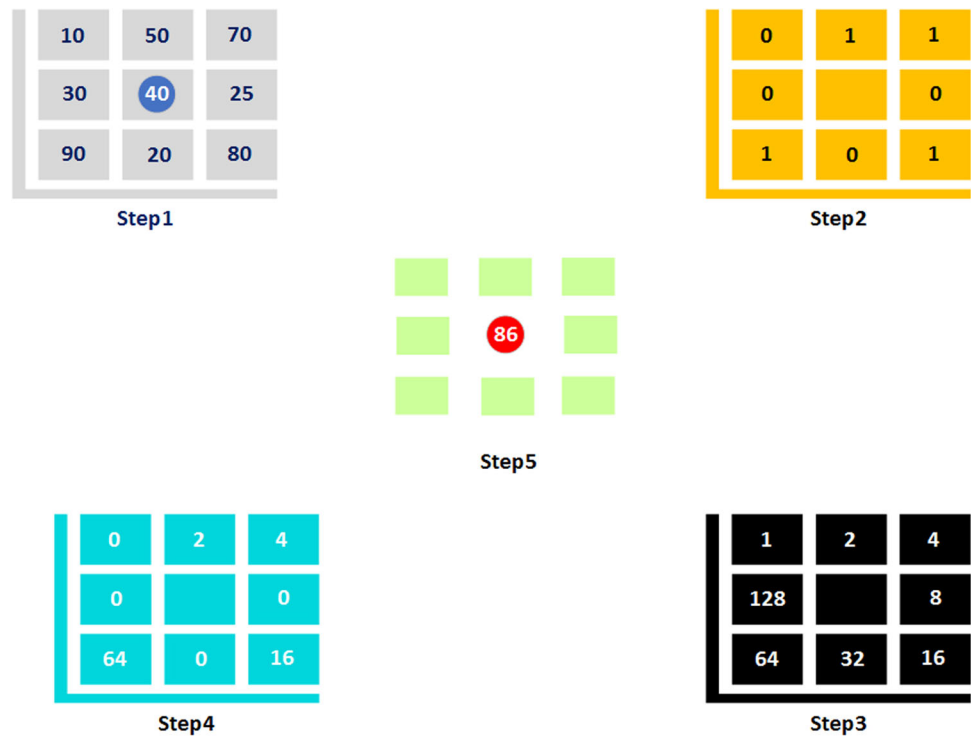
$$S_h = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \tag{3}$$

$$S_v = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{4}$$

$$G_h(x, y) = I(x, y) * S_h \tag{5}$$

$$G_v(x, y) = I(x, y) * S_v \tag{6}$$

Fig. 1 Basic LBP operator

Step 2—The magnitude ($|G(x, y)|$) and gradient orientation (θ) of each pixel (x, y) are calculated using Eqs. (7) and (8):

$$|G(x, y)| = \sqrt{G_h^2(x, y) + G_v^2(x, y)} \tag{7}$$

$$\theta = \arctan\left(\frac{G_v(x, y)}{G_h(x, y)}\right) \tag{8}$$

G_h and G_v represent the horizontal gradient and the vertical gradient at pixel (x, y), respectively.

Step 3—The histogram of the orientation based on the gradient inside each cell is calculated by quantizing unsigned gradients at each pixel in 9 channels (bins) orientations. The histograms are uniform from 0 to 180 (unsigned case) or from 0 to 360 (signed case).

Step 4—The characteristic vector for each cell is normalized using histograms in their recognized blocks. In this work, we use the L2-norm for the normalization of the blocks; the normalization factor is calculated using the following equation:

$$Hist_n = \frac{Hist}{\sqrt{\|Hist\|_2^2 + \epsilon}} \tag{9}$$

where $Hist$ is the non-normalized vector containing all the histograms in a block, $\|Hist\|_2$ is the L2 norm of the descriptor vector, and ϵ is a regularization term.

Step 5—The characteristic vector of each block is formed by concatenating the histogram vectors of all the cells in the block. The characteristic vector HOG is formed by concatenating the characteristic vectors of all the blocks for a given image.

3.3 Gray-level co-occurrence matrix (GLCM))

GLCM is a method widely used in image processing that belongs to the class of texture-based statistical methods. Textural content is expressed differently depending on distance d and orientation θ of the displacement considered between the pairs of sites, which provides 4 GLCMs due to the 4 orientations defined by : $\theta_1 = 0^o, \theta_2 = 45^o, \theta_3 = 90^o$ and $\theta_4 = 135^o$ (Vimal et al. 2021). Figure 3 illustrates the configuration of 4 GLCMs that correspond to the 4 directions with the distance fixed to $d = 1$. The extracted features from GLCMs include mainly the energy, the contrast, entropy, correlation, homogeneity, dissimilarity, the cluster shade, and the cluster prominence, which are explained briefly by:

- The energy (E_n): E_n expresses the regularity of the texture, which can be computed by:

$$E_n = \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} P_{(l,m)}^2 \tag{10}$$

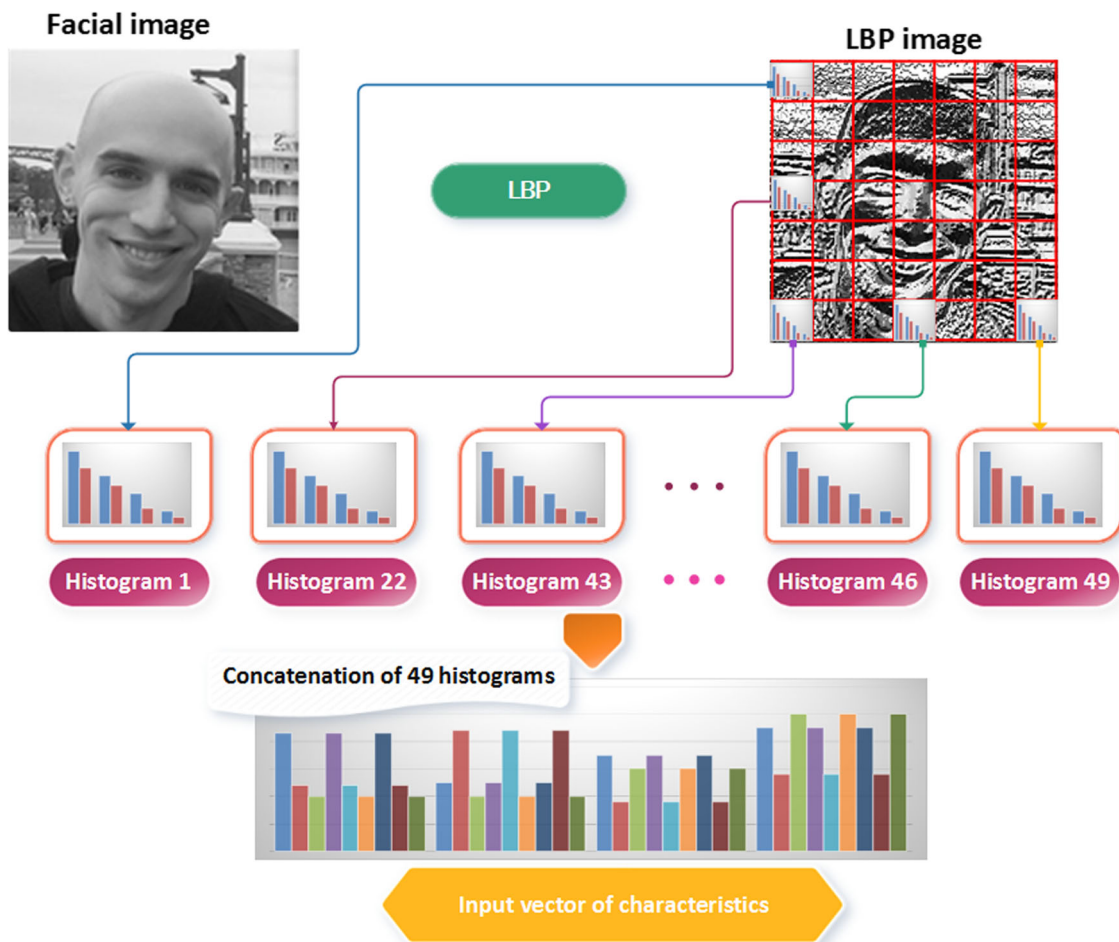It is important to note that a higher value of (E_n) signifies a completely homogeneous image.

Fig. 2 The coding face using a set of LBP histogram

- The contrast (C_n): It measures the rate of local variation in the picture (I). The formula of (C_n) is given by :

$$C_n = \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} P_{(l,m)} |l - m|^2 \tag{11}$$

- The entropy (E_t) : E_t is the inverse of energy and characterizes the irregular appearance of the image, hence a strong correlation between these two attributes. The formula of E_t is computed by :

$$E_t = \frac{1}{2log(N)} \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} P_{(l,m)} log_2 P_{(l,m)} \tag{12}$$

- Correlation (C_r): It can be compared to a measure of the linear dependence of gray levels in the image. It is calculated by:

$$C_r = \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} \frac{(l - \mu_l)(m - \mu_m)}{\sigma_l \sigma_m} P_{(l,m)} \tag{13}$$

- Homogeneity (H_m):The homogeneity changes inversely to the contrast and takes on high values if the differences between the analyzed pixel pairs are weak. Therefore, it is more sensitive to the element's diagonals of the GLCM, unlike the contrast, which depends more on the distant element's diagonal. It is measured by:

$$H_m = \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} \frac{P_{(l,m)}}{1 + |l - m|^2} \tag{14}$$

- Dissimilarity (D_s) : It expresses the same characteristics of the image as contrast to the difference that the weight of the GLCM inputs increases linearly as they move away from the diagonal rather than quadratically in the case of contrast.
It calculated by:

$$D_s = \frac{1}{(L-1)^2} \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} P_{(l,m)} |l - m| \tag{15}$$

- The cluster shade and the cluster prominence give information on the degree of symmetry of the GLCM.
  – The cluster shade is defined by :

$$C_s = \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} P_{(l,m)} (l - \mu_l + m - \mu_m)^3 \qquad (16)$$

whereas, the cluster prominence is given by:

$$C_p = \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} P_{(l,m)} (l - \mu_l + m - \mu_m)^4 \qquad (17)$$

Similar to multi-blocks LBP/HOG, the whole image is divided into 7×7 blocks, where from each block, the statistical moments GLCM are extracted and then combined to generate the descriptors vectors.

4 Archimedes optimization algorithm (AOA)

AOA is an algorithm inspired by physics, more particularly Archimedes' law. This algorithm is introduced by Fatma Hashim in 2020 and belongs to the class of metaheuristics (Hashim et al. 2020). The particularity of this algorithm lies in the encoding of the solution, which encompasses three auditory information: volume (V), density (D), and acceleration (Γ) to the basic agents. Hence, the group of agents is initially generated randomly in Dim dimensions. As additive data, random values of V, D and, Γ are provided. After, the evaluation process is realized for each object to determine the best object (O_{best}).

During the process of AOA, the update of density and volume is realized to change the acceleration based on the collision concept between objects, which play an important role in determining the novel position of current solution. The general steps of AOA are described as follows:

The first step–Initialization: This step aims to initialize randomly the real population that contains N objects using Eq. (18). Also, each object is characterized by their density (D_i), volume (V_i), and acceleration (Γ_i), which are defined in a random way using following Eqs. (19), (20), and (21):

$$O_i = O_i^{Min} + r_1 \times (O_i^{Max} - O_i^{Min}); i = 1, 2, ..., N \quad (18)$$
$$D_i = r_2 \qquad (19)$$
$$V_i = r_3 \qquad (20)$$
$$\Gamma_i = \Gamma_i^{Max} + r_4 \times (\Gamma_i^{Max} - \Gamma_i^{Min}); i = 1, 2, ..., N \quad (21)$$

where O_i represents the ith object, O_i^{Max} and O_i^{Min} are the maximal and minimal limits of the search space, respectively.

r_1, r_2, r_3, and r_4 are random vectors which belong to $[0, 1]^{Dim}$.

The population will be evaluated by computing the score for each object to determine the best object (O_{Best}) by joining their best values of density (D_{Best}), volume (V_{Best}), and acceleration (Γ_{Best}).

The second step–The update of densities and volumes In this step, the values of density and volume for each object are updated by the control of the best density and best volume using Eqs. (22) and ((23):

$$D_i^{t+1} = D^t + s_1 \times (D_{Best} - D_i^t) \qquad (22)$$
$$V_i^{t+1} = V_i^t + s_2 \times (V_{Best} - V_i^t) \qquad (23)$$

where s_1, s_2 are random scalars in [0, 1].

The third step–Transfer coefficient and density scalar: In this step, the collision between object is occurred until obtaining the equilibrium state. The principal role of transfer function (T_c) is to switch from exploration to exploitation mode, defined by Eq. (24):

$$T_c = \exp\left(\frac{t - T}{T}\right) \qquad (24)$$

The T_c increases exponentially over time until reaching 1. t is the current iteration, while T denotes the maximum number of iterations. Also, the decrease of density scalar d_s in AOA allows to find an optimal solution using Eq. (25):

$$d_s^{t+1} = \exp\left(\frac{t - T}{T}\right) - \left(\frac{t}{T}\right) \qquad (25)$$

The fourth step–Exploration phase: In this step, the collision between agents is occurred using a random selection of material (Mr). Thus, the update of acceleration objects is applied using Eq. (26) when the transfer function value is less or equal to 0.5.

$$\Gamma_i^{t+1} = \frac{D_{Mr} + V_{Mr} \times \Gamma_{Mr}}{D_i^{t+1} \times V_i^{t+1}} \qquad (26)$$

The fifth step–Exploitation phase: In this step, the collision between agents is not realized. Hence, the update of acceleration objects is applied using Eq. (27) when the transfer coefficient value is greater than 0.5.

$$\Gamma_i^{t+1} = \frac{D_{Best} + V_{Best} \times \Gamma_{Best}}{D_i^{t+1} \times V_i^{t+1}} \qquad (27)$$

where Γ_{Best} is the acceleration of the optimal object O_{Best}.

The sixth step–Normalization of acceleration: In this step, we normalize the acceleration to determine the rate of change using (28):

Fig. 3 An example of GLCMs based on different orientations
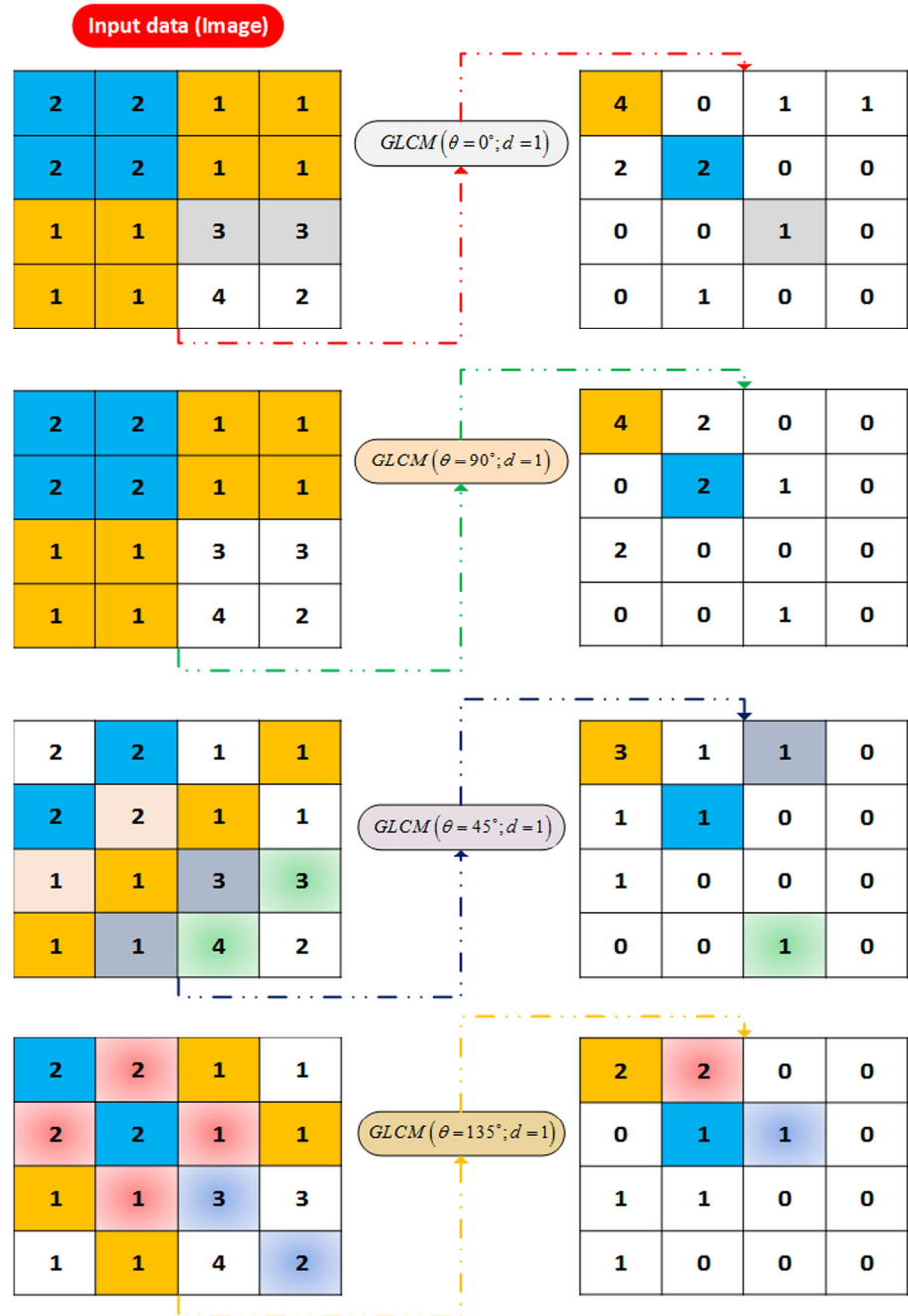
$$\Gamma_{i-norm}^{t+1} = \alpha \times \frac{\Gamma_i^{t+1} - \Gamma^{Min}}{\Gamma^{Max} - \Gamma^{Min}} + \beta \tag{28}$$

where α and β are fixed to 0.9 and 0.1, respectively. The Γ_{i-norm}^{t+1} shows the proportion of steps modified by each agent. A lower acceleration value indicates that the object is operating in the exploitation mode; otherwise, the object is operating in the exploration mode.

The seventh step–The Update process: For exploration phase ($T_c \leq 0.5$), the position of ith object in iteration $t+1$ is

modified by Eq. (29), whereas the object position is updated by Eq. (30) in exploitation phase ($T_c > 0.5$).

$$O_i^{t+1} = O_i^t + c_1 \times r_5 \times \Gamma_{i-norm}^{t+1} \times d_s \times (O_{rand} - O_i^t) \tag{29}$$

where c_1 is fixed to to 2.

$$O_i^{t+1} = O_{Best}^t + F \times c_2 \times r_6 \times \Gamma_{i-norm}^{t+1} \times d_s \\ \times (\delta \times O_{Best} - O_i^t) \tag{30}$$

where c_2 is fixed to 6.

The parameter δ is positively correlated with the time and this parameter is proportionally linked to the transfer coefficient T_c, i.e., $\delta = 2 \times T_c$. The main role of this parameter is to ensure a good balance between exploration and exploitation operations. During the first iterations, the margin between the best object and the other object is higher, which provides a high random walk. However, the margin will be reduced in last iterations and provide a low random walk.

F is employed for flagging which controls search direction using Eq. (31):

$$F = \begin{cases} +1 \ if \ \zeta \leq 0.5 \\ -1 \ if \ \zeta > 0.5 \end{cases} \quad (31)$$

where $\zeta = 2 \times rand - 0.5$.

The eighth step–The evaluation: In this step, we evaluate the novel population using score index Sc to determine the best object O_{Best} and the best additive information including D_{Best}, V_{Best}, and Γ_{Best}.

Algorithm 1 Pseudo code of AOA (Hashim et al. 2020).

1: **procedure** AOA(N,T, c_1 and c_2)
2:    Initialize N objects with their densities (D), volumes (V), and accelerations (Γ) using Eq.(18), Eq.(19), Eq.(20), and Eq.(21), respectively.
3:    Evaluate the score for each object.
4:    Determine the best object (O_{Best}) with their best density (D_{Best}), best volume (V_{Best}), and best acceleration (Γ_{Best}).
5:    $t = 1$
6:    **while** $t \leq T$ **do**
7:       **for** each object i **do**
8:          Adjust density and volume using Eq.(22) and Eq.(23)
9:          Adjust transfer coefficient (T_c) and density scalar (d_s) using Eq.(24) and Eq.(25), respectively.
10:          **if** $T_c \leq 0.5$ **then**   ▷ Exploration operation
11:             Adjust acceleration (Γ_i) using Eq.(26).
12:             Normalize acceleration (Γ_i) using Eq.(28)
13:             Adjust position using Eq.(29)
14:          **else**   ▷ Exploitation operation
15:             Adjust acceleration using Eq.(27).
16:             Adjust flagging control F using Eq.(31)
17:             Adjust position using Eq.(30)
18:          **end if**
19:          Compute the score of each object.
20:       **end for**
21:       Determine the best object (O_{Best}) with their best density (D_{Best}), best volume (Vl_{Best}) and best acceleration (Γ_{Best})..
22:       Set $t = t + 1$
23:    **end while**
24:    **return** Best object with their quality
25: **end procedure**

---

5 AOA-based FS for gender recognition

This section explains our system of gender recognition using Archimedes optimizer algorithm (AOA)-based feature selection. This system required three key points: the encoding solution, the evaluation of score, and the architecture system of gender recognition.

5.1 Structure of immersed object

This ingredient plays a vital role in the optimization process using physical or swarm algorithms.

The size of each object in AOA corresponds to the number of blocks, which is a 1-D vector with 49 elements. The vector components are randomly generated in the range [0, 1. At this stage, if the value is greater or equal to 0.5, thus the value is rounded to one. In this case, the bloc is considered as a relevant feature and coded by a histogram of (LBP, HOG, or GLCM). Contrarily, the block is ignored when the value is rounded to zero. This encoding aims to select the informative areas by realizing the concatenation of activated Multi-blocks-based histograms (LBP, HOG, or GLCM). The encoding object is shown in Fig. 4.

From Fig. 4, the first component(0.3) up to the 9th component (0.45) is less than 0.5. Thus, the current blocks will be inactivated. while the 10th (0.55), 11th, 12th, 46th, 47th, and 48th (0.62) components are greater than 0.5. As a result, the corresponding blocks will be selected to extract the features using HOG, LBP, or GLCM. This vector will be transformed to a matrix with a size of 7×7 to project the encoding solution on the image. The final vector of features represents the concatenation of activated histograms in the cases of LBP and HOG. However, a vector of parameters is used for GLCM.

According to this figure, we can see clearly that the current object selects fifteen blocks based histograms from 49 face areas, which will be concatenated and serve as the input of the neural network classifier.

5.2 Score evaluation

To apply the process of gender recognition using a wrapper feature selection assisted by AOA, a good compromise between accuracy and a lower number of features must be assured. Hence, the score for each object is computed by:

$$Sc = 0.99 \times Acc + 0.01 \times \left(\frac{D - d}{D} \right) \quad (32)$$

where (Acc) and (d) are the accuracy obtained by multilayer perceptron neural network (MLP) and the size of selected histograms, respectively.
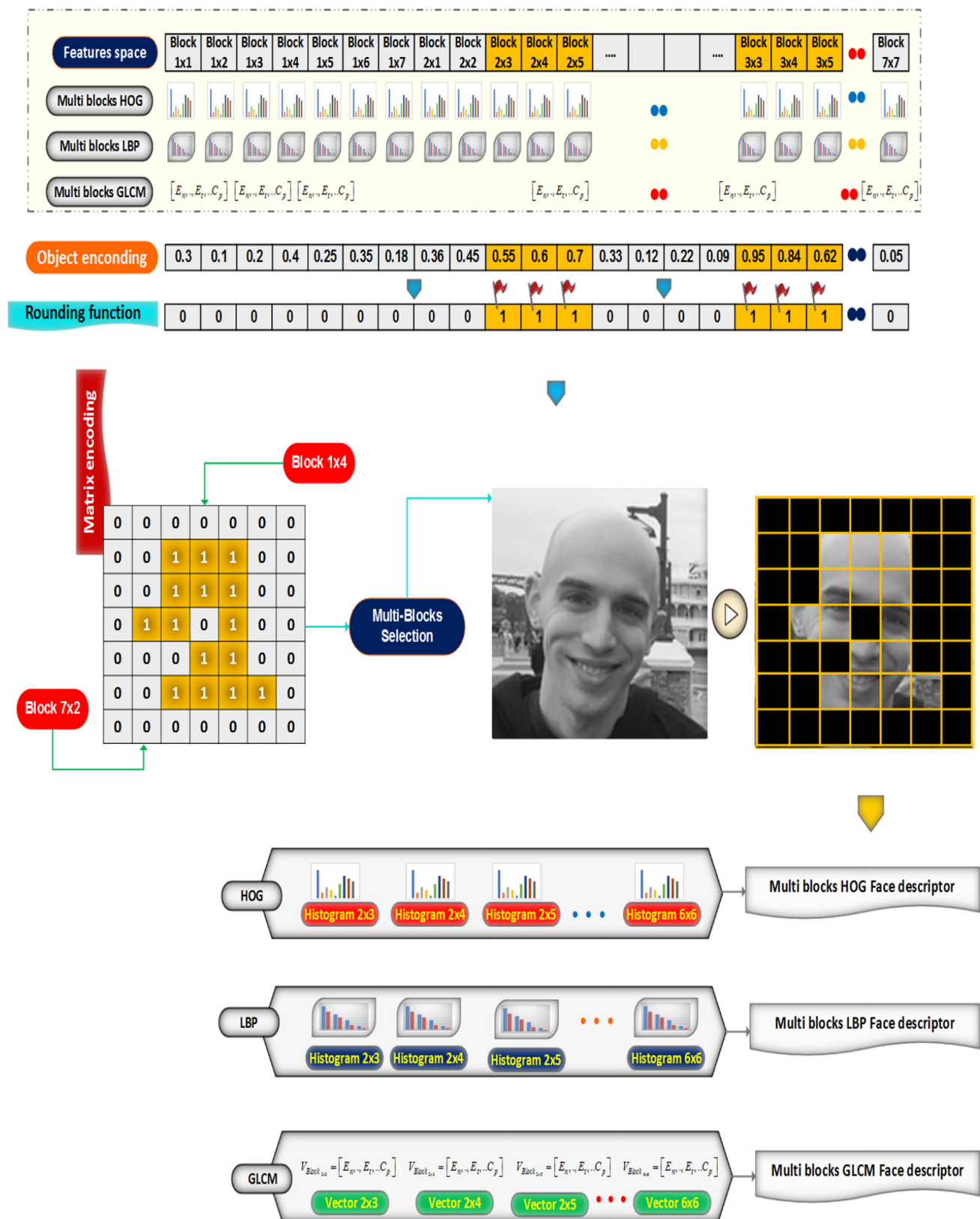
Fig. 4 Encoding solution

In Eq. (32), D is the total number of multi-blocks-based histogram extracted from original image.

The MLP is integrated as a classifier in the FS process using k-folds as a cross-validation strategy. In this study, the value of k is fixed to 5 to realize a fair comparison. Thus, 80% of samples is used in the training step, where the rest is used for testing. Additionally, the architecture of MLP is described in Fig. 5.

This architecture includes three layers:

Input layer It corresponds to the multi-blocks-based histograms (LBP/HOG) input features or GLCM vector. Hence, the number of neurons in this layer for LBP, HOG, and GLCM is equal to ($Blocks_{Selected} \times 59$) bins, ($Blocks_{Selected} \times 9$) bins and ($Blokcs_{Selected} \times 8$), respectively.

Hidden layer It contains the double of neurons used in the input layer.

Output layer It contains two neurons; the first corresponds to males, while the second is reserved for females.

In our study, the AOA aims to determine only the optimal patches for distinguishing between males and females. Thus, the MLP is designed for a classification task, knowing that the male output is coded by $[1, 0]^t$ and the female output is coded by $[0, 1]^t$.

The training configuration of MLP requires that the hidden neurons use the sigmoid function while the output neurons utilize softmax function. Also, the weights of MLP have been updated using a basic backpropagation algorithm, and the stop condition is indicated by the logical operator between the maximum number of iterations ((500) or the value of the mean square error (MSE), which is less than 10^{-5}.

It is important to indicate that the higher value of the computed score through all objects is assigned to the best object (O_{Best}).

5.3 Design framework

This part represents the core of our work, which consists of applying the AOA algorithm in gender recognition based on the selection of attributes. To better understand the proposed architecture, we preferred to explain the essential ingredients in bullet points:

- Initialization: This step is started by creating a random population of N immersed objects with 7×7 as dimension (49) elements.
- Encoding solution: This step aims to transform random objects into binary vectors to select the relevant blocks-based histograms or GLCM features.
- Selection of subset features: After decoding the object as illustrated in Fig. 4, the corresponding blocks are determined from the datasets based on LBP, HOG, or GLCM.

- Score evaluating: Each object generated by AOA indicates the subset of selected features must be assessed using an MLP classifier. The score represents the relationship between accuracy and the selected number of features, computed by Eq. 32.
- The update of position: The most important step in our architecture consists of applying a sequence of operators such as updating densities and volumes, exploring/ exploiting tasks, and norming of acceleration to produce better solutions with a higher score as shown in Algorithm 1.
- Stop condition: The cycle of AOA is an iterative process, controlled by the maximum number of iterations as a stop condition.

Figure 6 illustrates the overall steps of the AOA-based gender recognition and FS process. It is important to indicate that the MLP is performed using 5-fold cross-validation, which means that the MLP is trained 5 times and the average fitness evaluation is computed.

6 Experimental results

To realize a fair analysis, the efficiency of AOA is compared with different and recent computational algorithms inspired from swarm intelligence, mathematical algorithm, and physical algorithms, including HHO, MRFO, EPO, SCA, EO, HGSO, and MVO, tested on two scenarios.

The first scenario used smallest datasets captured in controlled conditions in the case of GT and FEI datasets.

The second scenario employed the challenger largest dataset, taken in a real environment in the Gallagher's dataset. All algorithms used the same conditions: the population size is fixed to ten solutions ($N = 10$), and the maximum number of iterations is fixed to 100 ($T = 10$). Also, three textural descriptors based on multi-blocks HOG, LBP, and GLCM, are implemented.

6.1 Simulation setup

6.1.1 Statistical metrics

To investigate the efficiency of the AOA algorithm in facial analysis-based FS, especially in gender recognition. The confusion matrix must be used and defined in Table 1. Next, some measures must be computed as Accuracy (Ac), Recall (Re), Precision (P), and F-score(F_{Sc}).

where: *TrP* The classifier identifies the person correctly knowing that their class is male;

TrN the classifier identifies the person correctly knowing that their class is female;

Fig. 5 The architecture of MLP
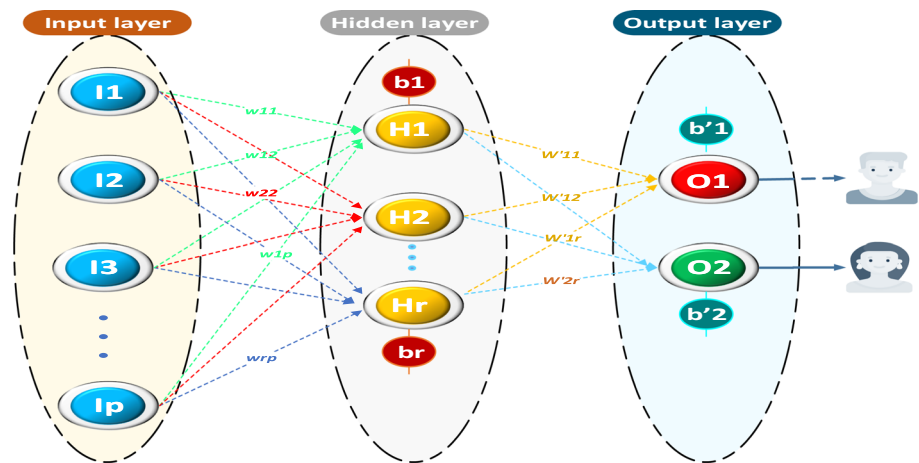
FaP the classifier assigns the person to the male class, knowing that the example belongs to the female class

FaN the classifier assigns the person to the female class, knowing that the example belongs to the male class.

- *The accuracy metric* (A) Among the most important metrics, we find the accuracy which measures the rate of correct data classification, defined by:

$$A = \frac{TrP + TrN}{TrP + FaN + FaP + TrN} \tag{33}$$

- *The recall metric* (R) This metric is also called true positive rate (TPR), which indicates the percentage of predicting male person as:

$$R = \frac{TrP}{TrP + FaN} \tag{34}$$

- *The precision* (P) It indicates the rate of true predicted samples as:

$$P = \frac{TrP}{FaP + TrP} \tag{35}$$

- *The fitness value metric* (Sc) This metric evaluates the performance of algorithms, which aims to maximize the accuracy and the rate of elimination of irrelevant features as computed in Eq. (32).
- *The size of selected features* (S_r) This metric implies the size of relevant features. It is computed as:

$$S_r = \frac{d}{D} \tag{36}$$

where d is the number of relevant blocks-based faces, which increase the performance of gender recognition.

- F-score (F_{Score}): In statistical F-score indicates the harmonic mean between recall and precision. It computed

Table 1 Confusion matrix

Actual	Predicted	
	Male	Female
Male	TrP	FaN
Female	FaP	TrN

by Eq. 37:

$$F_{Sc} = 2 \times \frac{R \times P}{R + P} \tag{37}$$

- CPU time (Cpu): It is the required time for each algorithm measured in second (S).

6.1.2 Parameters settings

It is important to enumerate the list of algorithms used for realizing the task of gender recognition from faces. This subsection defines all parameters used for each optimizer. The parameters settings of HHO, MRFO, EPO, SCA, EO, HGSO, and MVO are defined in Table 2.

6.2 Description of datasets

FEI dataset It is a Brazilian dataset that contains 200 individuals. It is important to note that each individual has 14 images, thus 2800. The images were captured on a white background and of color quality. The age category is between 19 and 40 years old. Some changes in the appearance of the face, such as hairstyle and adornments, have also been incorporated. The dataset is balanced because half of the examples are men and half are women.[3] We note that the resolution of each image is 640×480.

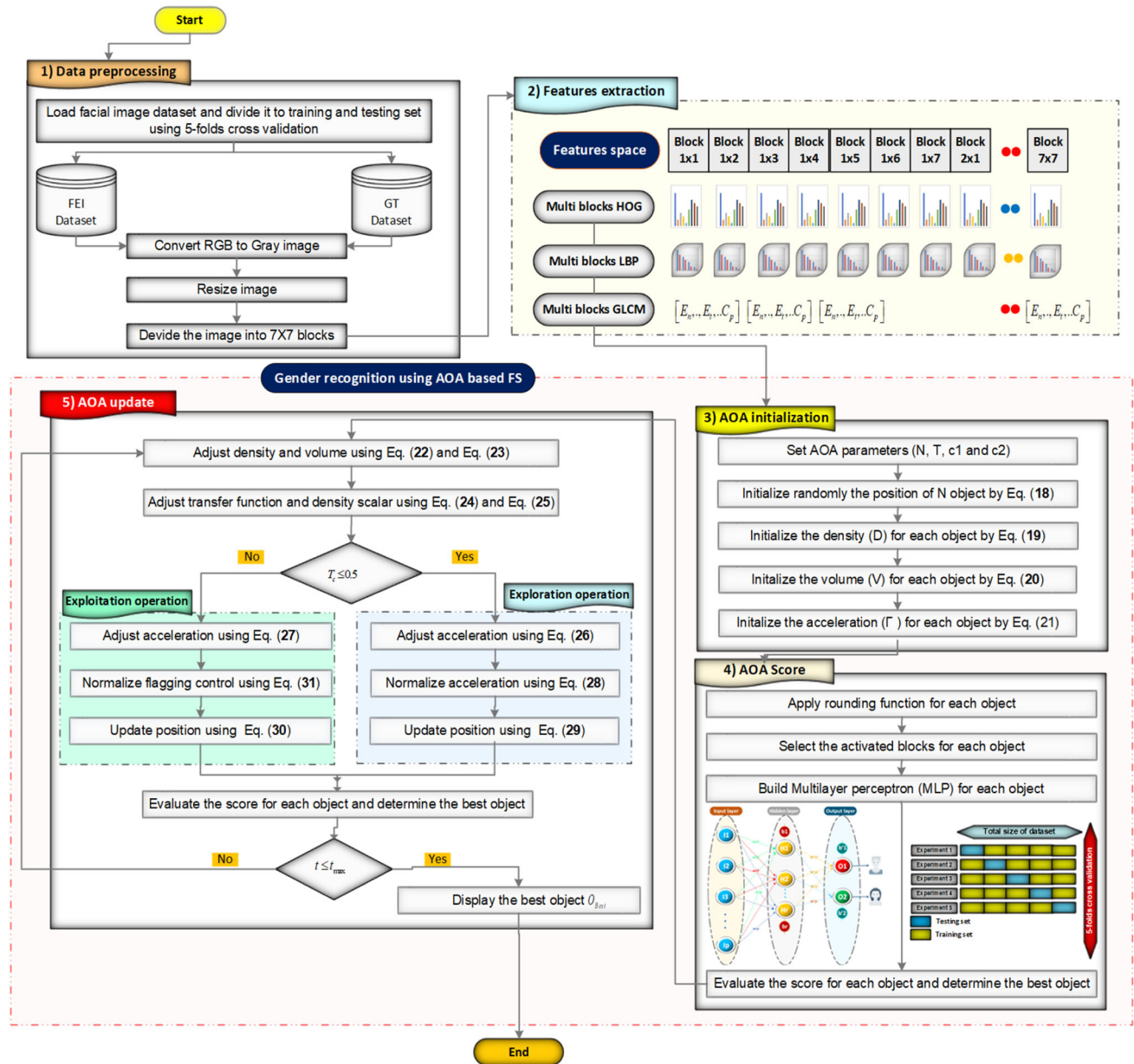[3] https://fei.edu.br/~cet/facedatabase.html.

Fig. 6 The design framework of AOA-based FS for gender recognition

Georgia Tech Face dataset (GT) It contains 50 people captured in two sessions between 04/06/99 and 11/15/99. Each person contains 15 images, hence a total of 750 images of size 640×480 pixels. The average size of a face is 150×150 pixels. Images are frontal expressed in light conditions, change of scale and expression. This dataset contains 7 women and 43 men. [4]

Gallagher's dataset This dataset is a massive, publicly accessible database that categorizes people by gender and age. It comprises almost 28,000 low-resolution photographs

collected from Flickr images taken in natural light [5]. All images are captured as accurately as possible to reflect real-world scenarios, including variances in appearance, pose, dark glasses, lighting conditions, and image quality (Gallagher and Chen 2009). According to the book of (Grother et al. 2019), this database presents a major challenge in determining gender and age.

Table 2 Parameters settings of physical, mathematical and swarm inspired algorithms

Algorithms	Parameters setting
Common settings	Population size ($N = 10$)
	Maximum number of iterations ($T = 100$)
	Maximal limit=1
	Minimal limit=0
	Dimension corresponds to the number of blocks ($D = 49$)
AOA (Hashim et al. 2020)	$c_1 = 2$ and $c_2 = 6$
	$\alpha = 0.9$ and $\beta = 0.1$
EO (Faramarzi et al. 2020)	$a_1 = 2$
	$a_2 = 1$
	Generation Probability ($GP = 0.5$)
MVO (Mirjalili et al. 2016)	Wormehole Existance Prob
	($WEP_{max} = 1$ and $WEP_{min} = 0.5$)
	Traveling Distance Rate ($TDR \in [0.5; 1]$)
HGSO (Hashim et al. 2019)	Clusters number=2
	$M_1 = 0.1$ and $M_2 = 0.2$
	$\alpha = \beta = 1$ and $K = 1$
	$l_1 = 5E - 03, l_2 = 1E + 02$ and $l_3 = 1E - 02$
EPO (Dhiman and Kumar 2018)	Temperature Profile ($T' \in [1; 1000]$)
	$A \in [-1, 1.5]$
	$M = 2$
	$f \in [2, 3]$
	$l \in [1.5, 2]$
	Function $S() \in [0, 1.5]$
SCA (Mirjalili 2016)	$a \in [2, 0]$
MRFO (Zhao et al. 2020)	$S = 2$
HHO (Heidari et al. 2019)	$\beta = 1.5$

6.3 Results and discussion of smallest datasets (FEI and GT datasets)

This subsection shows the first scenario, which applies wrapper FS using optimization algorithms, tested on standard benchmarking datasets (FEI and GT datasets) and taken in controlled situations with few images.

In terms of fitness and Cpu time Table 3 summarizes the results of the AoA algorithm based on wrapper FS in terms of fitness by varying the extraction methods (HOG, LBP, and GLCM), tested on two GT and FEI databases to achieve gender recognition.

By analyzing the results of Table 3, we firstly notice a precise performance obtained by the AOA algorithm based on the three extraction methods for the two GT and FEI databases. Secondly, AOA-based HOG achieves a higher value of of fitness than AOA-based on LBP and GLCM for both datasets.

Furthermore, the comparable results can be observed between AOA-based HOG and AOA-based GLCM for both datasets. Also, it is essential to show that EO is ranked second place for both datasets.

The time consumed by the optimization methods based on wrapper Fs varying the extraction methods is shown in Table 4. The SCA algorithm is the fastest for the FEI database compared to other competitors based on the three descriptors (HOG, LBP, and GLCM). EO requires less time for the GT database, based on LBP and GLCM than other optimizers. Note also that the HOG-based EPO algorithm is ranked first in terms of execution time compared to other algorithms.

In terms of Accuracy and Selection ratio Tables 5 and 6 illustrate the accuracy and selection ratio results based on wrapper AOA and other optimization algorithms by varying descriptors features such as HOG, LBP, and GLCM.

According to the results of Table 5, AOA based on the three descriptors (AOA-HOG, AOA-LBP, and AOA-GLCM) produces accurate results for both datasets in terms of accuracy compared to other algorithms, including HHO, SCA, EO, EPO, MRFO, HGSO, and MVO. Also, AOA-based HOG recognizes the gender of persons with a rate of 90.34% and

Table 3 The impact of features descriptors on the performance of AOA against other recent optimizers over fitness measures

Fitness	GT dataset			FEI dataset		
Algorithms	HOG	LBP	GLCM	HOG	LBP	GLCM
HHO	0.8947	0.8644	0.8913	0.9776	0.9262	0.9849
SCA	0.8914	0.8653	0.8873	0.9820	0.9518	0.9841
EO	0.9002	0.8642	0.8927	0.9853	0.9461	0.9829
EPO	0.8966	0.8658	0.8887	0.9804	0.9351	0.9780
MRFO	0.8950	0.8593	0.8902	0.9837	0.9412	0.9853
HGSO	0.8902	0.8543	0.8833	0.9837	0.9298	0.9834
MVO	0.8981	0.8589	0.8954	0.9805	0.9335	0.9825
AOA	**0.9015**	**0.8690**	**0.8969**	**0.9882**	**0.9534**	**0.9858**

Table 4 The impact of features descriptors on the performance of AOA against other recent optimizers over Cpu Time measures

CPU time	GT dataset			FEI dataset		
Algorithms	HOG	LBP	GLCM	HOG	LBP	GLCM
HHO	446.6000	466.0300	363.8200	249.3200	295.8900	258.3600
SCA	301.8600	211.4300	264.7200	**181.5400**	**118.8400**	**144.1000**
EO	324.7700	**182.8200**	221.4100	250.8200	214.4500	295.8000
EPO	**285.5800**	272.2800	324.0800	248.1200	126.0400	211.3700
MRFO	301.8600	445.4700	468.4100	302.8600	226.5200	218.7900
HGSO	454.8300	398.6100	312.1300	282.4900	199.3700	229.3200
MVO	368.0900	430.3200	369.4800	375.0500	272.1800	229.4900
AOA	391.3700	332.3200	388.8600	197.0700	231.6700	340.9200

99.16% for GT and FEI, respectively. This behavior can be interpreted by the excellent balance between exploration and exploitation of AOA and the efficiency of HOG descriptors based on orientation histogram.

From Table 6, The AOA-based HOG keeps the most area of face for GT, i.e., 14 blocks are selected from 49, which presents the best performance in blocks selection. Concerning the FEI dataset, AOA-based HOG is ranked second with keeping 17 blocks from 49, while SCA allows eliminating many irrelevant blocks in the case of HOG, LBP, and GLCM. Also, SCA-based LBP and GLCM obtain fewer informative gender faces area for the GT dataset than other optimizers.

In terms of recall and precision The comparison of eight wrapper FS algorithms using three descriptors (HOG, LBP, and GLCM) based on recall and precision are illustrated in Tables 7 and 8. By inspecting the results of precision measure, we can see that AOA- and MRFO-based HOG for the FEI dataset provides the same performance with 99.50% as precision. Also, HGSO- and AOA-based GLCM attains the same performance in terms of precision for the FEI dataset.

From Table 8, AOA-based HOG and GLCM provide better performance than other optimizers for both datasets. Moreover, MRFO- based HOG provides the same performance for FEI as AOA-based HOG, whereas it takes the second rank for the GT dataset.

In terms of F-score Table 9 indicates the values of F-score obtained by AOA and other optimizers by employing three descriptors features like HOG, LBP, and GLCM for both datasets. By inspecting the obtained results, we can see that F-score values over FEI are significantly higher than GT dataset due to the balanced samples of gender categories. For the GT dataset, a great competition between MRFO- and AOA-based HOG is highlighted with a slight advantage for MRFO because the margin is significantly lower with a value of 0.0008. In addition, AOA-based GLCM is still better for GT datasets than other algorithms including, HHO, SCA, EO, EPO, MRFO, HGSO, and MVO. Also, wrapper FS techniques based on LBP show lower F-score values than others descriptors for both datasets (GT and FEI).

For FEI-, AOA-, and MRFO-based HOG demonstrates a strong efficiency of Fscore with 99.50% as performance compared to other algorithms such as HHO, SCA, EO, EPO, HGSO, and MVO. Also, AOA-based GLCM reached the same performance as AOA and MRFO based HOG.

6.4 Results and discussion of largest dataset (Gallagher's dataset)

This subsection shows the second scenario, which applies wrapper FS using optimization algorithms, tested on real

Table 5 The impact of features descriptors on the performance of AOA against other recent optimizers over Accuracy measures

Accuracy	GT dataset			FEI dataset		
Algorithms	HOG	LBP	GLCM	HOG	LBP	GLCM
HHO	0.8990	0.8689	0.8960	0.9831	0.9320	0.9887
SCA	0.8914	0.8660	0.8887	0.9841	0.9534	0.9864
EO	0.9031	0.8651	0.8953	0.9889	0.9506	0.9864
EPO	0.8993	0.8669	0.8917	0.9840	0.9370	0.9815
MRFO	0.9028	0.8636	0.8944	0.9893	0.9456	0.9889
HGSO	0.8944	0.8578	0.8873	0.9881	0.9349	0.9886
MVO	0.9020	0.8624	0.8997	0.9865	0.9392	0.9875
AOA	**0.9034**	**0.8701**	**0.9014**	**0.9916**	**0.9561**	**0.9904**

Table 6 The impact of features descriptors on the performance of AOA against other recent optimizers over Selection ratio measures

Selection ratio	GT dataset			FEI dataset		
Algorithms	HOG	LBP	GLCM	HOG	LBP	GLCM
HHO	0.5306	0.5714	0.5714	0.5714	0.6122	0.3878
SCA	0.3673	**0.2041**	**0.2449**	**0.2245**	**0.2041**	**0.2449**
EO	0.3878	0.5306	0.3673	0.3673	0.4898	0.3673
EPO	0.3673	0.2449	0.4082	0.3673	0.2449	0.3673
MRFO	0.8776	0.5714	0.5306	0.5714	0.4898	0.3673
HGSO	0.5306	0.4898	0.5102	0.4490	0.5714	0.5306
MVO	0.4898	0.4898	0.5306	0.6122	0.6327	0.5102
AOA	**0.2857**	0.2449	0.5510	0.3469	0.6327	0.4694

Table 7 The impact of features descriptors on the performance of AOA against other recent optimizers over Recall measures

Recall	GT dataset			FEI dataset		
Algorithms	HOG	LBP	GLCM	HOG	LBP	GLCM
HHO	0.9040	0.8747	0.9000	0.9875	0.9425	0.9900
SCA	0.8960	0.8693	0.8907	0.9850	**0.9550**	0.9900
EO	0.9067	0.8693	0.8987	0.9925	0.9475	0.9900
EPO	0.9040	**0.8813**	0.8960	0.9875	0.9400	0.9825
MRFO	**0.9107**	0.8707	0.8973	**0.9950**	0.9500	0.9925
HGSO	0.9000	0.8613	0.8880	0.9925	0.9400	**0.9950**
MVO	0.9080	0.8693	0.9040	0.9925	0.9450	0.9925
AOA	0.9040	0.8733	**0.9080**	**0.9950**	0.9475	**0.9950**

Table 8 The impact of features descriptors on the performance of AOA against other recent optimizers over Precision measures

Precision	GT dataset			FEI dataset		
Algorithms	HOG	LBP	GLCM	HOG	LBP	GLCM
HHO	0.9014	**0.8677**	0.8949	0.9903	0.9347	0.9951
SCA	0.8891	0.8400	0.8870	0.9880	0.9560	0.9877
EO	0.9018	0.8414	0.8927	0.9928	**0.9642**	0.9904
EPO	0.8960	0.8475	0.8883	0.9882	0.9383	0.9881
MRFO	0.9052	0.8352	0.8913	**0.9951**	0.9510	0.9929
HGSO	0.8908	0.8361	0.8874	0.9928	0.9417	0.9927
MVO	0.8974	0.8399	0.8977	0.9927	0.9478	0.9927
AOA	**0.9099**	0.8529	**0.8980**	**0.9951**	0.9457	**0.9951**

Table 9 The impact of features descriptors on the performance of AOA against other recent optimizers over F-score measures

F-score	GT dataset			FEI dataset		
Algorithms	HOG	LBP	GLCM	HOG	LBP	GLCM
HHO	0.8977	0.8427	0.8953	0.9887	0.9379	0.9924
SCA	0.8892	0.8428	0.8866	0.9862	**0.9549**	0.9888
EO	0.9026	0.8400	0.8922	0.9925	0.9547	0.9900
EPO	0.8981	**0.8496**	0.8898	0.9875	0.9388	0.9848
MRFO	**0.9058**	0.8328	0.8896	**0.9950**	0.9499	0.9925
HGSO	0.8918	0.8359	0.8854	0.9925	0.9397	0.9938
MVO	0.9004	0.8351	0.8967	0.9925	0.9449	0.9925
AOA	0.9050	0.8422	**0.9009**	**0.9950**	0.9453	**0.9950**

challenger Gallagher's dataset, taken in uncontrolled situations with a high number of images.

Table 10 summarizes the results of GR using AOA-HOG against other optimizers, including HHO, SCA, EO, EPO, MRFO, HGSO, and MVO. It can be seen clearly that AOA-HOG generates precise results in terms of fitness, accuracy, recall, precision, Fscore, and selection ratio. Also, AOA based on the HOG descriptor ensures a good compromise between higher accuracy and a lower selection ratio. The numerical results indicate that the performance of correct gender classification is 95.51% by selecting a reduced number of patches, more precisely 20 patches out of 49, i.e., 40.82%. We can observe that the second best optimizer is MRFO, which achieved 95.01% of accuracy, but it contains a high number of selected patches, with 39 zones from 49.

Regarding Fscore, AOA-HOG reached a higher rate with 95.88%, which ensures a good balance between the binary classes (female and male).

In terms of CPU time, AOA-HOG is faster than other competitors, with 573.8 s, while the second-best optimizer is EO, with an 8490.305 s.

Table 11 indicates the results of different optimizers based on the LBP descriptor in terms of fitness, accuracy, recall, precision, Fscore, selection ratio, and CPU time. It can be remarked that AOA is the best algorithm in terms of correct gender classification compared to other competitors, with 96.08% accuracy. Also, AOA-LBP reached a lower number of selected patches, with 10 areas only from 49, i.e., 20.49%. Furthermore, the second rank is obtained by SCA in terms of accuracy and selection ratio, with 95.28% and 26.53%, respectively.

AOA still shows a better performance with 96.25% than other competitors in erms of Fscore.

In terms of time consuming, AOA-LBP is faster than the rest of the algorithms with 391.69 s.

Table 12 reports the gender recognition results tested on Gallagher's dataset using AOA, HHO, SCA, EO, EPO, MRFO, HGSO, and MVO based on the GLCM descriptor. From this table, we can see great competition between SCA and AOA in terms of fitness, accuracy, precision, and selection ratio.

For example, in terms of accuracy and fitness, SCA outperforms AOA with a low margin of 0.02% and 0.04%, respectively. Additionally, the selection ratio obtained by SCA is slightly higher than that of the AOA algorithm, with a margin of 2.4%. Furthermore, AOA outperforms SCA ith a low margin of 0.02% in terms of precision.

In terms of recall and Fscore, both algorithms AOA and SCA based on GLCM achieve the same performance with 95.25%.

In terms of CPU time, AOA is faster than SCA with a margin of 699.597s.

6.4.1 Graphical analysis

- Fitness curve analysis: The fitness curves obtained by the different optimizers, including AOA, HHO, SCA, EO, EPO, MRFO, HGSO, and MVO are shown in Fig. 7.

By analyzing the behavior of the convergence of the AOA algorithm for the smallest databases (FEI and GT) based on the different descriptors, a clear growth is illustrated by increasing the number of iterations compared to other algorithms, including EO, MVO, EPO, MRFO, HHO, SCA, and HGSO. This phenomenon is justified by a better balance between exploitation and exploration, making it possible to avoid convergence toward local minima. However, for largest database (Gallagher's dataset), AOA-HOG and AOA-LBP maintain a clear superiority during iterations, while AOA based on GLCM shows a clear growth until the 90th iteration, and the SCA-GLCM takes over between the 90th and 100th iteration, which will finish in first position with a margin of 0.04%.

For smallest datasets (FEI and GT), we can conclude that AOA based on the HOG descriptor provides a higher fitness value than other descriptors as GLCM and LBP.

Table 10 The performance results of AOA based on HOG descriptor against other recent optimizers for Gallagher's dataset

Algorithms	Gallagher's dataset – HOG Descriptor						
	Fitness	Accuracy	Recall	Precision	Fscore	Selection R	CPU tiime
HHO	0.9336	0.9405	0.9476	0.9477	0.9476	0.5510	18317.6000
SCA	0.9125	0.9137	0.9165	0.9143	0.9152	0.4490	19271.4500
EO	0.9387	0.9405	0.9416	0.9426	0.9420	0.4898	8490.3050
EPO	0.9100	0.912	0.9130	0.9144	0.9136	0.5306	16416.4600
MRFO	0.9451	0.9501	0.9549	0.9553	0.9550	0.5306	9257.9900
HGSO	0.9167	0.9183	0.9212	0.9198	0.9204	0.4898	1449.6810
MVO	0.9169	0.9196	0.9202	0.9230	0.9213	0.5918	1319.9640
AOA	**0.9515**	**0.9551**	**0.9600**	**0.9583**	**0.9588**	**0.4082**	**573.8000**

Table 11 The performance results of AOA based on LBP descriptor against other recent optimizers for Gallagher's dataset

Algorithms	Gallagher's dataset – LBP Descriptor						
	Fitness	Accuracy	Recall	Precision	Fscore	Selection R	CPU tiime
HHO	0.9364	0.9419	0.9475	0.9482	0.9475	0.6122	1564.6520
SCA	0.9506	0.9528	0.9550	0.9578	0.9550	0.2653	745.7800
EO	0.8754	0.8742	0.8693	0.8744	0.8708	0.3061	6460.6420
EPO	0.9461	0.9481	0.9500	0.9503	0.9500	0.3878	615.9800
MRFO	0.9383	0.9422	0.9450	0.9504	0.9463	0.4490	615.8400
HGSO	0.9192	0.9233	0.9275	0.9287	0.9276	0.4898	6460.6400
MVO	0.9310	0.9342	0.9375	0.9383	0.9375	0.3878	1449.0000
AOA	**0.9592**	**0.9608**	**0.9625**	**0.9626**	**0.9625**	**0.2041**	**391.6900**

Table 12 The performance results of AOA based on GLCM descriptor against other recent optimizers for Gallagher's dataset

Algorithms	Gallagher's Dataset – GLCM Descriptor						
	Fitness	Accuracy	Recall	Precision	Fscore	Selection R	CPU tiime
HHO	0.9425	0.9475	0.9525	0.9533	0.9524	0.5510	1861.5200
SCA	**0.9486**	**0.9505**	**0.9525**	0.9535	**0.9525**	**0.2449**	1106.1770
EO	0.9470	0.9501	0.9520	0.9530	0.9520	0.4490	1566.2000
EPO	0.9408	0.9441	0.9475	0.9486	0.9475	0.3878	1600.2400
MRFO	0.9445	0.9457	0.9495	0.9496	0.9495	0.5714	1560.8800
HGSO	0.9417	0.9452	0.9500	0.9483	0.9488	0.4082	453.2250
MVO	0.9437	0.9493	0.9524	0.9534	**0.9525**	0.6122	2045.6700
AOA	0.9482	0.9503	**0.9525**	**0.9537**	**0.9525**	0.2653	**406.5800**

Concerning the largest dataset (Gallagher's dataset), It can be seen that AOA-LBP reached a higher value of fitness with 95.92%, compared to AOA-HOG with 95.15% and AOA-GLCM with 94.82%.

- ROC curve analysis: In addition, we have graphically represented the ROC curve, as shown in Fig. 8, which relates true positive rate (TPR) as a function of false positive rate (FPR) by using the different algorithms based on the three descriptors (HOG, LBP, and GLCM) for all the images of the three corpora (FEI, GT, and Gallagher's datasets).

We notice that AOA shows a clear performance in terms of AUC compared to other optimizers. For the FEI dataset, the values of AUC obtained by AOA-based on HOG, LBP, and GLCM are 0.9951, 0.9606, and 0.9951, respectively. However, for the GT dataset obtained 0.9099, 0.8529, and 0.9027, respectively.

Concerning largest dataset (Gallagher's dataset), the values of AUC obtained by AOA-LBP, AOA-HOG, and AOA-GLCM are 0.9626, 0.9583, and 0.9537, respectively.

- Visual examples of selected patches: Fig. 9 shows the optimal patches obtained by the proposed method AOA based on three handcrafted descriptors: HOG, LBP, and GLCM for two individuals from Gallagher's dataset. It can be seen clearly that AOA-LBP finds a lower number of patches, with 10 zones out of 49, compared to AOA-HOG with 20 and AOA-GLCM with 13.

AOA-HOG appears to cover more important facial areas for gender recognition, such as the eyes, nose, and mouth, whereas LBP and GLCM detect sparse areas and focus on the edges. As shown in Table 5, these behaviors reflect HOG's superior overall accuracy.

6.4.2 Statistical analysis

To validate the efficiency of AOA to other competitive algorithms, a statistical study is required. Thus, this study is validated by Wilcoxon rank-sum test between the fitness values obtained by AOA and other algorithms, including HHO, SCA, EO, EPO, MRFO, HGSO, and MVO.

From Table 15, we can observe for the smallest datasets (FEI and GT) that AOA is statistically substantial to all competitors in the case of the GLCM descriptor ($p < 0.05$). Also, the same behavior is highlighted for the FEI dataset when the descriptor is HOG. Additionally, AOA is not substantial to MVO- based HOG and EPO-based LBP in the case of the GT datasets. Also, EO-based LBP in the case of the FEI dataset dominates AOA.

For largest dataset (Gallagher's dataset), AOA is statistically substantial in the case of HOG and LBP descriptors against HHO, SCA, EO, EPO, MRFO, HGSO, and MVO. Furthermore, in the case of GLCM, AOA is substantial to all optimizers except HHO and MRFO. In general, AOA shows a good performance in terms of Wilcoxon's test.

6.5 Comparative study

6.5.1 Comparative study with the existing works

A novel comparative study with the literature works has been realized further to explore the power of the proposed system AOA-based HOG. Mainly, we have selected some works from the literature based on machine learning, deep learning, and genetic-based FS. It is essential to show that the comparison is tough because the most authors used different conditions

FEI dataset A deep analysis of Table 13 indicates that the proposed method, AOA-based on multi-blocks HOG with BPNN as classifier achieves a higher accuracy performance, i.e., 99.16% as correct prediction rate of gender from faces. Most machine learning methods used mainly SVM classifier with multi-features based on handcrafted techniques inspired by LBP, LDP, and HOG. In this category, the SVM proposed by (Geetha et al. 2019) represents the best classifier that can predict the gender correctly with high accuracy of 99% compared to other ML methods. The deep-gender provided by Haider et al. (2019) shows a good accuracy with 98.75%. Furthermore, a wrapper FS based on a genetic algorithm is employed to predict gender from the face, using Eigen-space of features. The authors utilized BPNN as a classifier, whereas GA selects the significant Eigen vectors, which obtained 96% as accuracy. In conclusion, AOA-based HOG and AOA-based GLCM are the best classifier for predicting gender from the face compared to other approaches inspired by ML and deep CNN.

GT dataset For this dataset, a novel run is realized due to the value of *k*-folds. We note that some works of literature used *2*-fold. Hence, Table 14 highlights the performance results of AOA with some methods of machine learning (ML) methods using *2*-fold across the correct rate of gender identification.

The accuracy findings reveal that the AOA-based multi-blocks HOG approach outperforms other ML methods such as SVM based on combined DWT (discrete wavelet transform) and DCT (discrete cosine transform). Also, it is essential to show that the multi-blocks LBP based-AOA reached a lower performance with 97.25% in terms of the correct rate of gender identification. The AOA-based multi-blocks GLCM is ranked in the second position with 99.15% of accuracy, followed by the work of (Goel and Vishwakarma 2016a), which applied SVM based on DCT and reached 98.96% in accuracy.

Gallagher's dataset Table 16 shows the results obtained from the proposed methods (AOA-HOG, AOA-LBP, and AOA-GLCM) as well as from the state-of-the-art, including machine learning (ML) and deep learning (DL) techniques, tested on the largest Gallagher's dataset captured in uncontrolled conditions. From this table, it can be observed that AOA-LBP outperforms other variants of AOA, i.e., AOA-HOG and AOA-GLCM. Furthermore, AOA-LBP provides better performance in terms of accuracy compared to ML and DL methods. The numerical results indicate that AOA-LBP reached 96% of accuracy, while Ubunsa CNN and Bagging+LBP+HOG achieved 91.48% and 88.01%, respectively.

6.5.2 Comparative study with the basic MLP

To confirm the effectiveness of AOA, an ablation study is realized between AOA and MLP. Furthermore, the input of MLP considers all patches to extract features using HOG, LBP, or GLCM descriptors. The stop condition of backpropagation algorithm is the logical operator between max iteration (500) or mean square error ($MSE < 10^{-5}$)

Figure 10 illustrates the performance of AOA based on HOG, LBP, and GLCM against MLP in terms of accuracy, recall, precision, and Fscore.

In terms of accuracy for FEI dataset, MLP based on LBP, HOG, and GLCM achieve 90%, 90%, and 87%, respectively. While, AOA based on LBP, HOG, and GLCM reach 96%, 99%, and 99%, respectively.

Table 13 Statistical study using Wilcoxon's test (In bold best values $p < 0.05$, which implies that AOA is substantial against algorithm X)

AOA versus	GT dataset			FEI dataset			Gallagher's datase		
	HOG	LBP	GLCM	HOG	LBP	GLCM	HOG	LBP	GLCM
HHO	**4.40E-21**	**2.06E-14**	**5.00E-02**	**1.26E-32**	**3.59E-30**	**1.23E-39**	**9.57E-27**	**1.22E-21**	5.12E-01
SCA	**2.20E-02**	**1.71E-02**	**6.86E-31**	**9.74E-27**	**1.40E-04**	**2.62E-38**	**3.58E-36**	**4.20E-14**	**3.37E-04**
EO	**7.20E-03**	**3.37E-02**	**1.50E-05**	**2.70E-13**	2.64E-01	**1.33E-38**	**8.42E-25**	**4.76E-36**	**2.93E-04**
EPO	**1.04E-02**	8.09E-01	**6.97E-21**	**8.87E-23**	**2.86E-17**	**7.37E-42**	**4.43E-38**	**1.47E-10**	**2.77E-16**
MRFO	**5.11E-07**	**2.38E-10**	**1.13E-06**	**7.25E-19**	**1.91E-05**	**5.49E-39**	**3.58E-36**	**5.06E-20**	1.17E-01
HGSO	**3.58E-25**	**5.65E-26**	**1.11E-36**	**1.15E-19**	**4.36E-30**	**1.08E-39**	**7.35E-37**	**7.13E-35**	**2.30E-03**
MVO	9.91E-02	**2.20E-04**	**3.16E-17**	**1.11E-19**	**1.30E-17**	**1.31E-40**	**4.45E-38**	**1.21E-28**	**4.22E-05**

Table 14 Comparative performance in terms of accuracy with the existing approaches–FEI dataset

References	Classifier	Extracted features	Accuracy
Micheal and Geetha (2019)	SVM	DRLBP++RILPQ+PHOG	95.30%
Geetha et al. (2019)	SVM	8-LDP+LBP	99%
Ghojogh et al. (2018)	LDA+weighting vote	Intensity of lower part of face	94%
Haider et al. (2019)	Deepgender	*	98.75%
Zhou and Li (2019)	GA-BPNN	Eigen-features based on PCA	96%
Khan et al. (2019)	MSFS-CRFs	Segmentation based on Super-Pixels	93.70%
Kumar et al. (2019)	SVM	Multi-features (BoW+SIFT)	98%
Proposed method	AOA-BPNN	Multi-blocks HOG	**99.16%**
		Multi-blocks LBP	95.61%
		Multi-blocks GLCM	99.04%

Table 15 Comparative performance in terms of accuracy with the existing methods– GT dataset

References	Classifier	Extracted features	Accuracy
Goel and Vishwakarma (2016a)	SVM (2-folds)	DCT	98.96%
Goel and Vishwakarma (2016)	SVM (2-folds)	KPCA	97.38%
Goel and Vishwakarma (2016b)	SVM (2-folds)	DWT+DCT	99%
Proposed method	AOA-BPNN (2-folds)	Multi-blocks HOG	**99.50%**
		Multi-blocks LBP	97.25%
		Multi-blocks GLCM	99.15%

Table 16 Comparative performance in terms of accuracy with the existing approaches–Gallagher's dataset under Dago's protocol

| References	Classifier	Extracted features	Accuracy
Dago-Casas et al. (2011)	SVM	Gabor+PCA	86.01%
Castrillón-Santana et al. (2013)	Bagging	LBP+HOG	88.01%
Castrillón-Santana et al. (2016)	SVM	HOG+LBP^{u2}+LTP+WLD	82.91%
Mansanet et al. (2016)	Local DNN	*	90.58%
Orozco et al. (2017)	Ubunsa CNN	*	91.48%
Abdalrady and Aly (2020)	2-stage PCANet	*	89.65%
Proposed method	AOA-BPNN	Multi-blocks HOG	95.51%
		Multi-blocks LBP	**96.08%**
		Multi-blocks GLCM	95.03%
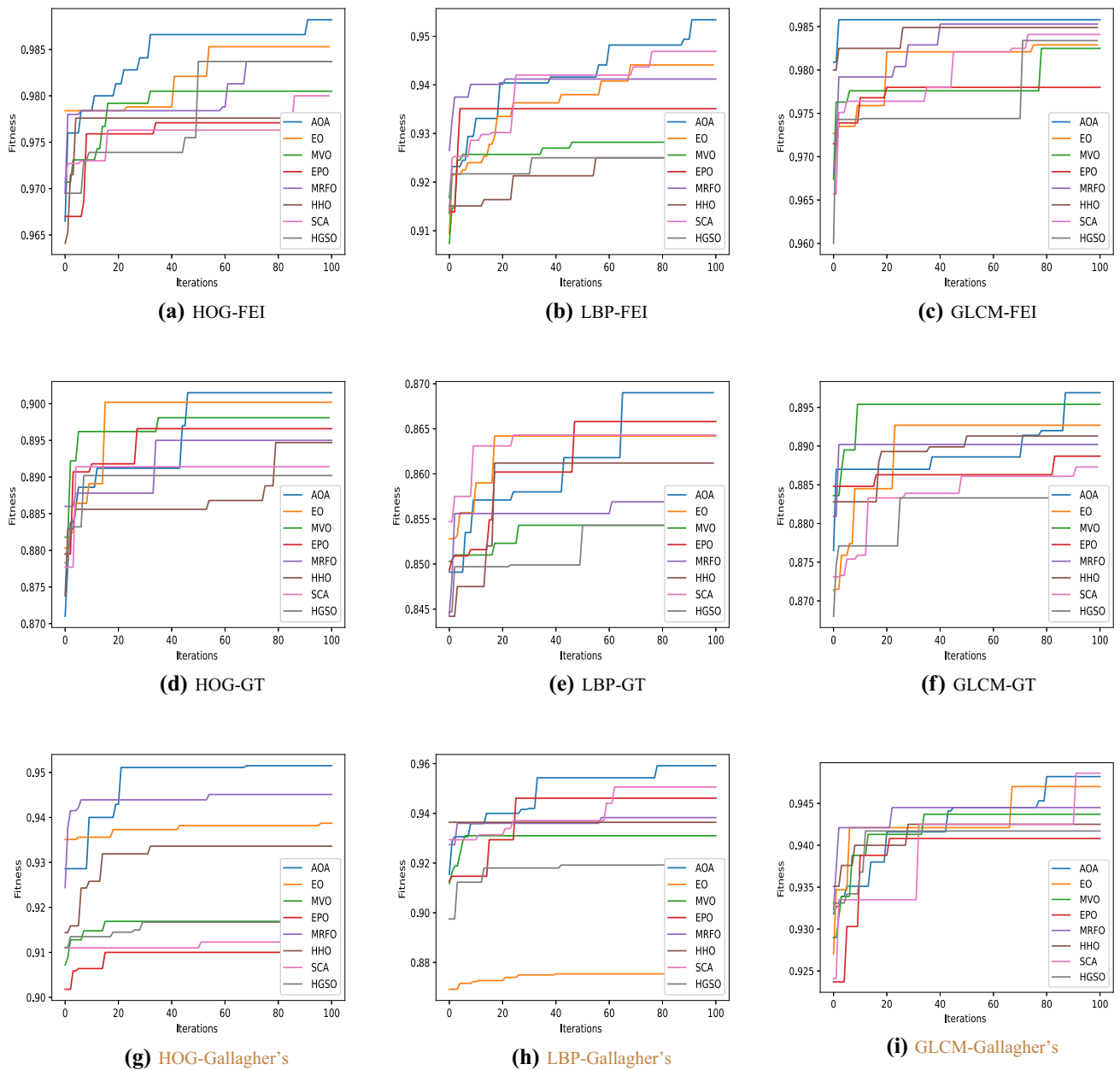
Fig. 7 Convergence curve of AOA versus other swarm intelligence algorithms over smallest and largest datasets

For GT dataset, in terms of accuracy MLP-LBP, MLP-HOG ,and MLP-GLCM achieve 82%, 85%, and 83%, respectively. Although, AOA-LBP, AOA-HOG, and AOA-GLCM attain 87%, 90% and 90%, respectively.

For the largest dataset (Gallagher's dataset), the margin of accuracy between AOA and MLP obtains 8%, 7%, and 5% using LBP, HOG and GLCM, respectively.

In terms of Fscore, the AOA-HOG and AOA-GLCM outperform the conventional MLP-HOG and MLP-GLCM for the smallest datasets (FEI and GT datasets). In addition, the values of Fscore obtained by AOA using the three descriptors

are very close to 96% for Gallagher's dataset, compared to MLP that achieves approximately 90%.

6.6 Results analysis and discussion

The main objective of this study is to design a new model within the framework of gender recognition based on a meta-heuristic inspired by Archimedes' law of physics named AOA. This study requires three descriptors like HOG, LBP, and GLCM, along with an MLP classifier. The experimental results showed the efficiency of the AOA algorithm compared to other competitors in the field of GR-based FS for
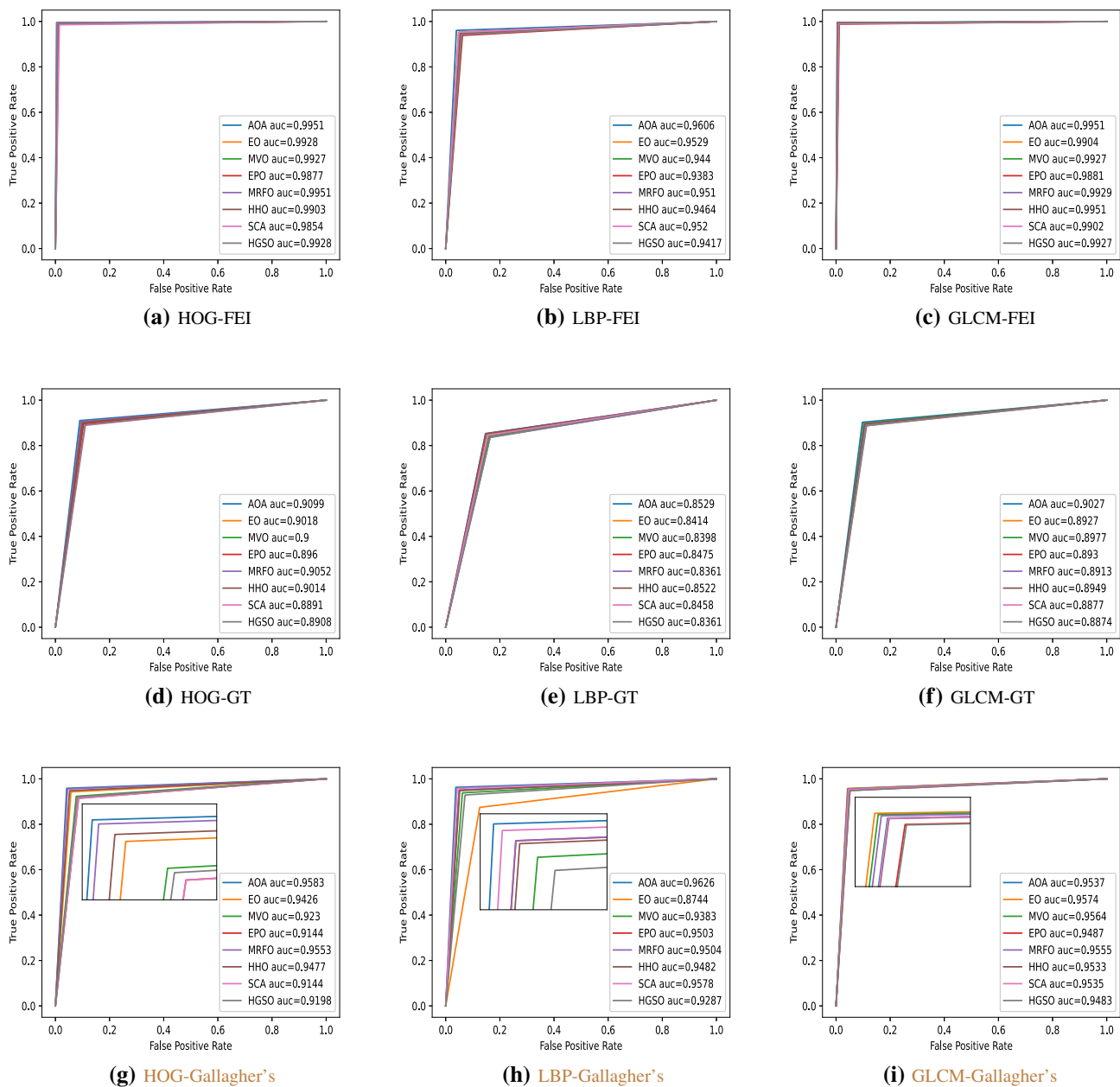
Fig. 8 ROC of AOA versus other swarm intelligence algorithms over smallest and largest datasets

the smallest (FEI and GT) and largest dataset (Gallagher's database). In addition, the way of integrating AOA to select the best patches from a face makes a real contribution.

The existing AOA has several benefits.

- For the smallest datasets (GT and FEI datasets), it can be seen that AOA based on three descriptors, HOG, LBP, and GLCM, provides better performance in terms of accuracy (see Table 5).
- Also, it is important to highlight that AOA-HOG achieved high accuracies for GT and FEI datasets.

- Concerning the optimal number of selected patches for the smallest datasets, AOA and SCA detect a lower number of selected patches (see Table 6).
- For the smallest dataset, EPO-HOG is the fastest algorithm for the GT dataset, and SCA-HOG is faster for the FET dataset. The main reason lies in using simple operators to explore the search space, i.e., EPO determines the temperature profile in the neighborhood of the huddle and locates the mover. It used only trigonometric operators based on sine and cosine functions to assess SCA (see Table 4).

Fig. 9 Visual examples of selected patches from Gallagher's dataset using AOA

(a) Original face from Gallagher's dataset

(b) Original face from Gallagher's dataset

(c) Selection of 20 patches using AOA-HOG

(d) Selection of 20 patches using AOA-HOG

(e) Selection of 10 patches using AOA-LBP

(f) Selection of 10 patches using AOA-LBP

(g) Selection of 13 patches using AOA-GLCM

(h) Selection of 13 patches using AOA-GLCM

- From the obtained results of the smallest datasets, we notice that the gender performance of the GT dataset causes less quality compared to FEI, and this is due to the complexity of the datasets, which depend on several challenging factors such as highlighting variation, facial expressions, different poses, and occluded eye area.
- For the largest dataset (Gallagher's dataset), AOA-LBP and AOA-HOG outperform AOA-GLCM in terms of accuracy. See Tables 10, 11, and 12.
- Also, AOA-LBP represents the best algorithm, which identifies the gender of Gallagher's dataset with 96.08%. Furthermore, this algorithm surpasses deep CNN and machine learning methods currently in the state of the art. See Table 16.
- For the largest dataset (Gallagher's dataset), AOA-LBP determines only 10 significant patches to distinguish between males and females. Furthermore, AOA-HOG identifies important facial features such as the eyes, nose,

and mouth, whereas LBP and GLCM detect sparse areas and concentrate on the edges. See Fig. 9.
- In terms of CPU time, AOA consumes less time for Gallagher's dataset than other optimizers. See Tables 10, 11 and 12.

According to the previous points, AOA ensures a good compromise between accuracy and the selection ratio, i.e., high accuracy and a low number of selected patches are obtained by AOA-HOG for the smallest datasets and AOA-LBP for the largest dataset (Gallagher's database). This phenomenon can be interpreted by the efficacy of physical rules inspired by Archimedes' theory, which ensures a good balance between exploration and exploitation mode. Also, AOA used extra information in the search space defined by volume and density, which enhanced the speed of convergence and ROC. See Figs. 7 and 8i.
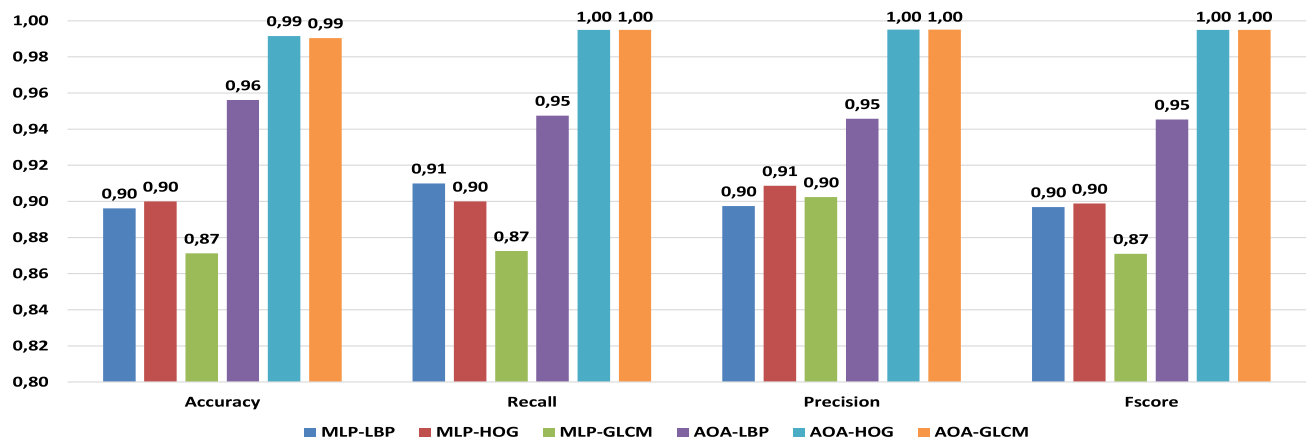
(a) FEI dataset
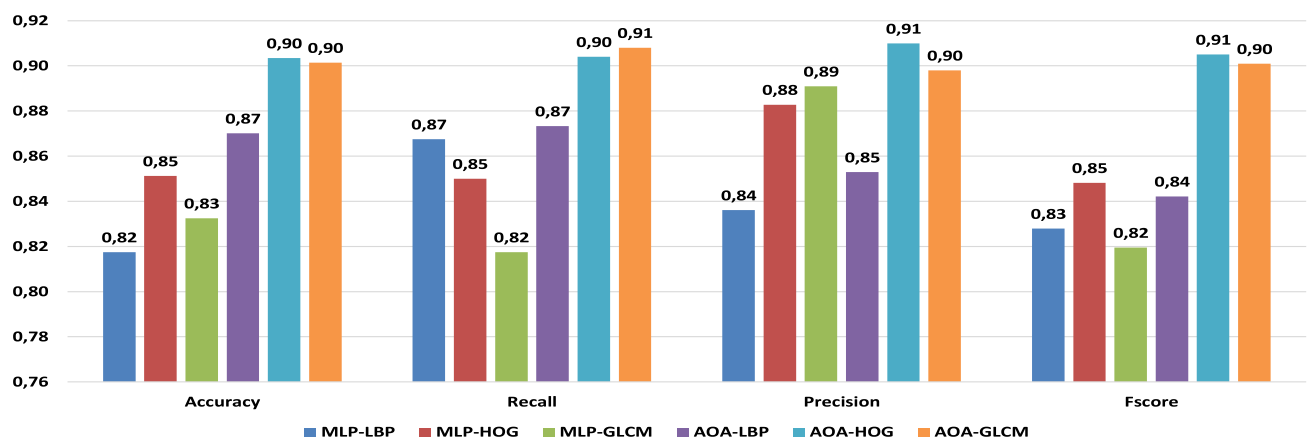
(b) GT dataset
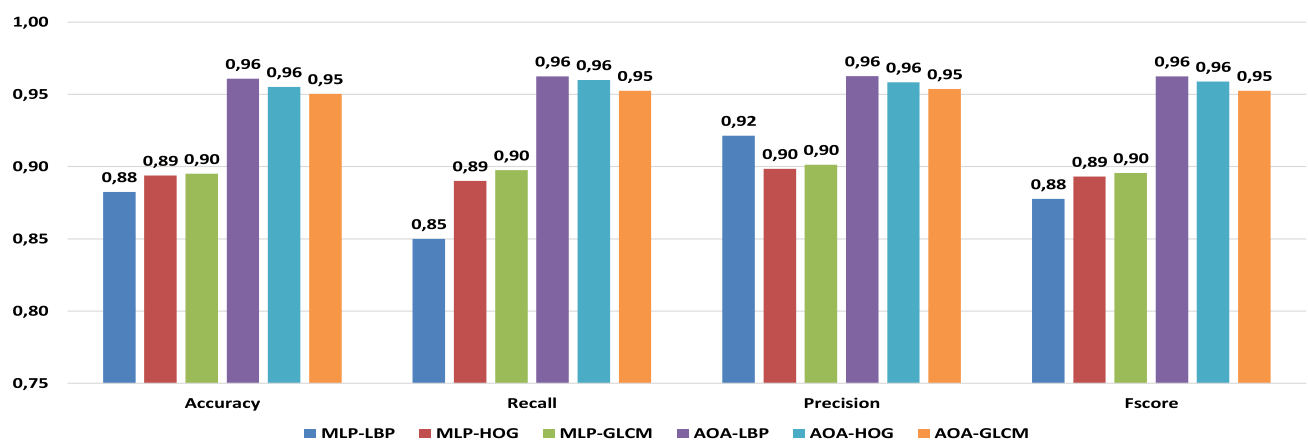
(c) Gallagher's dataset

Fig. 10 The performance of AOA against MLP

Besides the positives, the proposed AOA has certain drawbacks, which are mentioned below:

- This study employed MLP as a learning algorithm in a wrapper feature selection, but AOA aims to determine the significant patches of the face without optimizing the weights of MLP.
- The double use of random initialization in AOA and MLP allows the selection of different patches from one run to another.
- For handcrafted methods, their own parameters and the number of blocks can negatively influence the performance in terms of accuracy.

7 Conclusion

This paper presents a novel wrapper FS-based AOA for identifying the facial gender classes. For this, we have integrated three textural descriptors, including HOG, LBP, and GLCM tested over smallest datasets (GT and FEI) and largest dataset (Gallagher's database).

The obtained results by AOA-based HOG produced precise results in terms of accuracy and F-score for smallest datasets. Also, SCA allows keeping the smallest number of relevant blocks quickly. For largest dataset, AOA based on LBP achieved accurate results in terms of accuracy and the optimal number of selected patches in less time. As the advantage of the proposed method, the AOA ensures a good balance between the most relevant gender features from faces and the correct rate of gender classification. However, some drawbacks can be highlighted of AOA and the handcrafted methods. They are mainly several parameters defined randomly in the initialization steps of AOA, handcrafted methods, and the weights of MLP at initialization which is optimized by backpropagation algorithm. Also, the number of blocks for each handcrafted descriptor is fixed to 7×7 which has a higher impact on the performance. Furthermore, GLCM required us to define two parameters: the displacement (d) and orientation (θ). All parameters can be tuned automatically as hyper-heuristic AOA in the future.

The simultaneous optimization of weights of MLP and the automatic selection of patches can be explored in the future.

The new horizon can be explored, like the automatic fusion between textural descriptors assisted by FS wrapper using an improved version of AOA or another recent swarm algorithm which required fewer parameters with fast convergence. Also, combining handcrafted methods with deep features is conceivable and promising in automatic facial analysis, particularly gender recognition and age prediction or recognition of emotions.

Declarations

References

Abdalrady NA, Aly S (2020) Fusion of multiple simple convolutional neural networks for gender classification. In: 2020 International conference on innovative trends in communication and computer engineering (ITCE), IEEE, pp. 251–256

Abirami B, Subashini T, Mahavaishnavi V (2020) Gender and age prediction from real time facial images using cnn. Mater Today Proc 33:4708–4712

Acien A, Morales A, Vera-Rodriguez R, Bartolome I, Fierrez J (2018) measuring the gender and ethnicity bias in deep models for face recognition. in: iberoamerican congress on Pattern Recognition, Springer, pp. 584–593

Afifi M (2019) 11k hands: Gender recognition and biometric identification using a large dataset of hand images. Multimed Tools Appl 78:20835–20854

Agrawal B, Dixit M (2019) Age estimation and gender prediction using convolutional neural network. In: International conference on sustainable and innovative solutions for current challenges in engineering & technology, Springer, pp. 163–175

Alhichri H, Alswayed AS, Bazi Y, Ammour N, Alajlan NA (2021) Classification of remote sensing images using efficientnet-b3 cnn model with attention. IEEE Access 9:14078–14094

Al-Tashi Q, Rais HM, Abdulkadir SJ, Mirjalili S, Alhussian H (2020) A review of grey wolf optimizer-based feature selection methods for classification. Evolut Mach Learn Techniq 273–286

Althnian A, Aloboud N, Alkharashi N, Alduwaish F, Alrshoud M, Kurdi H (2021) Face gender recognition in the wild: an extensive performance comparison of deep-learned, hand-crafted, and fused features with deep and traditional models. Appl Sci 11:89

Anand S (2021) Archimedes optimization algorithm: Heart disease prediction: archimedes optimization algorithm: heart disease prediction. Multimed Res 4

Annrose J, Rufus N, Rex C, Immanuel DG (2021) A cloud-based platform for soybean plant disease classification using archimedes optimization based hybrid deep learning model. Wirel Pers Commun 1–23

Aspandi D, Martinez O, Sukno F, Binefa X (2021) Composite recurrent network with internal denoising for facial alignment in still and video images in the wild. Image Vis Comput 111:104189

Aspandi D, Mallol-Ragolta A, Schuller B, Binefa X (2020) Latent-based adversarial neural networks for facial affect estimations. In: 2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020), IEEE, 2020, pp. 606–610

Bekhouche SE, Ouafi A, Dornaika F, Taleb-Ahmed A, Hadid A (2017) Pyramid multi-level features for facial demographic estimation. Expert Syst Appl 80:297–310

Boon Ng C, Haur Tay Y, Goi BM (2012) Vision-based human gender recognition: a survey, arXiv e-prints: arXiv–1204

Castrillón-Santana M, Lorenzo-Navarro J, Ramón-Balmaseda E (2016) On using periocular biometric for gender classification in the wild. Pattern Recogn Lett 82:181–189

Castrillón-Santana M, Lorenzo-Navarro J, Ramón-Balmaseda E (2013) Improving gender classification accuracy in the wild. In: Iberoamerican congress on pattern recognition, Springer, pp. 270–277

Chen W-S, Jeng R-H (2020) A new patch-based lbp with adaptive weights for gender classification of human face. J Chin Inst Eng 43:451–457

Comas J, Aspandi D, Binefa X (2020) End-to-end facial and physiological model for affective computing and applications. In: 2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020), IEEE, pp. 93–100

Dago-Casas P, González-Jiménez D, Yu LL, Alba-Castro JL (2011) Single-and cross-database benchmarks for gender classification under unconstrained settings. In: 2011 IEEE international conference on computer vision workshops (ICCV Workshops), IEEE, pp. 2152–2159

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: (2005) IEEE computer society conference on computer vision and pattern recognition (CVPR'05), volume 1. IEEE 886–893

Dhiman G, Kumar V (2018) Emperor penguin optimizer: a bio-inspired algorithm for engineering problems. Knowl-Based Syst 159:20–50

Dhiman G, Oliva D, Kaur A, Singh KK, Vimal S, Sharma A, Cengiz K (2021) Bepo: a novel binary emperor penguin optimizer for automatic feature selection. Knowl-Based Syst 211:106560

Duan M, Li K, Yang C, Li K (2018) A hybrid deep learning cnn-elm for age and gender classification. Neurocomputing 275:448–461

Dwivedi N, Singh DK (2019) Review of deep learning techniques for gender classification in images. In: Harmony search and nature inspired optimization algorithms, Springer pp. 1089–1099

Faramarzi A, Heidarinejad M, Stephens B, Mirjalili S (2020) Equilibrium optimizer: a novel optimization algorithm. Knowl-Based Syst 191:105190

Fathy A, Alharbi AG, Alshammari S, Hasanien HM (2021) Archimedes optimization algorithm based maximum power point tracker for wind energy generation system. Ain Shams Eng J

Fitousi D, Rotschild N, Pnini C, Azizi O (2021) Understanding the impact of face masks on the processing of facial identity, emotion, age, and gender. Front Psychol 4668

Gallagher A, Chen T (2009) Understanding groups of images of people. In: IEEE conference on computer vision and pattern recognition, pp. 256–263

Gao Y, Zhou Y, Luo Q (2020) An efficient binary equilibrium optimizer algorithm for feature selection. IEEE Access 8:140936–140963

Gary BH, Manu R, Tamara B, Erik L et al (2007) Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In: Technical report 07-49, University of Massachusetts 1

Geetha A, Sundaram M, Vijayakumari B (2019) Gender classification from face images by mixing the classifier outcome of prime, distinct descriptors. Soft Comput 23:2525–2535

Ghazouani H (2021) A genetic programming-based feature selection and fusion for facial expression recognition. Appl Soft Comput 103:107173

Ghojogh B, Shouraki SB, Mohammadzade H, Iranmehr E (2018) A fusion-based gender recognition method using facial images. In: Electrical engineering (ICEE), Iranian conference on, IEEE, pp. 1493–1498

Ghosh KK, Guha R, Bera SK, Kumar N, Sarkar R (2021) S-shaped versus v-shaped transfer functions for binary manta ray foraging optimization in feature selection problem. Neural Comput Appl 1–15

Goel A, Vishwakarma VP (2016) Gender classification using kpca and svm. In: 2016 IEEE international conference on recent trends in electronics, information communication technology (RTEICT), 2016, pp. 291–295. 10.1109/RTEICT.2016.7807829

Goel A, Vishwakarma VP (2016a) Efficient feature extraction using dct for gender classification. In: 2016 IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), IEEE pp 1925–1928

Goel A, Vishwakarma VP (2016b) Feature extraction technique using hybridization of dwt and dct for gender classification. In: 2016 ninth international conference on contemporary computing (IC3), IEEE, . 1–6

Greco A, Saggese A, Vento M, Vigilante V (2020) A convolutional neural network for gender recognition optimizing the accuracy/speed tradeoff. IEEE Access 8:130771–130781

Greco A, Saggese A, Vento M, Vigilante V (2021) Gender recognition in the wild: a robustness evaluation over corrupted images. J Ambient Intel Human Comput 12:10461–10472

Greco A, Saggesea A, Vento M, Vigilante V (2020) Gender recognition in the wild: a robustness evaluation over corrupted images. J Intell Human Comput 1–12

Grother P, Grother P, Ngan M, Hanaoka K (2019) Face recognition vendor test (FRVT) part 2: identification. US Department of Commerce, National Institute of Standards and Technology

Haider KZ, Malik KR, Khalid S, Nawaz T, Jabbar S (2019) Deepgender: real-time gender classification using deep learning for smartphones. J Real-Time Image Proc 16:15–29

Hashim FA, Hussain K, Houssein EH, Mabrouk MS, Al-Atabany W (2020) Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems. Appl Intell 1–21

Hashim FA, Houssein EH, Mabrouk MS, Al-Atabany W, Mirjalili S (2019) Henry gas solubility optimization: a novel physics-based algorithm. Futur Gener Comput Syst 101:646–667

Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: algorithm and applications. Futur Gener Comput Syst 97:849–872

Hsu C-Y, Lin L-E, Lin CH (2021) Age and gender recognition with random occluded data augmentation on facial images. Multimed Tools Appl 80:11631–11653

Hung BT (2021) Face recognition using hybrid hog-cnn approach. In: Research in intelligent and computing in engineering, Springer, pp. 715–723

Huynh HT, Nguyen H (2020) Joint age estimation and gender classification of asian faces using wide resnet. SN Comput Sci 1:1–9

Ito K, Kawai H, Okano T, Aoki T (2018) Age and gender prediction from face images using convolutional neural network. In: (2018) Asia-Pacific signal and information processing association annual summit and conference (APSIPA ASC). IEEE 7–11

Jalali S, Boostani R, Mohammadi M (2021) Efficient fingerprint features for gender recognition. Multidim Syst Signal Process 1–17

Khalifa T, Şengül G (2018) Gender prediction from facial images using local binary patterns and histograms of oriented gradients

transformations. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi 7:14–22

Khan K, Attique M, Syed I, Gul A (2019) Automatic gender classification through face segmentation. Symmetry 11:770

Khan G, Samyan S, Khan MUG, Shahid M, Wahla SQ (2020) A survey on analysis of human faces and facial expressions datasets. Int J Mach Learn Cybern 11:553–571

Kumar S, Singh S, Kumar J (2019) Gender classification using machine learning with multi-feature method. In: (2019) IEEE 9th annual computing and communication workshop and conference (CCWC). IEEE 0648–0653

Lakshmi D, Ponnusamy R (2021) Facial emotion recognition using modified hog and lbp features with deep stacked autoencoders. Microprocess Microsyst 82:103834

Lapuschkin S, Binder A, Muller K-R, Samek W (2017) Understanding and comparing deep neural networks for age and gender classification. In: Proceedings of the IEEE international conference on computer vision workshops, pp. 1629–1638

Lee M, Lee J-H, Kim D-H (2022) Gender recognition using optimal gait feature based on recursive feature elimination in normal walking. Expert Syst Appl 189:116040

Lin C-J, Li Y-C, Lin H-Y (2020) Using convolutional neural networks based on a taguchi method for face gender recognition. Electronics 9:1227

Livieris IE, Pintelas E, Pintelas P (2019) Gender recognition by voice using an improved self-labeled algorithm. Mach Learn Knowl Extract 1:492–503

Mafarja M, Mirjalili S (2018) Whale optimization approaches for wrapper feature selection. Appl Soft Comput 62:441–453

Mane S, Shah G (2019) Facial recognition, expression recognition, and gender identification. In: Data management, analytics and innovation, Springer, pp. 275–290

Mansanet J, Albiol A, Paredes R (2016) Local deep neural networks for gender recognition. Pattern Recogn Lett 70:80–86

Micheal AA, Geetha P (2019) Combined feature extraction for multiview gender recognition. In: Smart Intelligent Computing and Applications, Springer, pp. 219–228

Mirjalili S (2016) Sca: a sine cosine algorithm for solving optimization problems. Knowl-Based Syst 96:120–133

Mirjalili S, Mirjalili SM, Hatamlou A (2016) Multi-verse optimizer: a nature-inspired algorithm for global optimization. Neural Comput Appl 27:495–513

Mirza AM, Hussain M, Almuzaini H, Muhammad G, Aboalsamh H, Bebis G (2013) Gender recognition using fusion of local and global facial features. In: International symposium on visual computing, Springer, pp 493–502

Nefian AV (2013) Georgia tech face database

Neggaz N, Houssein EH, Hussain K (2020) An efficient henry gas solubility optimization for feature selection. Expert Syst Appl 152:113364

Ng C-B, Tay Y-H, Goi B-M (2015) A review of facial gender recognition. Pattern Anal Appl 18:739–755

Nguyen H-T, Huong TTN (2017) Gender classification by lpq features from intensity and monogenic images. In: 2017 4th NAFOSTED conference on information and computer science, IEEE, pp. 96–100

Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24:971–987

Omer HK, Jalab HA, Hasan AM, Tawfiq NE (2019) Combination of local binary pattern and face geometric features for gender classification from face images. In: 2019 9th IEEE international conference on control system, computing and engineering (ICCSCE), IEEE, pp. 158–161

Orozco CI, Iglesias F, Buemi ME, Berlles JJ (2017) Real-time gender recognition from face images using deep convolutional neural network

Osman SM, Viriri S (2020) Dynamic local ternary patterns for gender identification using facial components. International Conference on Computer Vision and Graphics, Springer pp 133–141

Pai S, Shettigar R (2021) Gender recognition from face images using sift descriptors and trainable features. In: Advances in artificial intelligence and data engineering, Springer, pp. 1173–1186

Patil KA et al (2021) Features and methods of human age estimation: opportunities and challenges in medical image processing. Turkish J Comput Math Educ (TURCOMAT) 12:294–318

Pattnaik G, Parvathi K (2021) Automatic detection and classification of tomato pests using support vector machine based on hog and lbp feature extraction technique. In: Progress in advanced computing and intelligent engineering, Springer, pp. 49–55

Peimankar A, Puthusserypady S (2021) Dens-ecg: a deep learning approach for ecg signal delineation. Expert Syst Appl 165:113911

Peker M (2021) Classification of hyperspectral imagery using a fully complex-valued wavelet neural network with deep convolutional features. Expert Syst Appl 173:114708

Preeti RK (2021) Performance estimation of wireless sensor network using archimedes optimization algorithm. Des Eng 728–746

Sajja TK, Kalluri HK (2019) Gender classification based on face images of local binary pattern using support vector machine and back propagation neural networks. Adv Model Anal B 62:31–35

Savchenko AV (2019) Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output convnet. Peer J Comput Sci 5:e197

Silva DPd (2019) Age and gender classification: a proposed system, Ph.D. thesis

Simanjuntak F, Azzopardi G (2020) Fusion of cnn- and cosfire-based features with application to gender recognition from face images. In: Arai K, Kapoor S (eds) Advances in Computer Vision. Springer International Publishing, Cham, pp 444–458

Singh A, Rai N, Sharma P, Nagrath P, Jain R (2021) Age, gender prediction and emotion recognition using convolutional neural network, Available at SSRN 3833759

Sun X, Wang G, Xu L, Yuan H, Yousefi N (2021) Optimal estimation of the pem fuel cells applying deep belief network optimized by improved archimedes optimization algorithm. Energy 237:121532

Surinta O, Khamket T, Gender recognition from facial images using local gradient feature descriptors. In: (2019) 14th international joint symposium on artificial intelligence and natural language processing (iSAI-NLP). IEEE 1–6

Taghian S, Nadimi-Shahraki MH (2019) Binary sine cosine algorithms for feature selection from medical data, arXiv preprint arXiv:1911.07805

Thaher T, Heidari AA, Mafarja M, Dong JS, Mirjalili S (2020) Binary harris hawks optimizer for high-dimensional, low sample size feature selection. In: Evolutionary machine learning techniques, Springer, pp. 251–272

Thomaz CE, Giraldi GA (2010) A new ranking method for principal components analysis and its application to face image analysis. Image Vis Comput 28:902–913

Uçar MK, Uçar Z, Uçar K, Akman M, Bozkurt MR (2021) Determination of body fat percentage by electrocardiography signal with gender based artificial intelligence. Biomed Signal Process Control 68:102650

Vimal S, Robinson YH, Kaliappan M, Vijayalakshmi K, Seo S (2021) A method of progression detection for glaucoma using k-means and the glcm algorithm toward smart medical prediction. The J Supercomput 1–17

Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evol Comput 1:67–82

Yao B, Hayati H (2021) Model parameters estimation of a proton exchange membrane fuel cell using improved version of archimedes optimization algorithm. Energy Rep 7:5700–5709

Yıldız BS, Pholdee N, Bureerat S, Erdaş MU, Yıldız AR, Sait SM (2021) Comparision of the political optimization algorithm, the archimedes optimization algorithm and the levy flight algorithm for design optimization in industry. Mater Test 63:356–359

Yu H, Yang LT, Zhang Q, Armstrong D, Deen MJ (2021) Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. Neurocomputing

Zhang C, Ding H, Shang Y, Shao Z, Fu X (2018) Gender classification based on multiscale facial fusion feature. Math Probl Eng 2018

Zhao W, Zhang Z, Wang L (2020) Manta ray foraging optimization: an effective bio-inspired optimizer for engineering applications. Eng Appl Artif Intell 87:103300

Zhou Y, Li Z (2019) Facial eigen-feature based gender recognition with an improved genetic algorithm. J Intell Fuzzy Syst 37:4891–4902