



# Boosting Archimedes optimization algorithm using trigonometric operators based on feature selection for facial analysis

Imène Neggaz<sup>1</sup> · Nabil Neggaz<sup>1</sup> · Hadria Fizazi<sup>1</sup>

Received: 13 November 2021 / Accepted: 3 October 2022

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

## Abstract

Due to technical advancements and the proliferation of mobile applications, facial analysis (FA) of humans has recently become an important area for computer vision research. FA investigates a variety of difficulties, including gender recognition, facial expression recognition, age and race recognition, with the goal of automatically comprehending social interactions. Due to the dimensional challenge posed by pre-trained CNN networks, the scientific community has developed numerous techniques inspired by biology, swarm intelligence theory, physics, and mathematical rules. This article presents a gender recognition system based on scAOA, that is a modified version of the Archimedes optimization algorithm (AOA). The latest variant (scAOA) enhances the exploitation stage by using trigonometric operators inspired by the sine cosine algorithm (SCA) in order to prevent local optima and to accelerate the convergence. The main purpose of this paper is to apply scAOA to select the relevant deep features provided by two pretrained models of CNN (AlexNet & ResNet) to recognize the gender of a human person categorized into two classes (men and women). Two datasets are used to evaluate the proposed approach (scAOA): the Brazilian FEI dataset and the Georgia Tech Face dataset (GT). In terms of accuracy, Fscore and statistical test, the comparison analysis demonstrates that scAOA outperforms other modern and competitive optimizers such as AOA, SCA, Ant lion optimizer (ALO), Salp swarm algorithm (SSA), Grey wolf optimizer (GWO), Simple genetic algorithm (SGA), Grasshopper optimization algorithm (GOA) and Particle swarm optimizer (PSO).

**Keywords** Facial analysis · Gender recognition · Pretrained CNN · Sine cosine archimedes optimization algorithm (scAOA) · Trigonometric operators · Wrapper feature selection (FS)

## 1 Introduction

Human vision enables the performance of multiple tasks concurrently and rapidly, most notably facial detection, gender recognition, and distinguishing the state of mind that distinguishes a human being from others.

Gender recognition automation is a key difficulty for scientists, and it has a considerable impact on the

commercial sector and video surveillance. For example, shopping centers want to know the sales rate and the demographics of the individuals who purchase their items, notably their gender, age, and origins, in order to enhance sales. Additionally, another area necessitates the use of gender identification to identify suspicious individuals filmed by security cameras in broad locations such as airports, retail malls, and petrol stations. To help reduce the time spent searching for the suspected perpetrator, the gender identification application can make a significant contribution to resolving this issue, particularly in crucial scenarios such as suicide bombings or airport attacks. Additionally, the current scenario with the Covid-19 outbreak has compelled people to conceal the fact that gender identification automation plays a critical role in our lives.

According to the research, human detection rate is less than 95%, which complicates automatic gender recognition. Numerous problems are explored in this subject,

---

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

Nabil Neggaz  
nabil.neggaz@univ-usto.dz

Hadria Fizazi  
fizazi@univ-usto.dz

<sup>1</sup> Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf (USTO-MB), BP 1505, EL M'naouer, 3100 Oran, Algeria

including rotated, occluded faces and the human who resembles women with long hair, which has a significant impact on gender recognition performance.

Gender recognition, in general, necessitates two critical phases. The first stage extracts features from faces, while the second step performs classification using machine learning techniques.

In general, the extraction of characteristics falls into three broad categories: deep features, handcrafted techniques and combined methods.

Nowadays, deep features based on pre-trained CNNs are widely used in machine learning, particularly in computer vision [1], biomedical application [2, 3] and remote sensing [4, 5].

The great number of attributes generated by both methods (Handcrafted and deep features) prompts researchers to develop new selection methods called wrapper feature selection based on meta-heuristics (MHs). MHs are derived from different subjects, allowing the development of a large number of optimization algorithms that can be merged with machine learning techniques. In the early 1990 s, Holland developed Genetic algorithms [14] (GA), Differential evolution (DE) is introduced by Storn and Price [15], more recently, Evolutionary strategies (ES) is designed by Rudolph [16], and Genetic programming (GP) proposed by Koza [18], that used a tree representation. Then, another axis interests scientists: replicating the behavior of swarming through the introduction of the hypothesis of swarm intelligence (SI). This category has expanded significantly in recent years, with the development of the Practical Swarm Optimizer (PSO) by Eberhart and Kennedy [19]. Recently, the pioneer Mirjalili developed several algorithms including the Gray Wolf Optimizer (GWO) [20], Ant Lion Optimizer (ALO) [21], the Whale Optimization Algorithm (WOA) [22] and Salp swarm algorithm (SSA) [23]. Furthermore, Saremi et al. [24] realized a novel optimizer called Grasshopper Optimization Algorithm (GOA). More recently, Heidari et al. [25] simulated the behavior of hawks, which allows to create a new algorithm called the Harris hawks optimization (HHO). Zhao et al. [28] simulated the foraging of manta-raies, which created Manta-ray Foraging Optimizer (MRFO).

Last year, Remora optimization algorithm (ROA) has been introduced by Jia et al. [29]. In 2022, Honey badger algorithm (HBA) has been designed by Hashim et al. [27].

Additionally, we observe that physics and mathematics have inspired the development of new algorithms such as Archimedes optimization algorithm (AOA) and Henry Gas Solubility Optimization (HGSO), created by Hashim et al. [30, 33]. Also, we find other algorithms such as Equilibrium optimizer [32], Thermal exchange optimizer (TEO) [34], Arithmetic optimization algorithm, Runge kutta

optimizer (RUN) [35], and Sine cosine algorithm (SCA) [36].

Despite the significant development of metaheuristics generated by the imitation of animal behavior, mathematical and physical laws in the field of feature selection as MRFO introduced by Ghosh et al. [37], ALO developed by Emary et al. [38], HHO designed by Thaheret al. [39], GWO proposed by Al-Tashi et al. [40], WOA designed by Mafarja and Mirjalili. [41], HGSO introduced by Neggaz et al. [42], EO created by Gao et al. [43], GOA generated by Zakeri and Hokmabadi [44] and SCA produced by Taghian & Nadimi-Shahrak [45].

Recently, AOA has been published by Hashim et al. [30], which simulates the Archimedes physical rule in the field of global optimization. This algorithm is widely used in several fields like design of vehicle structure [46] and optimization of deep belief network [47]. In addition, an enhanced variant of AOA has been processed by Desuky et al. [48] in the field of feature selection. Also, SCA is employed for enhancing several swarm algorithms-like SSA and HHO in the field of FS [49, 50]. Also, Ewees et al. [51] investigated the use of SCA with artificial bee colony for image segmentation.

Due to the “No Free Lunch” theory, which states that no algorithm can solve all optimization problems, this work offers a novel variant of AOA dubbed scAOA for recognizing gender from faces. Thus, this approach seeks to improve upon the original AOA algorithm by employing sinusoidal operators in order to address some of its flaws, including delayed convergence, being trapped at local optima, and having an imbalanced algorithm. The major contributions of this paper are as follows:

- Designing a novel wrapper physical algorithm scAOA for predicting gender identification using an automatic selection of the optimal trainable deep features of face images.
- Comparing the performance of scAOA with several recent and robust optimizers as for facial analysis based on FS.
- Evaluating the impact of deep features based on AlexNet and ResNet.
- Testing the efficiency of scAOA for gender recognition (GR) over two datasets: FEI and GT.

This paper is structured around 5 sections. Section 2 explains some related work for gender recognition (GR) from faces. Section 3 describes trainable feature extractors based on AlexNet with 4096 features and ResNet with 2048 features. Section 4 gives details about the Archimedes optimizer algorithm and the concept of trigonometric operator (TO). In Sect. 5, we propose our system of GR using a modified version of AOA by employing TO as a mathematical operator to provide a new wrapper

algorithm for feature selection called scAOA. Section 6 represents the kernel of our paper, which includes datasets description, parameters of algorithms, and quantitative and graphical results proved by statistical analysis using Rank Sum Wilcoxon's. The last Sect. 7 shows our conclusion with some future horizons.

## 2 Related work

Facial analysis includes several area of research-like age and gender recognition, emotion and ethnicity detection.

Recently, deep CNN play an important role in computer vision area due to their double use, i.e., the deep CNN can provide the vector of characteristics (features) and classification task.

Recently, several architectures for gender recognition have been constructed utilizing pre-trained CNNs such as AlexNet, VGG16, ResNet, Inception, GoogleNet, and CaffeNet ([6–12]). Additionally, a gender recognition comparison research is conducted by Greco et al. [13], which used different pretrained CNNs, such as MobileNet, DensNet, Xception, and SqueezeNet.

This section highlights some works related to facial analysis, especially gender recognition. First, we review the trainable features with machine learning techniques for gender recognition. Second, we show some studies about gender identification based on handcrafted methods. Third, hybrid features are discussed. Finally, an outline of gender recognition via wrapper feature selection is presented.

### 2.1 Trainable features (Deep features)

Duan et al. [52] presented a novel synergy between convolutional neural network (CNN) and extrem learning machine (ELM) for age and gender identification. The CNN is employed as a feature extractor, while the ELM is used as a classifier to simultaneously estimate a person's age and discriminate between male and female face photos. The authors obtained higher performance compared to other studies in terms of accuracy for two popular datasets: Adience and Morph-II.

Acien et al. [53] measured ethnicity and gender using two pretrained-CNN architectures. For LFW dataset, VGG16 shown a good ability to recognize easily gender and race with a high performance compared to ResNet50.

Ito et al. [54] developed four pre-trained CNN architectures: AlexNet, VGG-16, ResNet-152, and Wide-ResNet-16-8 for predicting age and gender that used IMDB-WIKI datasets. The experiment demonstrated that Wide-ResNet outperformed other pre-trained CNNs in terms of accuracy.

Mane et al. [55] investigated three aspects of human facial analysis: face recognition, gender identification, and expression recognition. The experimental results indicated that the gender recognition rate is very higher to emotion recognition rate. The CNN achieved 96% as accuracy for gender identification task.

Recently, Agrawal and Dixit [56] used CNN as an extractor to predict the age and gender of individuals based on facial pictures. They further applied PCA to reduce the dimensional, however, the classification step is performed using a Feed-Forward Neural Network.

In 2019, Haider et al. [57] created a novel deep CNN architecture for real-time gender identification on smartphones. The design is comprised of four convolutional layers, three max-pooling layers, two fully connected layers, and a single regression layer. The training is conducted using two datasets: FEI and CAS-PEAL, which contain 200 persons with 2800 faces and 1040 individuals with 30,871 faces, respectively. The Deep-gender algorithm obtained 98.75% as accuracy on the FEI dataset by applying a novel strategy that includes alignment prior to shrinking the facial image. However, when applied to CAS-PEAL-R1 datasets, the proposed approach attained an accuracy of 97.73%

### 2.2 Handcrafted features

Recently, different sources dubbed handcrafted techniques have been created in the literature for human facial analysis, including face recognition, emotion recognition, age and gender estimation. The retrieved features are computed for the entire face or selected parts using a histogram of oriented gradients (HOGs) and a scale-invariant feature transform (SIFT) [58].

Surinta and Khamket [58] utilized a support vector machine (SVM) as a classifier to classify Color FERET samples into two classes: female and male. The authors conclude that when the size of the training data is decreased, HOG outperforms the SIFT descriptor based on SVM in the experimental investigation.

A few studies based on uncontrolled image conducted on Image of Groups dataset. As instance, the work of Zhang et al. [59] has been illustrated. The authors established an innovative method for gender recognition by fusing facial features. Combining local binary pattern (LBP), local phase quantization (LPQ), and a multiblock produces the vector of characteristics. The classification task is realized by SVM and validated on Image of Group datasets (IoG). The experimental results demonstrate that the proposed strategy outperforms existing fundamental variants (LBP and LPQ).

Moreover, an interesting study was published recently, dedicated to gender recognition [60]. In this work, Ghoghogh

et al. [60] created four frameworks for identifying a person's gender using facial photos. The first framework entails extracting features via a texturing approach based on LBP and reducing the dimension of vector features via PCA, which will then be used as input for a multi-layer perceptron (MLP). The second framework utilized Gabor filters to generate a vector of PCA-reduced features that was used as input to the kernel SVM classifier. The third framework extracts the lower portion of the face with a resolution of  $(30 \times 30)$ , which is then molded into a column vector with a resolution of  $(900 \times 1)$  and used as input for the kernel SVM classifier. The final framework extracts 34 landmarks from the face and classifies them using a linear discriminant method (LDA). All proposed frameworks are evaluated using FEI data, and experimental results indicate that the third framework surpasses the others by 90% in terms of accuracy. However, when a choice is made using the weighting vote, the accuracy jumps to 94%. Additionally, texture and geometric cues are used to identify gender, which can be detected using local binary patterns (LBP) and gray level co-occurrence matrices (GLCM).

Recent advances in face and gender identification have resulted in the development of various upgraded variants of LBP known as local directional pattern (LDP), local phase quantization, and local phase quantization (LPQ). In the same regard, Chen and Jeng. [61] proposed a novel variation of LBP dubbed Adaptive patch-weight LBP (APWLBP). Their method computed the gradient using a pyramid structure and weight parameters established by Eigen theory. The primary goal of (APWLBP) is to discover the best hyperplane projection with a high variance for gender recognition. On the 3D Adience and LFW datasets, the performance of APWLBP-based SVM is highly comparable to that of CNN.

### 2.3 Hybrid methods

Numerous studies have been conducted on deep-learned features and their impact on gender recognition when compared to handcrafted features and fused features [63].

Dwivedi and Singh [64] summarized first, some methods based on handcrafted data such as LBP, HOG, SIFT, weighted HOG, and CROSSFIRE Filter; and second, they discussed the role of CNN, which can be used for dual tasks, i.e., as a feature extractor and a classifier to recognize age and gender from the face.

More recently, Althnian et al. [65] presented three handcrafted features, which can be extracted from facial image using LBP, HOG, and PCA. Furthermore, the authors employed deep CNN features, and fused features based on three combinations dubbed LBP-DL, HOG-DL, and PCA-DL. After that, two classifiers, SVM and CNN,

are used to accomplish the task of gender identification. The experimental results indicated that fused features (LBP-DL) and an SVM classifier achieved a high rate of average accuracy of 88.1% when tested on two datasets, LFW and Adience.

Alghaili et al. [66] designed a novel framework for gender recognition, which treats camouflaged faces. Their network combines inception with variational feature learning (VFL). The experiment is evaluated using five public datasets. Our interest is focused on FEI performance, which attains 99.51% in terms of accuracy.

### 2.4 Wrapper FS methods

A few work used bio-inspired algorithm for gender recognition based on feature selection. As example, the work of Zhou and Li [67] employed genetic algorithm (GA) to categorize faces depending on their gender automatically. GA is used to discover the optimal set of PCA-extracted and neural network-classified eigenfeatures from faces. Two datasets FEI and FERET, are used to validate the experimental investigation. The obtained results attained accuracy rates of 96% and 94%, respectively. More recently, Neggaz and Fizazi [68] employed Archimedes optimization algorithm (AOA) for facial analysis. The authors used three handcrafted methods based on HOG, LBP and GLCM. Their idea consists to determine automatically the relevant blocs from face to recognize gender. The experimental results shown that AOA-LBP achieved a high performance in terms of accuracy with 96% for Gallagher dataset.

## 3 Features extraction using pre-trained CNN

CNNs are primarily composed of three types of layers: convolutional, pooling, and fully connected. The most critical layers are the convolutional and pooling layers. By convolving an image area with numerous filters, a convolution layer is utilized to extract features. Due to the higher layer count, a CNN can more precisely interpret the features in its input image. The pooling layer compresses the output mapping of the convolution. This section explains two types of pre-trained CNN: AlexNet and ResNet. It is important to highlight that Alexnet is chosen as a core component of the proposed model due to its capacity to perform rapid network training and its ability to reduce over-fitting. This model is more appropriate due to its straightforward structure, short training time, and low memory usage [69]. Additionally, ResNet was ranked first in ImageNet detection and localization [70].



### 3.1 AlexNet

AlexNet is a well-known image categorization model. The AlexNet architecture, as represented in Fig. 1, consists of five convolutional layers (Conv), three pooling layers (Pool), and three fully connected layers. The parameters of each layer are listed in Table 1.

- Convolution layer (CL) The convolution layer performs the convolution process on the original image by utilizing a special feature detector called the convolution kernel. During training, the image is scaled to  $227 \times 227 \times 3$ . In the first convolution layer, 96 number of  $11 \times 11 \times 3$  convolution kernels are employed to generate new pixels via convolution. The second layer's input data is the  $27 \times 27 \times 96$  pixel layer output by the previous layer. To aid in subsequent processing, each pixel's upper, bottom, left, and right margins are filled with two pixels. The second layer's convolution kernel has a size of  $5 \times 5$  and a stride of 1 pixel, which is identical to the first layer's computation formula. The third layer's input data is a collection of  $13 \times 13 \times 128$  pixel layers output by the second layer. To facilitate subsequent processing, each pixel layer's upper, bottom, left, and right margins are filled with a single pixel. The fourth layer's input data is similar to the third layer's: sets of  $13 \times 13 \times 192$  pixel layers output by the third layer. The fifth layer's input data is a collection of  $13 \times 13 \times 192$  pixel layers output by the fourth layer. To facilitate subsequent processing, each pixel layer's

upper, bottom, left, and right margins are filled with a single pixel. The convolution layer's  $K_s$ ,  $S$ , and  $P$  parameters are used to determine the sizes of the input and output images. More precisely, the convolution layer's computation formula is defined by:

$$O = \frac{(I - K_s + 2P)}{S} + 1 \tag{1}$$

- Pooling layer (PL) This architecture makes extensive use of max pooling, a technique that selects the matrix's maximum value. When an image is translated by a few pixels, the maximum pooling has little effect on the judgment. Max pooling's primary advantage is its efficacy in removing noise. After the first, second, and fifth convolution layers, the max pooling layers are added.
- Fully connected layer (FCL) AlexNet is composed of three fully connected layers. The first and second completely connected layers each contain 4096 convolution kernels with a size of  $6 \times 6 \times 256$ . This is because the convolution kernel's size is identical to the size of the feature map being processed. Each coefficient in the convolution kernel is multiplied by the pixel value of the feature map's size, resulting in a one-to-one relationship. Thus, the pixel layer after convolution has a size of  $4096 \times 1 \times 1$ , indicating that there are 4096 neurons. Due to the fact that this study has two categories, two neurons in the third fully linked layer are set.

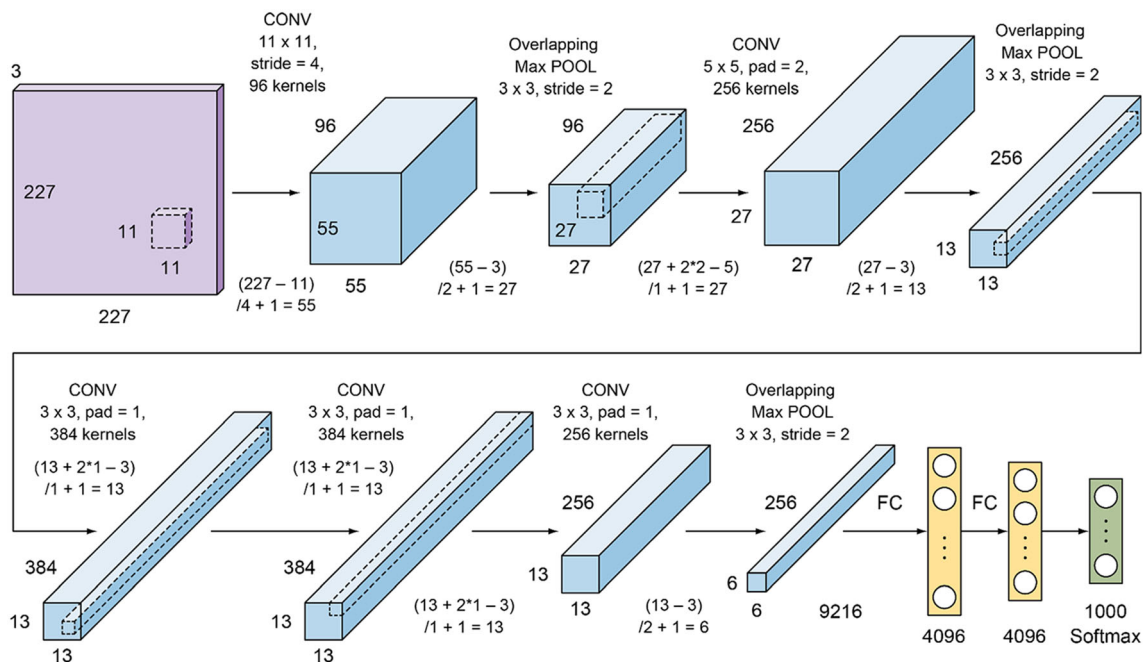


Fig. 1 The design of AlexNet. [72]

**Table 1** AlexNet layer parameters

	Conv1	Pool1	Conv2	Pool2	Conv3	Conv4	Conv5	Pool5
Kernel size ( $K_s$ )	$11 \times 11$	$3 \times 3$	$5 \times 5$	$3 \times 3$	$3 \times 3$	$3 \times 3$	$3 \times 3$	$3 \times 3$
Stride ( $S$ )	4	2	1	2	1	1	1	1
Padding ( $P$ )	0	0	2	0	1	1	1	0

### 3.2 ResNet

Figure 2 illustrates a ResNet residual block. As illustrated in the figure, stacked layers execute residual mapping in residual networks by establishing shortcut connections that perform identity mapping ( $x$ ). Their outputs were added to the residual function  $F$  of the stacked layers ( $x$ ).

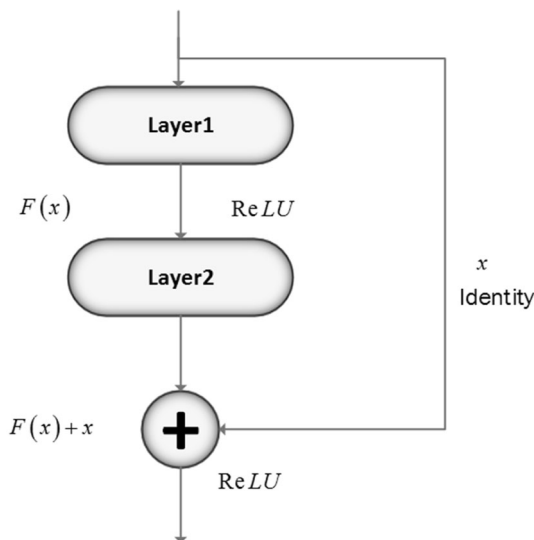
The gradient of error was determined and propagated to the shallow layers during the deep network’s backpropagation training. This inaccuracy becomes smaller and smaller as one progresses deeper into the levels, until it eventually vanishes. This phenomenon is referred to as the gradient vanishing problem in really deep networks. As illustrated in Figs. 2 and 3, the problem can be handled via residual learning [71].

The initial residual branch, or unit  $l$ , is depicted in Fig. 3 within the residual network. Weights, batch normalization ( $BN$ ), and corrected linear unit are depicted in the figure ( $ReLU$ ). The following equations were used to determine the input and output of a residual unit:

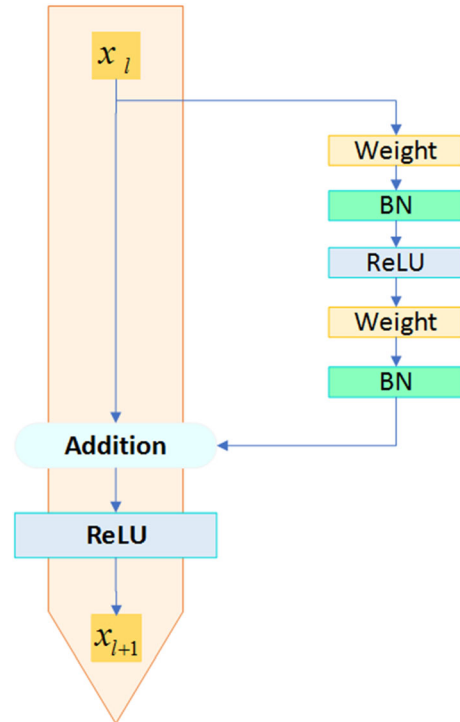
$$y_l = h(x_l) + F(x_l + W_l) \tag{2}$$

$$x_{l+1} = f(y_l)$$

where  $h(x_l)$  represents the identity mapping,  $F$  represents the residual function,  $x_l$  represents the input, and  $W_l$  represents the weight coefficient. The identity mapping is denoted by the symbol  $h(x_l) = x_l$  is the foundation of the



**Fig. 2** ResNet residual block



**Fig. 3** Original residual unit

ResNet architecture. The residual networks were created for networks with a layer count of 34, 50, 101, and 152. ResNet-50 was employed in this investigation. The network is made up of 50 layers.

## 4 Procedure & methodology

### 4.1 Archimedes optimization algorithm (AOA)

AOA is a physics-inspired algorithm, more specifically Archimedes’ law. Fatma Hashim invented this method in 2020 and it is classified as a meta-heuristic [30]. The uniqueness of this algorithm is in the solution’s encoding, which includes three pieces of auditory information for the fundamental agents: Volume ( $V$ ), Density ( $D$ ), and Acceleration ( $Gamma$ ). Thus, the first set of agents in  $Dim$  dimensions is formed randomly. We offer random values of  $V$ ,  $D$ , and  $Gamma$  as additive data. Following that, an evaluation process is carried out on each object to decide which one is the best ( $O_{best}$ ). During the AOA process,

density and volume are updated to adjust the acceleration based on the collision notion between objects, which is critical in determining the novel position of the current solution. The following are the general key steps in AOA:

- *The first stage - Initialization*

Using Eq. (3), this phase seeks to randomly initialize the real population of  $N$  objects. Additionally, each object's density ( $D_i$ ), volume ( $V_i$ ) and acceleration ( $\Gamma_i$ ) are randomly constructed using the given equations: Eqs. (4), (5) and (6):

$$O_i = O_i^{Min} + r_1 \times (O_i^{Max} - O_i^{Min}); i = 1, 2, \dots, N \quad (3)$$

$$D_i = r_2 \quad (4)$$

$$V_i = r_3 \quad (5)$$

$$\Gamma_i = \Gamma_i^{Max} + r_4 \times (\Gamma_i^{Max} - \Gamma_i^{Min}); i = 1, 2, \dots, N \quad (6)$$

Where  $O_i$  represents the  $i$ th object,  $O_i^{Max}$  and  $O_i^{Min}$  are the search-maximal space's and minimal bounds, respectively.

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

The population will be analyzed by assigning a score to each candidate in order to decide which object is the best ( $O_{Best}$ ) by joining their best values of density ( $D_{Best}$ ), volume ( $V_{Best}$ ) and acceleration ( $\Gamma_{Best}$ ).

- *The second stage - Density and volume adjustments*

The density and volume values for each object are adjusted in this stage by controlling the best density and volume using Eqs. (7) and (8):

$$D_i^{t+1} = D^t + s_1 \times (D_{Best} - D_i^t) \quad (7)$$

$$V_i^{t+1} = V_i^t + s_2 \times (V_{Best} - V_i^t) \quad (8)$$

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

- *The third stage - Transfer coefficient & density scalar*

This process involves colliding objects until the equilibrium state is reached. The transfer function ( $T_c$ ) is primarily responsible for switching from exploration to exploitation phase, as defined by Eq. (9):

$$T_c = \exp\left(\frac{It - It_{Max}}{It_{Max}}\right) \quad (9)$$

The  $T_c$  increases exponentially over time until reaching 1.  $It$  is the current iteration, while  $It_{Max}$  denotes the maximum number of iterations. In contrast, the density factor  $d_s$  is supposed to gradually decrease in order to convert the global search to a local search using Eq.(10):

$$d_s^{t+1} = \exp\left(\frac{It - It_{Max}}{It_{Max}}\right) - \left(\frac{It}{It_{Max}}\right) \quad (10)$$

- *The fourth stage - Exploration phase*

The collision between agents occurs in this step via a random selection of material ( $Mr$ ). Thus, when the transfer coefficient value is less than or equal to 0.5, the acceleration objects are updated using Eq. (11).

$$\Gamma_i^{t+1} = \frac{D_{Mr} + V_{Mr} \times \Gamma_{Mr}}{D_i^{t+1} \times V_i^{t+1}} \quad (11)$$

- *The fifth stage - Exploitation phase*

This phase omits the possibility of agents colliding. Thus, when the transfer coefficient value is greater than 0.5, the acceleration objects are updated using Eq. (12)

$$\Gamma_i^{t+1} = \frac{D_{Best} + V_{Best} \times \Gamma_{Best}}{D_i^{t+1} \times V_i^{t+1}} \quad (12)$$

Where  $\Gamma_{Best}$  is the acceleration of the optimal object  $O_{Best}$ .

- *The sixth stage - Normalization of acceleration*

This stage consists to normalize the acceleration in order to determine the rate of change using (13):

$$\Gamma_{i-norm}^{t+1} = \lambda \times \frac{\Gamma_i^{t+1} - \Gamma^{Min}}{\Gamma^{Max} - \Gamma^{Min}} + \mu \quad (13)$$

Where  $\lambda$  and  $\mu$  are fixed to 0.9 and 0.1, respectively. The  $\Gamma_{i-norm}^{t+1}$  determines the percentage of step that each agent will change. A greater acceleration value indicates that the candidate is operating in the exploration mode; otherwise, the object is operating in the exploitation mode.

- *The seventh stage - The Update process*

For exploration phase ( $T_c \leq 0.5$ ), Eq. (14) modifies the position of the  $i$ th object in iteration  $it + 1$ , however, Eq. (15) updates the object position in exploitation phase ( $T_c > 0.5$ ).

$$O_i^{t+1} = O_i^t + c_1 \times r_5 \times \Gamma_{i-norm}^{t+1} \times d_s \times (O_{rand} - O_i^t) \quad (14)$$

Where  $c_1$  is equal to 2.

$$O_i^{t+1} = O_{Best}^t + F.c_2.r_6.\Gamma_{i-norm}^{t+1}.d_s.(\delta.O_{Best} - O_i^t) \quad (15)$$

where  $c_2$  is fixed to 6.

The parameter *delta* has a positive correlation with time and is proportional to the transfer coefficient  $T_c$ , i.e.,  $\delta = 2 \times T_c$ . This parameter's primary function is to maintain a good balance between exploration and exploitation operations. During the initial iterations, the margin between the best and worst objects is greater, resulting in a high random walk. However, in the final iterations, the margin will be lowered and a low random walk will be provided.

$F$  is used for flagging, which regulates the direction of the search via Eq. (16):

$$F = \begin{cases} +1 & \text{if } \zeta \leq 0.5 \\ -1 & \text{if } \zeta > 0.5 \end{cases} \quad (16)$$

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

• *The eighth stage - The evaluation*

This step, the novel population is evaluated using score index  $Sc$  in order to determine the best object  $O_{Best}$  and the best additive information including  $D_{Best}$ ,  $V_{Best}$ , and  $\Gamma_{Best}$ .

---

**Algorithm 1** Pseudo code of AOA [30].

---

```

1: procedure AOA( $N, It_{Max}, \lambda, \mu, c_1$  and  $c_2$ )
2:   Initialize  $N$  objects with their densities ( $D$ ), volumes ( $V$ ) and accelerations ( $\Gamma$ ) using Eq.(3), Eq.(4), Eq.(5) and Eq.(6), 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 ( $\Gamma_{Best}$ ).
5:    $It = 1$ 
6:   while  $It \leq It_{Max}$  do
7:     for each object  $i$  do
8:       Adjust density and volume using Eq.(7) and Eq.(8)
9:       Adjust transfer coefficient ( $T_c$ ) and density scalar ( $d_s$ ) using Eq.(9) and Eq.(10), respectively.
10:      if  $T_c \leq 0.5$  then           ▷ Exploration operation
11:        Adjust acceleration ( $\Gamma_i$ ) using Eq.(11).
12:        Normalize acceleration ( $\Gamma_i$ ) using Eq.(13)
13:        Adjust position using Eq.(14)
14:      else                       ▷ Exploitation operation
15:        Adjust acceleration using Eq.(12).
16:        Adjust flagging control  $F$  using Eq.(16)
17:        Adjust position using Eq.(15)
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 ( $V_{Best}$ ) and best acceleration ( $\Gamma_{Best}$ ).
22:    Set  $It = It + 1$ 
23:  end while
24:  return Best object with their quality
25: end procedure

```

---

## 4.2 Trigonometric operator (TO)

Trigonometric operator is inspired from SCA algorithm, developed by [36]. This operator can enhance deeply the exploitation step of AOA using sine and cosine functions, as shown in Eqs. (17) or (18), respectively:

$$O_i^{It+1} = O_i^{It} + \alpha_1 \times \sin(\alpha_2) \times |\alpha_3 O_{Best}^{It} - O_i^{It}| \quad \text{if } \alpha_4 < 0.5 \quad (17)$$

$$O_i^{It+1} = O_i^{It} + \alpha_1 \times \cos(\alpha_2) \times |\alpha_3 O_{Best}^{It} - O_i^{It}| \quad \text{if } \alpha_4 \geq 0.5 \quad (18)$$

where  $O_{Best}^{It}$  is the best object at  $It$  iteration,  $O_i^{It}$  is the current solution in  $i$ th dimension at  $t$  iteration,  $||$  denotes the absolute cost.  $\alpha_1, \alpha_2, \alpha_3$  and  $\alpha_4$  are random numbers.

The parameter  $\alpha_1$  regulates the ratio of exploration to exploitation. This parameter is updated using the following formula:

$$\alpha_1 = s - It \frac{s}{It_{Max}} \quad (19)$$

where  $It$  is the current iteration,  $It_{Max}$ , is the maximum number of iteration and  $s$  is fixed to 2.  $\alpha_2$  controls whether the following solution will move in the direction of the target or away from it.  $\alpha_3$  denotes the weight assigned to the optimal solution in order to stochastically emphasize it. ( $\alpha_3 > 1$ ) or downplay ( $\alpha_3 < 1$ ) the effect of the destination on the distance definition [73]. Using Eqs. (17) and (18), the parameter  $\alpha_4$  enables switching between sine and cosine or vice versa. The process of TO is illustrated in Algorithm 2

---

**Algorithm 2** Trigonometric operator (TO)

---

```

1: procedure TO( $s, It_{Max}$ )
2:    $It = 1$ 
3:   while  $It \leq It_{Max}$  do
4:      $\alpha_1 = s - It \frac{s}{It_{Max}}$ 
5:     for  $1 := 1 : N$  do
6:        $\alpha_2 = 2 \times pi \times rand_2$ 
7:        $\alpha_3 = 2 \times rand_3$  &  $\alpha_4 = rand_4$ 
8:       if  $\alpha_4 \leq 0.5$  then
9:          $O_i^{It+1} = O_i^{It} + \alpha_1 \times \sin(\alpha_2) \times |\alpha_3 O_{Best}^{It} - O_i^{It}|$ 
10:       else
11:          $O_i^{It+1} = O_i^{It} + \alpha_1 \times \cos(\alpha_2) \times |\alpha_3 O_{Best}^{It} - O_i^{It}|$ 
12:       end if
13:       Compute the score for each object
14:     end for
15:     Set  $It = It + 1$ 
16:   end while
17:   return the best object
18: end procedure

```

---

## 5 scAOA-based FS for gender recognition (GR)

To improve the exploitation phase of the original AOA method and avoid convergence to local minima, we developed a new approach called scAOA that incorporates the trigonometric operator (TO). By utilizing the sine and cosine mathematical functions, this operator ensures a more balanced approach to exploration and exploitation. The integration of this operator (TO) provides a good solution to escape from local optima. The design of GR based on scAOA is depicted in Fig. 4 and it contains five basic processes, which are detailed as follows:



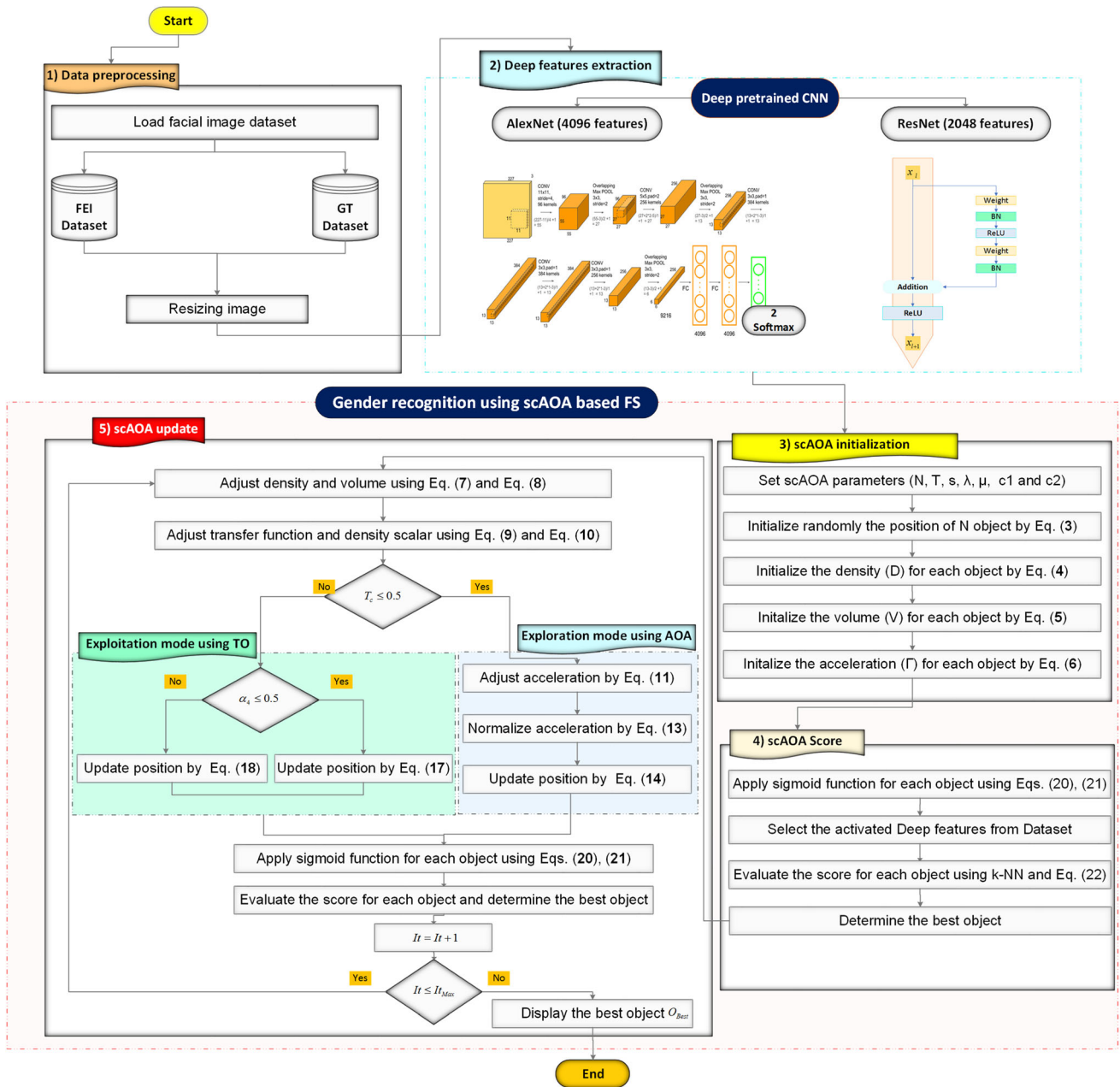


Fig. 4 The design framework of scAOA-based FS for gender recognition

- (1) *Preprocessing data* This stage consists to load the datasets of facial images, which will be divided to  $k$ -folds. All images must be resized to  $227 \times 227 \times 3$  for AlexNet and  $224 \times 224 \times 3$  for ResNet.
- (2) *Deep feature extraction* In this stage, two pre-trained CNN are used to extract the trainable features, which are more efficient compared to handcrafted descriptors. AlexNet extracts 4096 features, while ResNet provides 2048.
- (3) *Initialization* As the case of the majority of computational algorithms, scAOA begins by generating an initial population of  $N$  objects; each object has a

dimension  $Dim$  in the search space that is constrained by the maximal and minimal limits, as defined by Eq. (3). The initial values of Volume ( $V$ ), Density ( $D$ ), and Acceleration ( $\Gamma$ ) are generated by Eqs. (4), (5) and (6). The process of FS required to convert the real values to binary using the sigmoidal function, defined by the following equations:

$$O_i^{t+1} = \begin{cases} 0, & \text{if } rand < Sig(O_i^t) \\ 1, & \text{if } rand \geq Sig(O_i^t) \end{cases} \quad (20)$$

Where:

$$Sig(O_i^t) = \frac{1}{1 + e^{(-10*(O_i^t - 0.5))}} \tag{21}$$

Any solution is represented as a one-dimensional vector; the length is specified by the number of deep features. Any cell may have one of two values, 0 or 1, where 1 indicates that the appropriate feature has been selected and 0 indicates that it has not been selected.

- (4) *Score (Fitness) evaluation* Generally, feature selection seeks to decrease both the number of features and the error rate of classification. In other words, classification accuracy is maximized by deleting superfluous and redundant traits and maintaining only the most pertinent ones. The *k*-NN classifier was used in this investigation due to its ease in evaluating the score. So, the score (Fitness) for each object is evaluated by:

$$Sc = 0.99 \times (1 - Cr) + 0.01 \times \frac{|Sel_f|}{|Tot_f|} \tag{22}$$

where *Cr*, *Sel<sub>f</sub>* are the accuracy obtained by *k*-NN (*k* = 5) and the size of selected deep features, respectively. *Tot<sub>f</sub>* is the total number of trainable features provided by AlexNet/ResNet.

- (5) *Updating process* First, sAOA seeks to update the density, volume, transfer coefficient, and density scalar using Eqs. (7), (8), (9) and (10), respectively. Second, the update of objects is realized by the exploration mode (when *T<sub>c</sub>* ≤ 0.5), which applies the adjustment of acceleration using Eq. (11), the normalization of acceleration using Eq. (13) and the adjustment of position using Eq. (14). The exploitation mode is realized by the integration of trigonometric operators (TO), which ensures a good balance between exploration and exploitation modes. More exactly, TO utilized two strategies: sine and cosine strategies based on the condition (*α<sub>4</sub>* ≤ 0.5), using Eqs. (17) and (18), respectively. This integration enhances deeply the convergence to global solution. The third step consists of evaluating the score for each object using Eq. (22), in order to find the best candidate *O<sub>Best</sub>*. The evaluating and updating stages are repeated indefinitely until the termination condition is satisfied. This condition is utilized in this study to determine the quality of the suggested approach for locating the optimal subset of features within the given number of iterations.

The process of gender recognition using sAOA algorithm is outlined in Algorithm 3.

**Algorithm 3** Pseudo code of sAOA.

```

1: procedure sAOA(N, ItMax, λ, μ, s, c1 and c2 )
   ▷ I) Data preprocessing
2:   Load facial image dataset( GT/FEI).
3:   Resize image to 227 × 227 × 3 for AlexNet and 224 × 224
   × 3 for ResNet.
   ▷ II) Deep features extraction
4:   Extract deep features using pre-trained models: AlexNet
   (4096) and ResNet(2048).
5:   Split data to training and testing set.
   ▷ III) sAOA initialization
6:   Set the initial parameters of sAOA (N = 10, ItMax =
   100, λ = 0.9 , μ = 0.1 s = 2, c1 = 2 and c2 = 6).
7:   Initialize N objects with their densities (D), volumes
   (V) and accelerations (Γ) using Eq.(3), Eq.(4), Eq.(5) and
   Eq.(6), respectively.
   ▷ IV) sAOA score evaluation
8:   Apply sigmoid function for each object using Eq. (20)
   and Eq. (21).
9:   Select the activated deep features from dataset (Train
   and test sets).
10:  Compute the score of each object using k-NN and Eq.
   (22).
11:  Determine the best object (OBest) with their best density
   (DBest), best volume (VBest) and best acceleration
   (ΓBest).
   ▷ V) sAOA update
12:  It = 1
13:  while It ≤ ItMax do
14:    α1 = s - It *  $\frac{s}{It_{Max}}$ 
15:    for each object i do
16:      Adjust density and volume using Eq.(7) and
   Eq.(8)
17:      Adjust transfer coefficient (Tc) and density scalar
   (ds) using Eq.(9) and Eq.(10), respectively.
18:      if Tc ≤ 0.5 then ▷ Exploration operation
19:        Adjust acceleration (Γi) using Eq.(11).
20:        Normalize acceleration (Γi) using Eq.(13)
21:        Adjust position using Eq.(14)
22:      else ▷ Exploitation operation using TO
23:        α2 = 2 × pi × rand2
24:        α3 = 2 × rand3
25:        α4 = rand4
26:        if α4 ≤ 0.5 then
27:          Adjust position using Eq.(17).
28:        else
29:          Adjust position using Eq.(18)
30:        end if
31:      end if
32:      Apply sigmoid function for each object using Eq.
   (20) and Eq. (21).
33:      Compute the score of each object using k-NN and
   Eq. (22).
34:    end for
35:    Determine the best object (OBest) with their best
   density (DBest), best volume (VBest) and best accelera-
   tion (ΓBest).
36:    Set It = It + 1
37:  end while
38:  return Best object with their quality
39: end procedure

```

## 6 Experimental results

In order to realize a fair analysis, the efficiency of AOA is compared with different and recent computational algorithms inspired from swarm intelligence, mathematical algorithm, genetic evolution and physical algorithms, tested on two datasets (GT & FEI datasets) in the same condition by employing two deep descriptors including AlexNet and ResNet.

### 6.1 Simulation setup

In this study, we set the population size and maximum number of iterations to 10 and 100, respectively. Due to the stochastic nature of the computational algorithms, each algorithm is run 30 times separately. To recognize GR using FS, we randomly split datasets using  $k$ -fold cross-validation. The computer’s CPU is an Intel(R) Core(TM) i7-5500U processor running at 2.40 GHz, and the RAM is 8 GB.

#### 6.1.1 Statistical metrics

To confirm the efficiency of the scAOA algorithm in the field of facial analysis-based FS, 8 meta-heuristics, including GWO, SCA, GOA, ALO, PSO, SGA, SSA and basic AOA, were used to compare the performance of the proposed algorithm under the same conditions. In order to compute the conventional metrics as correct rate of gender classification ( $Cr$ ), Recall ( $Re$ ), precision ( $Pr$ ) and F-score( $F_{sc}$ ), we define the confusion matrix, which is illustrated in Table 2. Also, other measures are shown as the score (Fitness), the selection ratio, and CPU time. As a result, all metrics are expressed in terms of the mean and standard deviation, which are described by:

Where TP The classifier determines correctly the person knowing that their class is male , TN the classifier determines the person correctly knowing that their class is female, FP the classifier assigns the person to the male class knowing that the subject is female, FN: the classifier assigns the person to the female class knowing that the subject is male

- Mean Correct rate of classification ( $\mu_{Cr}$ ): It represents the accuracy, which is defined by:

$$CR = \frac{TP + TN}{TP + FN + FP + TN} \tag{23}$$

$\mu_{CR}$  is measured as follows:

$$\mu_{CR} = \frac{1}{30} \sum_{r=1}^{30} CR_r^{best} \tag{24}$$

- Average recall ( $\mu_{Re}$ ): It indicates the percentage of predicting positive patterns (TPR), that is defined by:

$$Re = \frac{TP}{TP + FN} \tag{25}$$

The  $\mu_{Re}$  is calculated from the best object ( $O_{best}$ ) using:

$$\mu_{Re} = \frac{1}{30} \sum_{r=1}^{30} Re_r^{best} \tag{26}$$

- Average precision ( $\mu_{Pr}$ ): It indicates the frequency of true expected samples as follows:

$$Pr = \frac{TP}{FP + TP} \tag{27}$$

and the mean precision(  $\mu_{Pr}$ ) can be calculated as follows::

$$\mu_{Pr} = \frac{1}{30} \sum_{r=1}^{30} Pr_r^{best} \tag{28}$$

- Average cost value ( $\mu_{Score}$ ): The score is related to the error rate of classification and selection ratio measures, as shown in Eq. (22). A lower value of the score signifies better performance. Its average is computed by:

$$\mu_{Score} = \frac{1}{30} \sum_{r=1}^{30} Score_r^{best} \tag{29}$$

- Average selection ratio ( $\mu_{Sr}$ ):This metric denotes the ratio of selected features compared to the total number of features. It is calculated as follows:

$$\mu_{Sr} = \frac{1}{30} \sum_{r=1}^{30} \frac{Sel_f^{best}(r)}{Tot_f} \tag{30}$$

where  $Sel_f^{best}r$  is the cardinal of the best agent’s selected features for  $r^{th}$  execution.

- Mean F-score ( $\mu_{F_{Score}}$ ): This metric is used to calculate the harmonic average of recall and precision. It is already in use for balanced data, which can be calculated using Eq. (31):

$$F_{Score} = 2 \times \frac{Re \times Pr}{Re + Pr} \tag{31}$$

and the mean F-score can be calculated using:

**Table 2** Confusion matrix

Actual	Predicted	
	Male	Female
Male	TP	FN
Female	FP	TN

**Table 3** Parameters settings of scAOA and other computational algorithms

Algorithms	Parameters setting
Common settings	Population size ( $N = 10$ ) Maximum number of iterations ( $It_{Max} = 100$ ) Maximal limit is fixed to 1 Minimal limit is fixed to 0
AOA	$c_1 = 2$ and $c_2 = 6$ $\lambda = 0.9$ and $\mu = 0.1$
scAOA	$c_1 = 2$ and $c_2 = 6$ $\lambda = 0.9$ and $\mu = 0.1$ $s=2$
GOA	$C_{max} = 0.1$ and $C_{min} = 4E - 05$
SGA	$P_c = 0.8$ and $P_m = 0.1$ <i>Elitismratio</i> = 0.9
SCA	$A = 2$
GWO	$a \in [2, 0]$
SSA	$c_1 = 2 * \exp(-4t/It_{Max})^2$ , $c_2 = rand$ and $c_3 = rand$
ALO	-
PSO	$W_{max} = 0.9$ and $W_{min} = 0.2$ $c_1 = 2$ and $c_2 = 2$

$$\mu_{F_{Score}} = \frac{1}{30} \sum_{r=1}^{30} F_r^{Score-best} \quad (32)$$

- Average CPU time ( $\mu_{Cpu}$ ): It is the average time required to compute each process, that is:

$$\mu_{Cpu} = \frac{1}{30} \sum_{r=1}^{30} T_r^{best} \quad (33)$$

- Standard deviation ( $\sigma$ ): is the algorithm's quality and the analysis of the obtained results through several executions and metrics. It is computed for all previously mentioned steps.

### 6.1.2 Parameters settings

This subsection defines each optimizer's parameters. To ensure a fair comparison, it is necessary to enumerate the algorithms utilized to complete the task of gender recognition from faces. The suggested technique scAOA and other computational algorithms such as AOA, SCA, GOA, SSA, ALO, SGA, PSO, and GWO have their parameter settings defined in Table 3.

## 6.2 Description of datasets

– *FEI dataset* There are 200 individuals in this Brazilian dataset. It is critical to keep in mind that each individual has 14 photos, totaling 2800. The photographs were taken

on a white background and are of high quality in terms of color. Between the ages of 19 and 40, this category applies. Additionally, modifications to the face's look, such as hairstyles and adornments, have been integrated. Because half of the cases are male and half are female, the dataset is balanced.<sup>1</sup> We notice that each image has a resolution of  $640 \times 480$  in original format.

– *Georgia Tech Face dataset (GT)* It comprises 50 individuals photographed between 04/06/99 and 11/15/99 over two sessions. Each individual is represented by fifteen photos, for a total of 750 pictures of  $640 \times 480$  pixels, in original format. A face is typically  $150 \times 150$  pixels in size. In terms of light, scale, and expression, images are frontal. Seven women and 43 males are represented in this dataset.<sup>2</sup>

## 6.3 Results & discussion

– *In terms of average fitness and average accuracy*

The average fitness and standard deviation results obtained by the scAOA approach and the other competitors algorithms are shown in Table 4. Two corpora and two pre-trained CNN models (AlexNet & ResNet) were employed to extract the deep features. The deep analysis of results across the two corpora used, showed that the quantitative results obtained by the proposed approach scAOA achieved

<sup>1</sup> <https://fei.edu.br/cet/facedatabase.html>.

<sup>2</sup> [http://www.anefian.com/research/face\\_reco.html](http://www.anefian.com/research/face_reco.html).

**Table 4** The impact of deep features models on the performance of scAOA over average fitness and standard deviation

Fitness algorithms	FEI dataset		Georgia tech daraset	
	AlexNet	ResNet	AlexNet	ResNet
scAOA	<b>0.0055</b> ± 0.0039	<b>0.0225</b> ± 0.0082	<b>0.0015 ± 0.0011</b>	<b>0.0050 ± 0.0020</b>
AOA	0.0281 ± 0.0033	0.0246 ± 0.0050	0.0071 ± 0.029	0.0060 ± 0.0020
SCA	0.0130 ± 0.0039	0.0271 ± 0.0071	0.0033 ± 0.0018	0.0180 ± 0.0024
GOA	0.0421 ± 0.0044	0.0586 ± 0.0054	0.0195 ± 0.0018	0.0340 ± 0.0040
SSA	0.0296 ± 0.0030	0.0403 ± 0.0057	0.0140 ± 0.0024	0.0200 ± 0.0030
ALO	0.0278 ± 0.0032	0.0334 ± 0.0043	0.0116 ± 0.0024	0.0170 ± 0.0040
SGA	0.0333 ± 0.0041	0.0424 ± 0.0056	0.0158 ± 0.0021	0.0240 ± 0.0050
PSO	0.0289 ± <b>0.0015</b>	0.0375 ± <b>0.0033</b>	0.0121 ± 0.0012	0.0190 ± 0.0020
GWO	0.0362 ± 0.0044	0.0498 ± 0.0077	0.0147 ± 0.0026	0.0290 ± 0.0040

better performance using the two pretrained CNN models (AlexNet & ResNet) compared to optimization algorithms such as basic AOA, SCA, GOA, SSA, ALO, SGA, PSO, and GWO. It is clear that the results of the proposed scAOA approach based on AlexNet as deep features are significantly better compared to ResNet-based scAOA. The lower values of fitness for FEI and GT datasets are equal to 0.005 and 0.0015, respectively. The obtained performance can be justified by two reasons: The first one is focused on the integration of trigonometric operators to AOA which allows to keep a significant deep features, while the second reason lies in the type of pre-trained network, which is used at the level of features extraction step. According to the results of Table 4, the pre-trained CNN models based on AlexNet provides a good performance of fitness indicated by Eq. (22), which ensure a compromise between lower rate of error classification and lower rate of selection ratio. Concerning the measurement of the standard deviation, we notice that the scAOA algorithm obtained lower values than other algorithms at the level of the base GT, while the PSO algorithm obtained low values at the level of the base FEI. It is important to mention that the second best place was obtained by AOA using Alexnet at the FEI base level, while SCA took the second place at the GT base. In

addition, for ResNet deep features, AOA is ranked second for both databases (FEI & GT).

The results of the gender categorization are summarized in Table 5 in terms of mean accuracy and standard deviation for both datasets (FEI and GT) utilizing two distinct types of deep characteristics (AlexNet and ResNet). When compared to other algorithms such as AOA, SCA, GOA, SSA, GWO, SGA, PSO, and ALO in terms of average accuracy, the suggested scAOA algorithm based on the two deep descriptors demonstrated a clear performance. We see that AlexNet-based scAOA is much superior for both corpora (FEI and GT datasets), with classification rates of 99.5% and 99.95% percent, respectively, for FEI and GT. This phenomenon is explained by the fact that trigonometric operators are integrated at the operational step level. Additionally, scAOA is ranked first in terms of STD for the GT base, whereas SCA is ranked first for the FEI dataset.

*– In terms of recall and precision*

The recall and precision of the proposed method scAOA with eight wrapper FS algorithms using two deep descriptors (AlexNet & ResNet) are listed in Tables 6 and 7.

By inspecting the average recall and average precision values shown in Tables 6 and 7, we can see clearly that

**Table 5** The impact of deep features models on the performance of scAOA over average accuracy and standard deviation

Accuracy (%) algorithms	FEI Dataset		Georgia tech daraset	
	AlexNet	ResNet	AlexNet	ResNet
scAOA	<b>99.50</b> ± 0.39	<b>97.92</b> ± 0.77	<b>99.95 ± 0.11</b>	<b>99.69 ± 0.17</b>
AOA	97.47 ± 0.35	97.88 ± 0.50	99.48 ± 0.28	99.64 ± 0.23
SCA	98.73 ± 0.39	97.33 ± 0.71	99.72 ± 0.18	98.27 ± 0.44
GOA	96.25 ± 0.45	94.58 ± 0.54	99.52 ± 0.18	97.02 ± 0.40
SSA	97.48 ± 0.31	96.42 ± 0.57	99.05 ± 0.25	98.44 ± 0.30
ALO	97.65 ± 0.33	97.10 ± 0.42	99.28 ± 0.24	98.74 ± 0.36
SGA	97.12 ± 0.41	96.20 ± 0.57	98.88 ± 0.21	98.11 ± 0.53
PSO	97.55 ± <b>0.15</b>	96.70 ± <b>0.34</b>	99.25 ± 0.13	98.56 ± 0.21
GWO	96.72 ± 0.41	95.37 ± 0.78	98.80 ± 0.25	97.54 ± 0.35



**Table 6** The impact of deep features models on the performance of scAOA over average recall and standard deviation

Recall (%) algorithms	FEI dataset		Georgia tech daraset	
	AlexNet	ResNet	AlexNet	ResNet
scAOA	<b>99.54</b> ± 0.36	<b>97.98</b> ± 0.73	<b>99.81</b> ± <b>0.39</b>	<b>98.86</b> ± <b>0.67</b>
AOA	97.47 ± 0.35	97.86 ± 0.47	98.22 ± 0.92	98.61 ± 0.85
SCA	98.70 ± 0.40	97.33 ± 0.71	99.05 ± 0.68	94.74 ± 1.45
GOA	96.10 ± 0.48	94.58 ± 0.54	94.83 ± 0.59	91.29 ± 1.35
SSA	97.39 ± 0.32	96.42 ± 0.58	96.52 ± 1.02	95.33 ± 1.03
ALO	97.56 ± 0.35	97.11 ± 0.42	97.36 ± 0.89	96.26 ± 1.04
SGA	97.00 ± 0.43	96.20 ± 0.57	95.91 ± 0.81	94.37 ± 1.53
PSO	97.46 ± <b>0.16</b>	96.71 ± <b>0.34</b>	97.23 ± 0.50	95.88 ± 0.76
GWO	96.60 ± 0.42	95.36 ± 0.78	95.87 ± 0.93	92.68 ± 0.99

**Table 7** The impact of deep features models on the performance of scAOA over average precision and standard deviation

Precision (%) algorithms	FEI dataset		Georgia tech daraset	
	AlexNet	ResNet	AlexNet	ResNet
scAOA	<b>99.45</b> ± 0.43	<b>97.88</b> ± 0.78	<b>99.97</b> ± <b>0.06</b>	<b>99.79</b> ± <b>0.18</b>
AOA	97.49 ± 0.35	97.88 ± 0.55	99.19 ± 0.70	99.77 ± 0.25
SCA	98.76 ± 0.38	97.36 ± 0.71	99.76 ± 0.29	98.71 ± 0.50
GOA	96.46 ± 0.40	94.60 ± 0.54	98.76 ± 0.53	97.37 ± 0.58
SSA	97.59 ± <b>0.28</b>	96.42 ± 0.57	99.37 ± 0.27	98.76 ± 0.47
ALO	97.75 ± 0.29	97.11 ± <b>0.42</b>	99.53 ± 0.29	98.96 ± 0.55
SGA	97.26 ± 0.37	96.21 ± 0.56	99.24 ± 0.35	98.45 ± 0.63
PSO	97.25 ± 0.46	95.86 ± 0.57	99.21 ± 0.42	98.55 ± 0.62
GWO	96.86 ± 0.39	95.38 ± 0.77	98.92 ± 0.51	97.98 ± 0.63

scAOA outperforms all competitors advanced algorithms based on both deep features (ResNet & AlexNet) for both datasets. For FEI dataset, the recall and precision values reached 99.54 and 99.45% using scAOA with AlexNet deep features, which indicated an excellent performance. TPR of 0.9954 showed that 99.54% of observation is classified correctly to the male class. Concerning precision value it achieved 0.9945 for FEI dataset, which indicated that 99.45% of observation is predicted correctly to the male class. For GT dataset, scAOA shown a higher values of recall and precision using deep features extracted by AlexNet. 0.9981 of TPR and 0.9997 of precision showed that 99.81% of observation is assigned correctly to the male class, while 99.97% is predicted correctly. The higher rate of TPR and precision can be explained by the significant deep trainable features and the efficiency of the wrapper scAOA to select the relevant features. Also, we can observe that the performance of average recall and average precision obtained by scAOA based on AlexNet is more better than scAOA based on ResNet. In addition, SCA based on deep AlexNet descriptor ranked in second position in terms of average recall and average precision for both Datasets. It can be seen that scAOA based on deep descriptors has a strong stability for GT dataset due to the

lower values of STD in terms of precision and recall metrics.

– *In terms of F-score and selection ratio* From Table 8, we can observe that the proposed method scAOA based on pretrained CNN (AlexNet & ResNet) outperforms all other competitors in terms of F-score. The value of F-score is more significant when the data are unbalanced like the case of GT dataset. 99.89% of F-score for GT dataset indicated a good balance performance of both categories: male and female. Furthermore, a great competition between AOA based on ResNet and SCA based on AlexNet is highlighted for second rank. Also, GOA based on deep features reached lower values of F-score for both datasets (GT & FEI).

According to the results of Table 9, which displays the average rate of selection ratio along with its standard deviation. A wrapper FS-based SCA demonstrated a high performance in terms of selecting relevant deep features for both datasets (FEI & GT datasets). Also, we can observe that the the proposed method scAOA provides a good behavior to select the optimal set of relevant deep features. Its performance is ranked second with a difference of 0.92 and 3.64% for FEI and GT based on AlexNet deep features, respectively. Additionally, the conventional AOA is ranked in third position. So, the integration of trigonometric

**Table 8** The impact of deep features models on the performance of scAOA over average F-score and standard deviation

F-score (%) algorithms	FEI dataset		Georgia tech daraset	
	AlexNet	ResNet	AlexNet	ResNet
scAOA	<b>99.49</b> ± 0.40	<b>97.91</b> ± 0.77	<b>99.89</b> ± <b>0.23</b>	<b>99.32</b> ± <b>0.38</b>
AOA	97.47 ± 0.35	97.87 ± 0.51	98.96 ± 0.68	99.18 ± 0.52
SCA	98.73 ± 0.39	97.33 ± 0.71	99.40 ± 0.39	96.58 ± 0.91
GOA	96.22 ± 0.46	94.58 ± 0.54	96.67 ± 0.41	94.00 ± 0.87
SSA	97.47 ± 0.31	96.42 ± 0.57	97.88 ± 0.58	96.94 ± 0.62
ALO	97.64 ± 0.33	97.10 ± 0.42	98.40 ± 0.55	97.55 ± 0.72
SGA	97.10 ± 0.41	96.20 ± 0.57	97.49 ± 0.50	96.26 ± 1.09
PSO	97.54 ± <b>0.15</b>	96.70 ± <b>0.34</b>	98.35 ± 0.29	97.20 ± 0.41
GWO	96.70 ± 0.41	95.37 ± 0.78	97.32 ± 0.57	95.09 ± 0.72

**Table 9** The impact of deep features models on the performance of scAOA over average selection ratio and standard deviation

Selection ratio (%) algorithms	FEI Dataset		Georgia tech daraset	
	AlexNet	ResNet	AlexNet	ResNet
scAOA	5.54 ± 2.08	18.46 ± 10.18	9.78 ± 3.01	16.09 ± 3.58
AOA	30.19 ± 1.43	26.48 ± 2.08	19.50 ± 3.99	23.82 ± 4.13
SCA	<b>4.62</b> ± 2.97	<b>7.34</b> ± 3.33	<b>16.4</b> ± 4.05	<b>12.42</b> ± 4.64
GOA	49.39 ± 1.78	49.41 ± 1.55	48.83 ± <b>1.23</b>	49.95 ± 1.28
SSA	47.15 ± 1.82	47.78 ± <b>1.31</b>	46.26 ± 1.89	47.54 ± 1.63
ALO	44.88 ± 1.44	47.22 ± 1.35	44.75 ± 1.55	47.91 ± 1.34
SGA	47.32 ± 1.61	48.04 ± 1.78	46.63 ± 1.60	49.03 ± 1.71
PSO	46.84 ± <b>1.25</b>	48.21 ± 1.45	47.00 ± 1.34	49.17 ± <b>1.13</b>
GWO	36.51 ± 11.24	39.54 ± 12.40	28.69 ± 7.88	45.15 ± 13.07

operators to the basic AOA enhances deeply the performance of correct classification rate by keeping the adequate deep features of pretrained CNN across FEI and GT datasets.

– *In terms of CPU time* The time consumed by the proposed scAOA method and the other competitors is recorded in Table 10. it is clear that the SCA algorithm is the fastest compared to other algorithms because SCA uses

simple sine / cosine-based math functions to fit the solutions. Furthermore, scAOA takes more times compared to SCA but it consumes less time compared to the the conventional AOA for both datasets using deep features. Also, It can be seen that ALO consumed a lot of time to carry out the task of gender classification based on the selection of attributes. This is due to the complexity of the operators used in ALO.

**Table 10** Cpu time

CPU time algorithms	FEI Dataset		Georgia tech daraset	
	AlexNet	ResNet	AlexNet	ResNet
scAOA	13.50	13.94	22.10	22.08
AOA	20.15	21.14	29.25	29.09
SCA	<b>12.05</b>	<b>12.53</b>	<b>17.94</b>	<b>19.85</b>
GOA	21.97	21.68	49.57	50.39
SSA	21.13	21.43	47.71	48.85
ALO	21.96	21.75	49.75	51.01
SGA	16.92	16.71	37.76	39.08
PSO	21.76	21.60	49.57	51.03
GWO	21.33	20.61	45.04	47.21

### 6.3.1 Statistical analysis

A statistical analysis is necessary to compare the efficiency of scAOA to that of other competitive algorithms. Thus, the Wilcoxon ranksum test is used to compare the accuracy values acquired by scAOA and those obtained by other algorithms, such as basic AOA, SCA, GOA, SSA, ALO, SGA, PSO, and GWO. As shown in Table 11, scAOA is statistically significant for all competitors in both datasets in the case of the AlexNet deep descriptor. Additionally, AOA-based ResNet dominates scAOA for both datasets. In general, scAOA based on deep descriptor ResNet is statistically significant with 87.5% of algorithms.

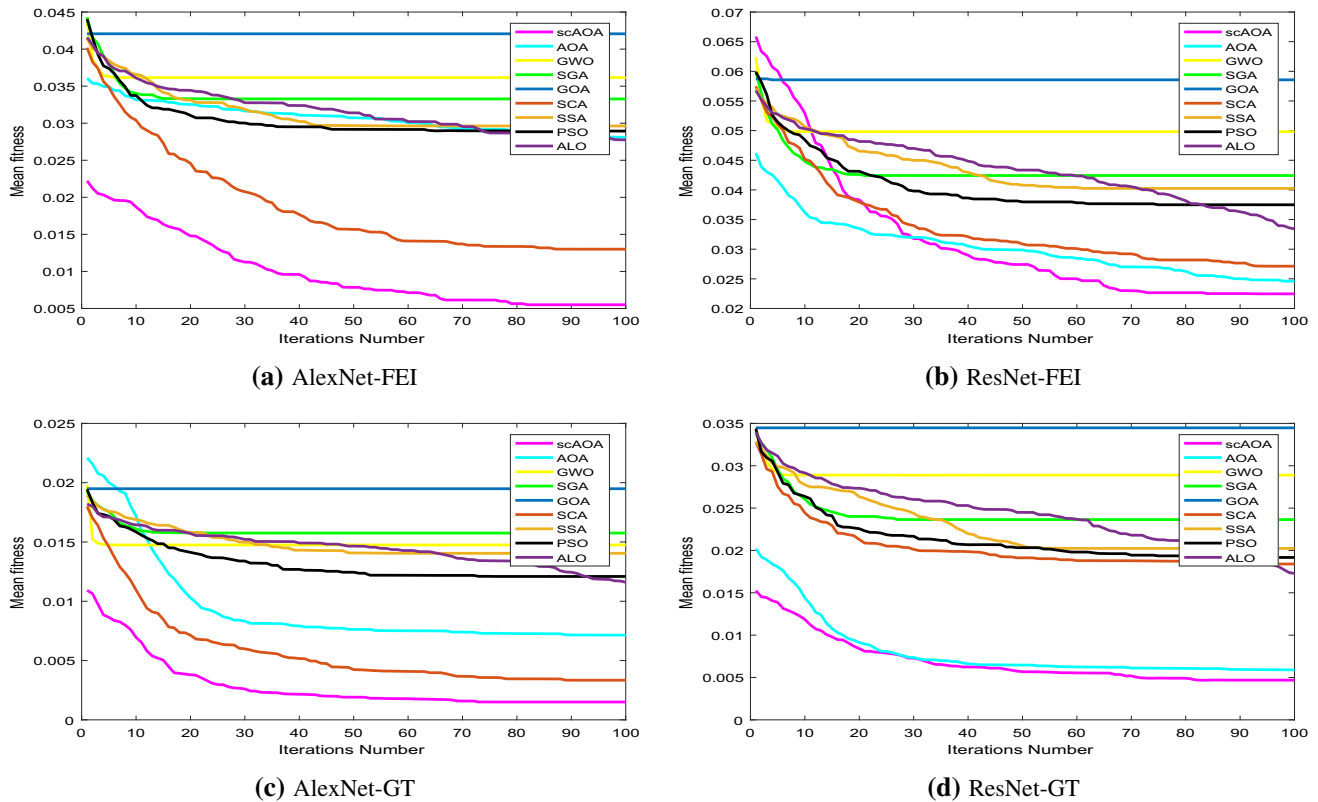
**Table 11** Statistical study using Wilcoxon’s test (In bold  $p > 0.05$ )

scAOA versus	FEI Dataset		Georgia tech daraset	
	AlexNet	ResNet	AlexNet	ResNet
AOA	1.28E-11	<b>7.47E-01</b>	6.36E-09	<b>3.30E-01</b>
SCA	3.80E-08	2.51E-03	2.03E-06	1.45E-11
GOA	1.60E-11	1.96E-11	2.62E-12	1.40E-11
SSA	9.73E-12	4.88E-09	4.68E-12	1.09E-11
ALO	1.05E-11	4.93E-06	1.34E-11	3.37E-11
SGA	1.40E-11	7.02E-10	4.13E-12	1.47E-11
PSO	2.10E-12	1.25E-08	1.59E-12	8.92E-12
GWO	1.46E-11	6.80E-11	2.64E-12	1.29E-11

### 6.3.2 Graphical analysis

Fig. 5 depicts the fitness curves of derived by the various optimizers based on AlexNet and ResNet for datasets (FEI and GT). By analyzing the behavior of the convergence of the scAOA algorithm for the two databases based on AlexNet deep descriptor, a speed convergence is illustrated by increasing the number of iterations compared to other algorithms, including basic AOA, SCA, GOA, SSA, ALO, SGA, PSO, and GWO.

For both datasets, we can see that scAOA and the conventional AOA based on ResNet descriptor highlight a great competition in the first iterations, but after 30 iterations scAOA start to be more efficient. This behavior can be interpreted by the use of trigonometric operators, which allow to enhance deeply the exploitation process. Additionally, as shown in Fig. 6, we have plotted the boxplot of the proposed method scAOA against other algorithms such as conventional AOA, SCA, GOA, SSA, ALO, SGA, PSO, and GWO in terms of accuracy. As illustrated in this figure, the suggested method scAOA based on deep features achieved greater mean and median accuracy values than other advanced algorithms for both datasets. The collected results demonstrate the proposed method’s efficacy in maintaining the highest classification accuracy, especially for AlexNet deep features. According to the obtained results, we observe that the gender performance of the FEI dataset results in a proportionately lower quality than that of the GT dataset. This is due to the complexity of the datasets, which is dependent on several challenging factors, including highlighting variation, facial expressions, different poses, and the dataset’s size.



**Fig. 5** Convergence curve of scAOA versus other swarm intelligence algorithms over all datasets

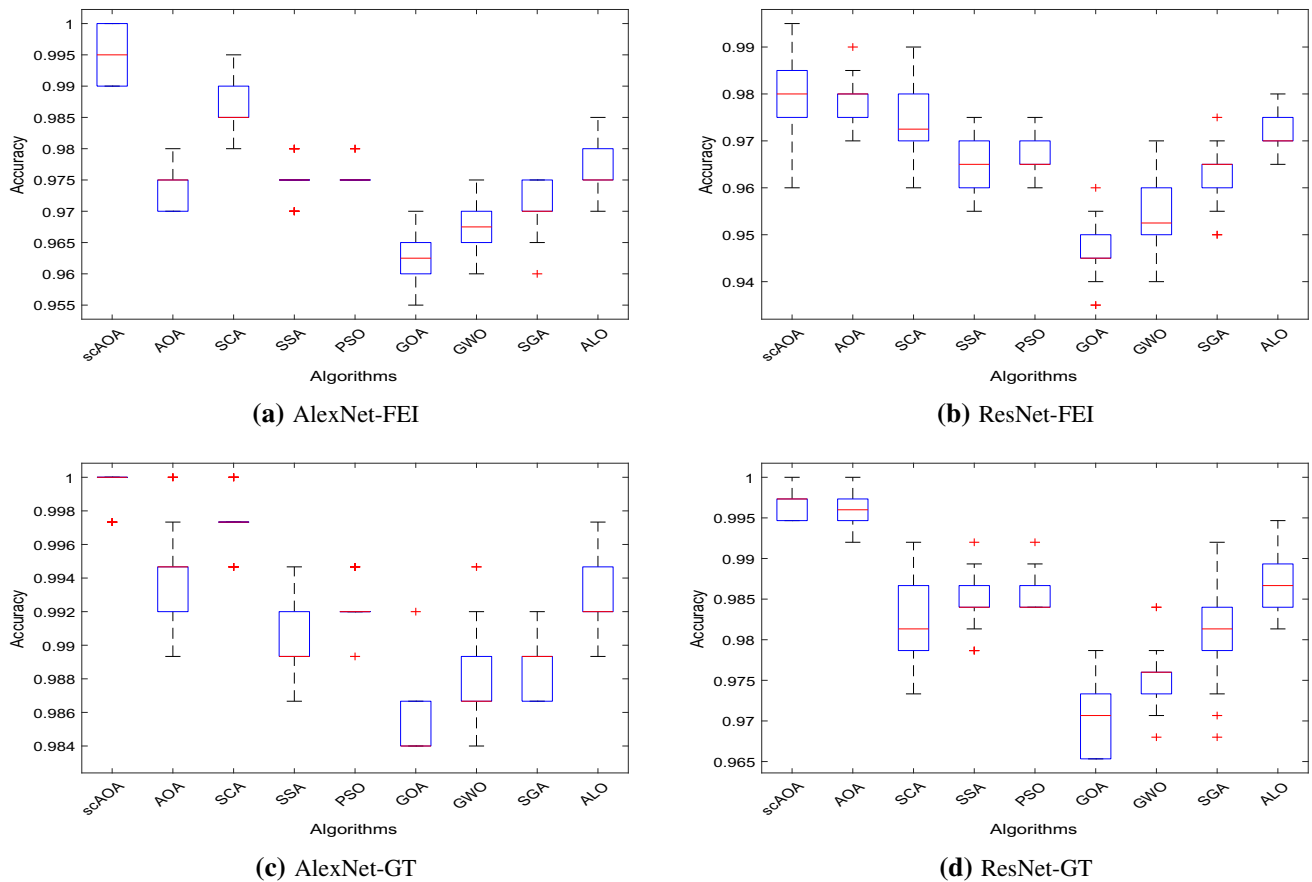


Fig. 6 Boxplot of scAOA versus other swarm intelligence algorithms over all datasets

### 6.4 Comparative study

#### 6.4.1 Comparative study between scAOA and based on Pretrained CNN (AlexNet & ResNet)

This study aims to compare the efficiency of scAOA wrapper feature selection approach based on two deep features provided by AlexNet/ResNet and *k*-NN. From the results of Table 12, it can be seen that scAOA based on *k*-NN as classifier and AlexNet as deep features outperforms *k*-NN using AlexNet as deep features with a margin of 2.21% i.e., 99.50% for FEI dataset and a margin of 1.91% for GT dataset i.e., 99.95%.

#### 6.4.2 Comparative study with the existing works

To demonstrate the efficacy of the suggested technique scAOA, numerous algorithms from the literature were chosen for a fair comparison, including machine learning, deep learning, and genetic algorithms.

Figs. 7 and 8 illustrate the proper rate of classification’s performance across both datasets.

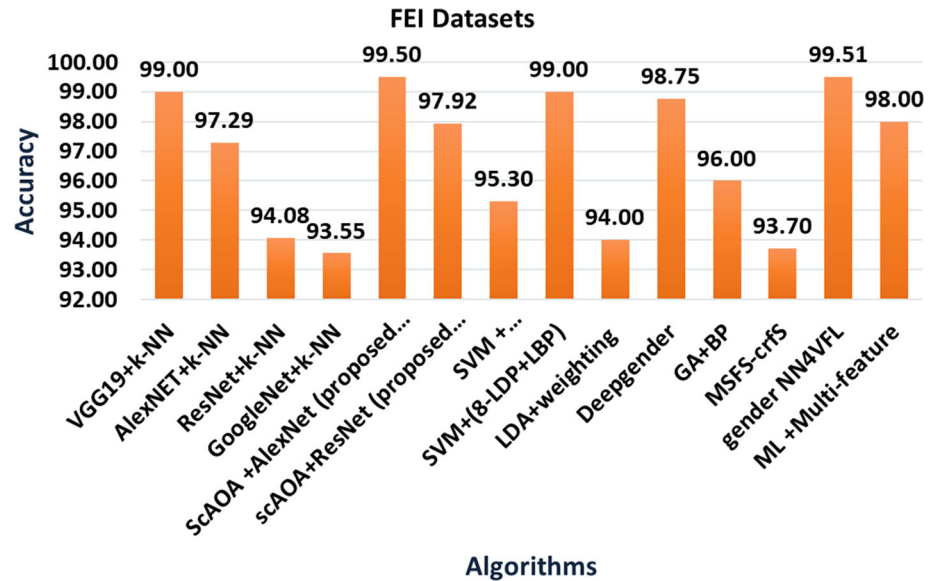
For *FEI* dataset

– As illustrated in Fig. 7, the first four algorithms combined trainable deep features with a *k*-NN classifier. That includes VGG19, AlexNet, ResNet, and GoogleNet. The best performance is obtained by VGG19+*k*-NN with 99%, which is lower than the performance of the proposed method scAOA+AlexNet, which attains 99.5% in terms of

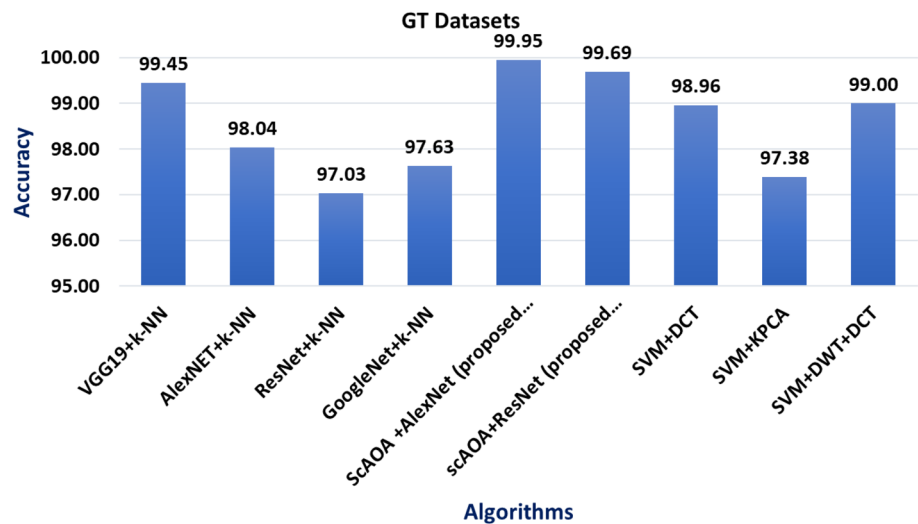
Table 12 Comparative study

Datasets algorithms	FEI dataset Accuracy	GT dataset Accuracy
AlexNet (Deep features)+ <i>k</i> -NN (classifier)	97.29	98.04
scAOA+ <i>k</i> -NN+AlexNet (Deep features)	<b>99.50</b>	<b>99.95</b>
ResNet (Deep features)+ <i>k</i> -NN (classifier)	94.08	97.03
scAOA+ <i>k</i> -NN+ResNet (Deep features)	97.92	99.69

**Fig. 7** The comparative study between scAOA based on pretrained CNN with existing algorithms – FEI dataset



**Fig. 8** The comparative study between scAOA based on pretrained CNN with existing algorithms – GT dataset



accuracy. Also, scAOA+k-NN outperforms Deepgender with a margin of 0.75% in terms of accuracy [57]. In addition, great competitiveness is shown between the proposed method scAOA and gender NN4VFL (combination between inception network and VFL) because the margin is insignificant with 0.01%.

Concerning the performance of scAOA based on AlexNet against ML and GA-based FS, scAOA still better than all previous algorithms illustrated in Fig. 7. The margin between scAOA+AlexNet and SVM based on 8-LDP+LBP is equal to 0.5% [57], whereas, the margin reached 3.5% compared to GA-based FS [67].

For *GT dataset*

– We start by comparing this dataset to pre-trained CNNs based on k-NN. Second, we compared the proposed scAOA + AlexNet and scAOA + ResNet approaches to machine learning methods in terms of performance. From the first comparison, we can see clearly that both proposed methods based on scAOA outperform all pre-trained CNN with k-NN including VGG19, AlexNet, ResNet and GoogleNet. The performance between scAOA+ AlexNet and VGG19+k-NN is significant because the margin is equal to 0.5%, which means scAOA allows to eliminate perfectly the irrelevant features.



Concerning the second comparison, scAOA based on AlexNet and ResNet provide a high performance because the margins are 0.95% and 0.69% compared to the best ML classifier, which used SVM as classifier and Combined transform as DCT and DWT [?].

## 7 Conclusion

Deep gender has recently experienced wide use in the field of video surveillance. However, the high number of attributes produced by pre-trained CNNs, prompted us to integrate MHs to select the optimal set of deep learning attributes (AlexNet and ResNet). The majority of metaheuristics suffer from the problem of exploitation. We have solved this problem at the level of the AOA algorithm by integrating trigonometric operators inspired by the SCA algorithm, applied to facial analysis, in particular gender identification. By analyzing the results obtained, we noted that the proposed scAOA algorithm manages to improve the performance of gender classification in terms of accuracy for the two selected datasets, particularly for AlexNet deep features. A great competitiveness was underlined between the proposed scAOA approach and the SCA algorithm in terms of selected attributes and execution time. As a perspective, the self-checking of the parameters of the AOA algorithm can be considered as a new avenue to explore. Also, the process of large datasets and the choice of CNN architecture can be processed in future.

## Declarations

**Conflict of interest** No conflict of interest is declared by all authors.

## References

- 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
- Peimankar A, Puthusserypady S (2021) DENS-ECG: A deep learning approach for ECG signal delineation. *Expert Syst Appl* 165:113911
- 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* 444:92–110
- Peker M (2021) Classification of hyperspectral imagery using a fully complex-valued wavelet neural network with deep convolutional features. *Expert Syst Appl* 173:114708
- 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
- Huynh HT, Nguyen H (2020) Joint age estimation and gender classification of Asian faces using wide ResNet. *SN Comput Sci* 1(5):1–9
- Savchenko AV (2019) Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet. *PeerJ Comput Sci* 5:e197
- Lapuschkin S, Binder A, Muller KR, 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
- Silva DPD (2019) Age and gender classification: a proposed system (Doctoral dissertation)
- Abirami B, Subashini TS, Mahavaishnavi V (2020) Gender and age prediction from real time facial images using CNN. *Mater Today: Proc* 33:4708–4712
- Lin CJ, Li YC, Lin HY (2020) Using convolutional neural networks based on a Taguchi method for face gender recognition. *Electronics* 9(8):1227
- Swaminathan A, Chaba M, Sharma DK, Chaba Y (2020) Gender classification using facial embeddings: a novel approach. *Procedia Comput Sci* 167:2634–2642
- Greco A, Saggese A, Vento M, Vigilante V (2020) Gender recognition in the wild: a robustness evaluation over corrupted images. *J Amb Intell Human Comput* 12(12):10461–72
- Holland JH (1992) Genetic algorithms. *Scientific American* 267(1):66–73
- Storn R, Price K (1997) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Global Optim* 11(4):341–359
- Rudolph G (2000) Evolution strategies. *Evol Comput* 1:81–88
- Bäck T, Hoffmeister F, Schwefel HP (1991) A survey of evolution strategies. In: *Proceedings of the 4th international conference on genetic algorithms*
- Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection (Vol. 1). MIT press
- Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: *MHS'95. Proceedings of the 6th international symposium on micro machine and human science, IEEE*, pp. 39–43
- Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61
- Mirjalili S (2015) The ant lion optimizer. *Adv Eng Softw* 83:80–98
- Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67
- Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017) Salp Swarm algorithm: a bio-inspired optimizer for engineering design problems. *Adv Eng Softw* 114:163–191
- Saremi S, Mirjalili S, Lewis A (2017) Grasshopper optimisation algorithm: theory and application. *Adv Eng Softw* 105:30–47
- 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
- Wang GG, Deb S, Cui Z (2019) Monarch butterfly optimization. *Neural Comput Appl* 31(7):1995–2014
- Hashim FA, Houssein EH, Hussain K, Mabrouk MS, Al-Atabany W (2022) Honey badger algorithm: new metaheuristic algorithm for solving optimization problems. *Math Comput Simul* 192:84–110
- 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
- Jia H, Peng X, Lang C (2021) Remora optimization algorithm. *Expert Syst Appl* 185:115665
- Hashim FA, Hussain K, Houssein EH, Mabrouk MS, Al-Atabany W (2021) Archimedes optimization algorithm: a new

- metaheuristic algorithm for solving optimization problems. *Appl Intell* 51(3):1531–1551
31. Li S, Chen H, Wang M, Heidari AA, Mirjalili S (2020) Slime mould algorithm: a new method for stochastic optimization. *Futur Gener Comput Syst* 111:300–323
  32. Faramarzi A, Heidarinejad M, Stephens B, Mirjalili S (2020) Equilibrium optimizer: a novel optimization algorithm. *Knowl-Based Syst* 191:105190
  33. 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
  34. Kaveh A, Dadras A (2017) A novel meta-heuristic optimization algorithm: thermal exchange optimization. *Adv Eng Softw* 110:69–84
  35. Ahmadianfar I, Heidari AA, Gandomi AH, Chu X, Chen H (2021) RUN beyond the metaphor: an efficient optimization algorithm based on Runge Kutta method. *Expert Syst Appl* 181:115079
  36. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. *Knowl-Based Syst* 96:120–133
  37. 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* 33(17):11027–41
  38. Emary E, Zawbaa HM, Hassanien AE (2016) Binary ant lion approaches for feature selection. *Neurocomputing* 213:54–65
  39. 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, Singapore, pp. 251–272
  40. Al-Tashi Q, Rais HM, Abdulkadir SJ, Mirjalili S, Alhussain H (2020) A review of grey wolf optimizer-based feature selection methods for classification. *Evolutionary machine learning techniques*, pp. 273–286
  41. Mafarja M, Mirjalili S (2018) Whale optimization approaches for wrapper feature selection. *Appl Soft Comput* 62:441–453
  42. Neggaz N, Houssein EH, Hussain K (2020) An efficient henry gas solubility optimization for feature selection. *Expert Syst Appl* 152:113364
  43. Gao Y, Zhou Y, Luo Q (2020) An efficient binary equilibrium optimizer algorithm for feature selection. *IEEE Access* 8:140936–140963
  44. Zakeri A, Hokmabadi A (2019) Efficient feature selection method using real-valued grasshopper optimization algorithm. *Expert Syst Appl* 119:61–72
  45. Taghian S, Nadimi-Shahraki MH (2019) Binary sine cosine algorithms for feature selection from medical data. *arXiv preprint arXiv:1911.07805*
  46. Yıldız BS, Pholdee N, Bureerat S, Erdaş MU, Yıldız AR, Sait SM (2021) Comparison of the political optimization algorithm, the Archimedes optimization algorithm and the Levy flight algorithm for design optimization in industry. *Materials Testing* 63(4):356–359
  47. 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
  48. Desuky AS, Hussain S, Kausar S, Islam MA, El Bakrawy LM (2021) EAOA: an enhanced archimedes optimization algorithm for feature selection in classification. *IEEE Access* 9:120795–120814
  49. Neggaz N, Ewees AA, Abd Elaziz M, Mafarja M (2020) Boosting Salp swarm algorithm by sine cosine algorithm and disrupt operator for feature selection. *Expert Syst Appl* 145:113103
  50. Hussain K, Neggaz N, Zhu W, Houssein EH (2021) An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection. *Expert Syst Appl* 176:114778
  51. Ewees AA, Abd Elaziz M, Al-Qaness MA, Khalil HA, Kim S (2020) Improved artificial bee colony using sine-cosine algorithm for multi-level thresholding image segmentation. *IEEE Access* 8:26304–26315
  52. Duan M, Li K, Yang C, Li K (2018) A hybrid deep learning CNN-ELM for age and gender classification. *Neurocomputing* 275:448–461
  53. Acien A, Morales A, Vera-Rodriguez R, Bartolome I, Fierrez J (2018, November) Measuring the gender and ethnicity bias in deep models for face recognition. In: *Iberoamerican congress on pattern recognition*, Springer, Cham (pp. 584–593)
  54. 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, pp. 7–11
  55. Mane S, Shah G (2019) Facial recognition, expression recognition, and gender identification. In *Data management, analytics and innovation*. Springer, Singapore, pp. 275–290
  56. 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, Cham, pp. 163–175
  57. 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(1):15–29
  58. Surinta O, Khamket T (2019) 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, pp. 1–6
  59. Zhang C, Ding H, Shang Y, Shao Z, Fu X (2018) Gender classification based on multiscale facial fusion feature. *Math Prob Eng*
  60. Ghoghj B, Shouraki SB, Mohammadzade H, Iranmehr E (2018, May) A fusion-based gender recognition method using facial images. In: *Electrical Engineering (ICEE), Iranian conference on*, IEEE, pp. 1493–1498
  61. Chen WS, Jeng RH (2020) A new patch-based LBP with adaptive weights for gender classification of human face. *J Chin Inst Eng* 43(5):451–457
  62. 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, Singapore, pp. 1173–1186
  63. Simanjuntak F, Azzopardi G (2019) Fusion of CNN-and Cosfire-based features with application to gender recognition from face images. In: *Science and information conference*, Springer, Cham, pp. 444–458
  64. Dwivedi N, Singh DK (2019) Review of deep learning techniques for gender classification in images. In: *Harmony search and nature inspired optimization algorithms*, Springer, Singapore, pp. 1089–1099
  65. 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(1):89
  66. Alghaili M, Li Z, Ali HA (2020) Deep feature learning for gender classification with covered/camouflaged faces. *IET Image Proc* 14(15):3957–3964
  67. Zhou Y, Li Z (2019) Facial Eigen-Feature based gender recognition with an improved genetic algorithm. *J Intell Fuzzy Syst* 37(4):4891–4902

68. Neggaz I, Fizazi H (2022) An Intelligent handcrafted feature selection using Archimedes optimization algorithm for facial analysis. *Soft Computing*, 1-30
69. Yao X, Wang X, Wang SH, Zhang YD (2020) A comprehensive survey on convolutional neural network in medical image analysis. *Multimedia Tools and Applications*, 1-45
70. Nagpal C, Dubey SR (2019) A performance evaluation of convolutional neural networks for face anti spoofing. In: 2019 international joint conference on neural networks (IJCNN), IEEE, pp. 1-8
71. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. In: European conference on computer vision, Springer, Cham, pp. 630-645
72. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Farhan L (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J big Data* 8(1):1–74
73. Zhao S, Wang P, Heidari AA, Zhao X, Ma C, Chen H (2021) An enhanced Cauchy mutation grasshopper optimization with trigonometric substitution: engineering design and feature selection. *Eng Comput*, pp. 1-34

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.