Intelligent Automation & Soft Computing

Tech Science Press

DOI: 10.32604/iasc.2023.039440
Article

Advanced Guided Whale Optimization Algorithm for Feature Selection in BlazePose Action Recognition

Motasem S. Alsawadi[1,*], El-Sayed M. El-kenawy[2] and Miguel Rio[1]

[1]Electronic and Electrical Engineering Department, University College London, London, WC1E7JE, England
[2]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt

*Corresponding Author: Motasem S. Alsawadi. Email: motasem.alsawadi.18@ucl.ac.uk
Received: 30 January 2023; Accepted: 24 May 2023; Published: 11 September 2023

Abstract: The BlazePose, which models human body skeletons as spatiotemporal graphs, has achieved fantastic performance in skeleton-based action identification. Skeleton extraction from photos for mobile devices has been made possible by the BlazePose system. A Spatial-Temporal Graph Convolutional Network (STGCN) can then forecast the actions. The Spatial-Temporal Graph Convolutional Network (STGCN) can be improved by simply replacing the skeleton input data with a different set of joints that provide more information about the activity of interest. On the other hand, existing approaches require the user to manually set the graph's topology and then fix it across all input layers and samples. This research shows how to use the Statistical Fractal Search (SFS)-Guided whale optimization algorithm (GWOA). To get the best solution for the GWOA, we adopt the SFS diffusion algorithm, which uses the random walk with a Gaussian distribution method common to growing systems. Continuous values are transformed into binary to apply to the feature-selection problem in conjunction with the BlazePose skeletal topology and stochastic fractal search to construct a novel implementation of the BlazePose topology for action recognition. In our experiments, we employed the Kinetics and the NTU-RGB+D datasets. The achieved actiona accuracy in the X-View is 93.14% and in the X-Sub is 96.74%. In addition, the proposed model performs better in numerous statistical tests such as the Analysis of Variance (ANOVA), Wilcoxon signed-rank test, histogram, and times analysis.

Keywords: BlazePose; metaheuristics; convolutional networks; feature selection; action recognition

1 Introduction

BlazePose is an architecture for human posture prediction using a lightweight convolutional neural network optimized for real-time inference on mobile devices. During the inference process, the neural network generates 33 crucial body points for a single individual and runs at a rate of over 30 frames per second on a Pixel 2 phone [1]. This is appropriate, particularly in real-time use cases such as measuring one's fitness and recognizing one's sign language. In various applications, such as health

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

monitoring, sign language recognition, and gestural control, estimating the human body's position from photographs or video is extremely important. This activity is challenging since there are many poses, degrees of freedom, and obstructions. In the most recent study [2], pose estimation has shown significant progress. The standard method generates heatmaps for every joint while simultaneously adjusting offsets for every position. Although this choice of heatmaps can be used by many people with only a little additional work required, it makes the model for a single individual significantly more complex than what is needed for real-time inference on mobile phones. A specific application scenario and demonstrate considerable speedups to the model with minimal to no loss in quality [3].

In contrast to methods based on heatmaps, regression-based approaches, despite being less computationally intensive and more scalable, attempt to forecast the mean coordinate values. Still, they frequently cannot resolve the ambiguity at the root of the problem. It has been demonstrated in [4] that the stacked hourglass architecture dramatically improves the prediction accuracy, even when using a reduced number of factors. Expanding on this idea, our previous work used an encoder-decoder network architecture to predict heatmaps for all joints, followed by an encoder that regresses straight to the coordinates of all joints.

2 Literature Review

We divided the review into two categories, the first for action recognition and the second for feature selection.

2.1 Convolutional Networks Action Recognition

Primitive methods for recognizing actions performed by a skeleton typically used artificially-created features and took advantage of relative 3D joint rotations and translations. Deep learning introduced new algorithms that can improve robustness and achieve previously unattainable levels of performance, ushering in a new era of innovation in activity recognition [5,6]. Such schemes depend on the skeleton data in numerous ways, including the following: Strategies utilizing RNNs.

Graph Neural Networks, in deep learning, the term "geometric deep learning" refers to all developing techniques that generalize deep learning models to non-Euclidean domains like graphs. The concept of a Graph Neural Network (GNN), was first described in [7]. The hunch that underpins GNNs is that the edges of a graph indicate the links between items or concepts, while the nodes represent the objects or concepts themselves. GNNs process the graph in an iterative manner, each time representing nodes as the outcome of applying a transformation on the features of nodes as well as the features of their neighbors. The author [8] is credited with the original formulation of CNNs on graphs. He adapted convolution to signals by employing a spectral construction in his work. After the success of convolutional neural networks, the idea of convolution was eventually modified such that it could be applied to graph data instead of the grid as shown in Fig. 1.

Computer vision transformers First proposed in [9] as an alternative to recurrent networks, the Transformer has quickly become the dominant neural model for NLP. It was created to solve two major problems: the difficulty in processing very long sequences using LSTM and RNNs, and (ii) the difficulty in processing sentences using standard RNN architectures, typically done sequentially, word by word. The goal of its creation was to solve these problems. The transformer functions similarly to an encoder-decoder but wholly relies on multi-head self-attention. The transformer self-attention method has seen widespread use in recent years, finding applications in several well-established computer vision applications as shown in Fig. 2.

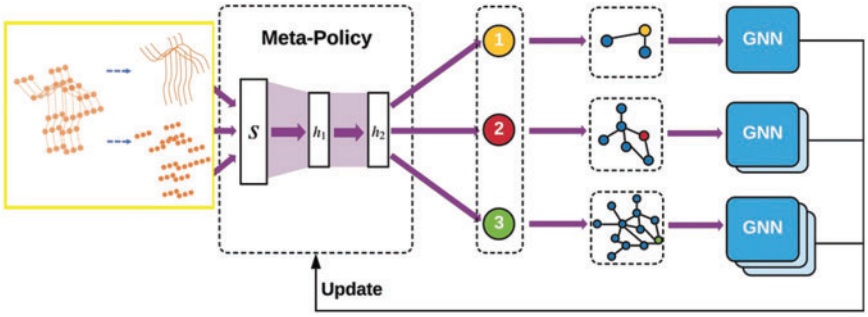

Figure 1: Action recognition process using convolutional neural network

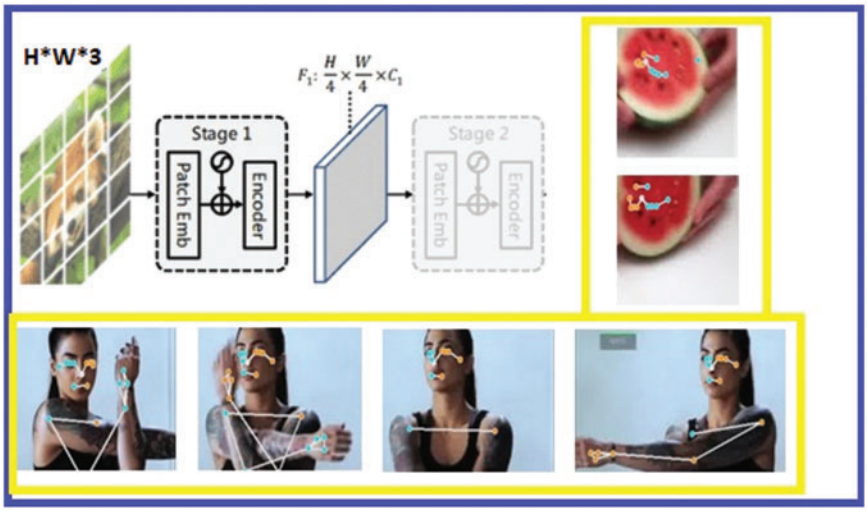

Figure 2: The Pose values detection by the neural network

2.2 Metaheuristics Algorithms for Feature Selection

Determining the best combination of characteristics is difficult and time-consuming to compute. Recently, metaheuristics have been helpful and dependable methods for tackling various optimization issues [10]. Some examples of these challenges include machine learning, data mining difficulties, engineering design, and feature selection. Metaheuristics offer superior performance compared to precise search methods since they do not have to look through the total search space. This frees them from this burden. These search algorithms are only partially functional [11,12]. However, exact approaches are comprehensive and ensure the discovery of the most effective solution to a problem, provided that sufficient time and memory are available. They are ineffective solutions for issues requiring a high level of computing complexity. Recently, scholars, particularly those working on hybrid metaheuristics, have shown significant interest in optimization. For example, the first proposed technique for feature selection utilizing a hybrid metaheuristic and used local search methods in conjunction with the Genetic Algorithm (GA) algorithm. In the research that has been done on Particle Swarm Optimization (PSO) for continuous search space problems, it has been combined with several different metaheuristics. For instance, the PSOGA algorithm was proposed in [13]. This algorithm combines PSO with GA. Other comparable works include a PSO combined with Differential Evolution (DE) (PSODE; referenced in [14]), a hybrid PSO, and Gravitational Search Algorithm (GSA) [15].

In addition, the PSO algorithm was hybridized with the Bacterial Foraging Optimization method to improve the power system's stability [16]. These hybrid strategies try to share the strengths of both of their components to increase the capability of exploitation while simultaneously minimizing the likelihood of falling into an optimal local state. In a similar vein, GWO has received a great deal of attention on the subject of hybrid metaheuristics. For example, the authors in [17] and [18] have combined GWO and DE to test scheduling and continuous optimization. The authors of [19] combined GWO and GA to reduce the value of potential energy functions.

The authors' goal in [20] was to improve the performance of complex systems, where they proposed a hybridization of GWO with Artificial Bee Colony (ABC). GWOSCA, presented in [21] and combines GWO and the Sine Cosine Algorithm, is yet another hybrid technique (SCA). According to this research's findings, the hybrid approaches' performance was significantly superior to that of other global or local search methods. On the subject of feature selection, metaheuristics have also gained a lot of traction in recent years. For example, in [22], a hybrid filter feature selection approach has been proposed. This approach combines SA and GA to improve the searchability of GA. The performance of this approach was tested on eight datasets obtained from UCI, and it received satisfactory results, taking into account the attributes chosen. GA and SA were hybridized in another investigation, and the resulting strain was evaluated based on its ability to hand-print Farsi characters [23]. In addition, a hybrid PSO that utilizes an innovative local search approach founded on information correlation was suggested in [24]. A hybrid genetic algorithm with particle swarm optimization called GPSO was used to pick wrapper features using SVM as a classifier for microarray data [25]. The authors presented a hybrid mutation operator for an enhanced multi-objective PSO [26], published in the same journal as unreliable data. It was proposed in [27] that a hybrid GA with PSO may be used to improve the feature set of digital mammogram datasets. Two hybrids, one utilizing ACO and the other using GA to select features, were proposed in [28] and [29].

Another somewhat similar method can be found in reference [30]. As a feature selector, the authors of [31] utilized a combination of DE and ABC. The authors in [32] proposed a hybrid harmony search method that combines a local stochastic search with their original search for the same objective. Recently, in [33], a hybrid WOA and SA were proposed to select wrapper features. In addition, in [34], there was a proposal for feature selection that utilized a combination of GWO and antlion optimization (ALO). By the (No Free Lunch) theorem, there has never been, there currently is not, and there will never be an optimization method that can handle all problems related to optimization. Despite the excellent performance of the approaches described above, it is safe to say that none of them can handle all of the issues associated with the feature selection process [35–38]. An algorithm's performance may get worse when applied to datasets comparable to the ones it was designed for or datasets of a different sort. As a result, the solutions to problems involving feature selection can be improved by making enhancements to the approaches that are now in use [39–46].

3 The Proposed Methodology

The BlazePose system gives more information about joints than its predecessors and makes tracking more accurate. We think that the BlazePose system's increase in the number of joints in the skeleton compared to other skeleton topologies (like OpenPose) will give us more information to help improve the ST-GCN model's performance. The first difference between the two systems is how they figure out the pose from an image. OpenPose works from the bottom up, while BlazePose works from the top down. The first method identifies the body parts in the image and then maps them to the right person. The second method determines the person's location and then estimates the main joints. We propose a novel skeleton topology that can help improve the performance of the ST-GCN model even more. The goal of the Enhanced-BlazePose topology is to make an even more accurate representation

of the actions by adding more edges to the existing BlazePose topology. By adding feature selection layers with SFS-Guided WOA, we hope to understand how the shoulders and head move together during the activities. For the Kinetics dataset (D1) and the NTU-RGB+D dataset (D2). Fig. 3 shows the proposed topology definition for action recognition.

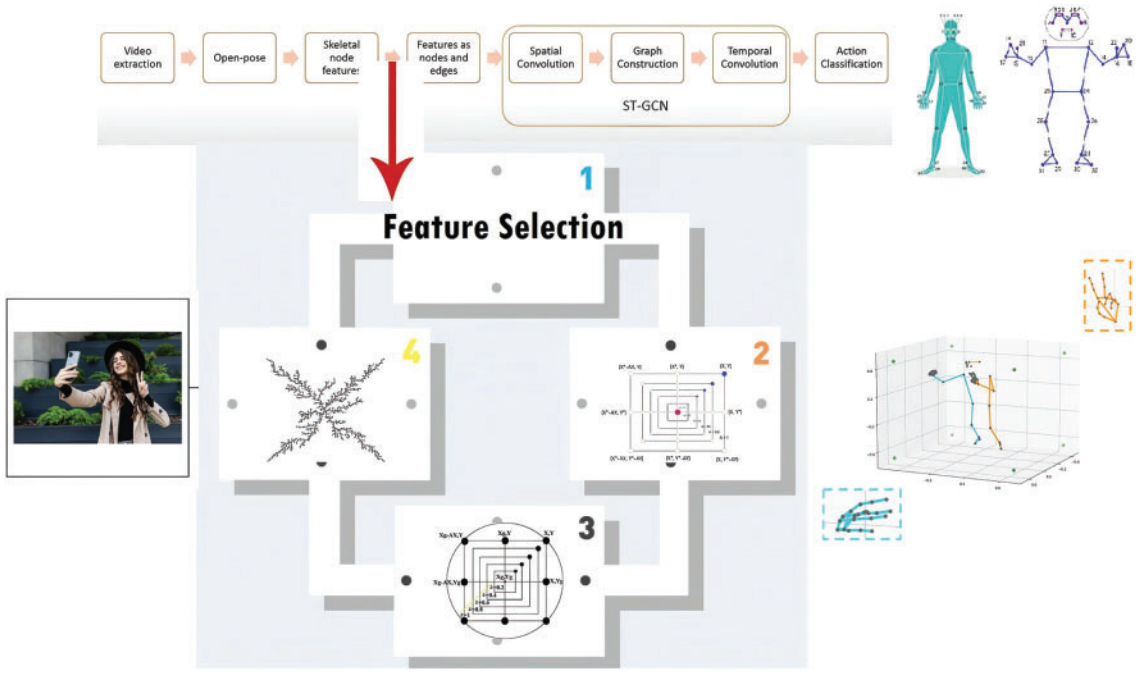

Figure 3: The architecture of the proposed approach

3.1 SFS-Guided WOA for Feature Selection

Inspired optimizers have recently been introduced for feature selection optimization, and these optimizers have been tested for their efficacy and capacity to move optimization problems from local to global optimization. The wrapper-based approach and the filter-based approach, two standard methods for assessing feature quality, are utilized. Using a population-based heuristic random intelligence algorithm, the "whale optimization algorithm" introduces a new type of algorithm. The algorithm's local search ability is improved through a shrinking encircling mechanism and a spiral ascending mechanism, both of which are inspired by the predation behavior of humpback whales. In contrast, its global search ability is improved through a random learning method. It benefits from having few control parameters, easy calculation, and robust optimal solution searching capability. The Guided WOA is an improved version of the original WOA. Like the global search, the early WOA would cause whales to swim aimlessly in circles. However, this method isn't without flaws; for instance, a more sophisticated system can replace a random whale's search strategy and direct the whales to the best possible solution or prey much more quickly. To boost exploration efficiency, the Guided-WOA algorithm lets a whale follow not just one but three random whales. This can encourage whales to expand their range without compromising their authority. The SFS algorithm uses a diffusion method that generates random walks to find the best solution. The answer can serve as the basis for these. This improves the exploratory power of the Guided WOA, and the diffusion process is employed to arrive at the best option. Diffusion around the most up-to-date position includes Gaussian random walks as shown in Fig. 4.

Algorithm: Proposed SFS Guided WOA

Initialize WOA population Gi (i = 1, 2, ..., n), size n, objective function Fn.
    Collection WOA configuration parameters
Initialize GWOA population
    Function Fitness Func (Solution s)
      Fn = h1 Error(O) + h2 |s|/|f|
    End Function
    While t < iter_Num
        Calculate objective function F_n for each agent G_i
    Set Q = best agent position
        $G(t+1) = (o_1 * g\ o_1\ (t) + f * o_2 * (g\ o_2\ (t) - g\ o_3\ (t)) + (1-f) * o_3 * (leader(t) - gr1(t))$
where g o_1, g o_2, g o_3 are three random solutions, o_1 = rand (0, 0.5), o_2 = rand (0, 1), o_3 = rand (0, 1),
  F = decrease exponentially from 1 to 0
In each group, **Update** *the number of solutions*
If the best fitness from the previous 2 iterations did not improve,
        Increase in the *exploration* group solutions number
end if
for each solution in the **exploration** group
        update g o_1, g o_2, g o_3 and Q
    The best solutions were elitism
If $Q <$ any of the best solutions
    mutate the solution by
$$g(t+1) = k + \left(\frac{\sum((g\ o1 + g\ o2 + g\ o3)}{f + iters\_count}\right)$$
$$k = 1 - \frac{2 \times t^2}{(iters\_count)^2}$$
k decreases exponentially from 1 to 0 over the course of iterations,
else
    search around *current* solution
        $G(t+1) = (o_1 * g\ o_1\ (t) + f * o_2 * (g\ o_2\ (t) - g\ o_3\ (t)) + (1-f) * o_3 * (led(t) - g\ o_1\ (t))$
end if
end for
for each solution in the **exploitation** group
  The best solutions were elitism
      update g o_1, g o_2, g o_3 and Q
If $Q <$ any of the best solutions
    move towards the best solution
        $g(t+1) = (o_1 * g\ o_1\ (t) + f * o_2 * (g\ o_2\ (t) - g\ o_3\ (t)) + (1-f) * o_3 * (led(t) - go_1(t))$
else
    search around *the best* solution
$$G(t+1) = f\cos(2\pi) + Gt * \left(1 - \frac{t}{(iters\_count)^2}\right)$$

(Continued)

Figure 4: (Continued)

Algorithm (continued)

Calculate $\mathrm{Gt} = Gaussian\ (\mu, \sigma) + (\rho i - \mathrm{G(t+1)})$
end if
end for
amend solutions
update fitness
end while
return best agent position Q

Figure 4: The SFS Guided WOA algorithm employed in feature selection

We resized each movie till it had the proportions 340 × 256 pixels. This did not consider any video frames in which the BlazePose system did not identify a skeleton as being present. Limiting the number of frames in the series of skeletons to just 300. As a result of this limitation, the majority of videos featured a limited number of frames. Because of this, if a sequence contained fewer than 300 frames, we repeated the first few until we reached the appropriate length. On the other hand, if the sequence contained more than 300 frames, we arbitrarily removed some excess frames. Using spatial configuration partitioning for joint label mapping, the model is trained for 80 epochs.

3.2 Fitness Function

Each solution to the hybrid SFS Guided WOA is evaluated based on how well it meets a fitness function. The fitness function is proportional to the percentage of incorrect classifications and chosen features. Solutions are deemed effective if they reduce the number of features selected while maintaining or improving classification accuracy. The following equation is used to determine how effective each solution is:

$$Fitness = h_1 E(D) + h_2 \frac{|s|}{|f|} \tag{1}$$

where $E(D)$ is the classification error rate for each dimension, s is the number of selected features, f is the number of features and $h_1 \in [0, 1]$, $h_2 = 1 - h_1$ are constants that manage the importance of classification error rate and the number of the selected feature.

4 Results and Discussion

The results are shown in two stages: First, the performance achieved in SFS-Guided WOA with all feature selection performance metrics. The second SFS-Guided WOA, in conjunction with the BlazePose skeletal topology and stochastic fractal search, constructs a novel implementation of the BlazePose topology for action recognition.

4.1 Evaluation Metrics

Classification average error shows how accurate the classifier is given the selected feature set. The classification average error can be calculated in Eq. (2).

$$AvgError = \frac{1}{M} \sum_{j=1}^{M} \frac{1}{N} \sum_{i=1}^{N} mse(C_i, L_i) \tag{2}$$

where M is the number of the run to the algorithm, N is the number of test points, C_i the classifier output label for the point i, L_i is the class label for data point i, and Match is the function that calculates if two inputs are matched or not.

The Best Fitness function is the smallest fitness value for a particular optimizer throughout all M optimization steps. As the most optimistic of the solutions found, the best can be expressed mathematically in Eq. (3).

$$best = min_{i=1}^{M} g_*^i \tag{3}$$

where g_*^i is the optimal solution resulting from a run number i.

To execute an optimization process M times, the worst solution is the one with the worst Fitness. It is possible to express the worst-case scenario mathematically as the Eq. (4).

$$worst = Max_{i=1}^{M} g_*^i \tag{4}$$

The average Fitness size is a measure of the typical proportion of selected features to all available features. The equation for this metric is Eq. (5).

$$AVGSelectionSize = \frac{1}{M} \sum_{i=1}^{M} \frac{size\left(g_*^i\right)}{D} \tag{5}$$

Given a dataset of $size(x)$, the number of features is x, where size(x) is the number of unique elements in the vector x. The mean represents the average of all solutions obtained by an optimization algorithm under Mean M running conditions. The mean performance of any stochastic optimizer can be expressed as a Eq. (6).

$$Mean = \frac{1}{M} \sum_{M}^{1} g_*^i, \tag{6}$$

Standard deviation represents the variation of the best solutions found for running a stochastic optimizer for M different runs. Std is used as an indicator for optimizer stability and robustness. In contrast, Std is smaller, which means that the optimizer always converges to the same solution, while larger values for Std mean many random results. Std is formulated as in Eq. (7).

$$Std = \sqrt{\frac{1}{M-1} \sum (g_1^i - Mean)^2}, \tag{7}$$

where M is the number of times to run the optimization algorithm to select the feature subset, g_1^i is the optimal solution resulting from run number i, and the Mean is the average defined in Eq. (6).

4.2 Feature Selection Results

Table 1 displays the results-based evaluation measures that were used. There are six different evaluation measures shown here in the table. The results of the proposed feature selection algorithm are compared to those of six alternative feature selection methods. It is clear from this table that the proposed strategy achieves higher recorded values than the competing methods.

Table 1: Evaluation metrics of the proposed feature selection with comparison to other methods

	Average error						
Dataset	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
D1	**0.27276**	0.28648	0.2768	0.285465	0.27374	0.2806	0.27374
D2	**0.21014**	0.22314	0.25091	0.230256	0.23596	0.24151	0.23083

(Continued)

Table 1 (continued)

Average select size							
Dataset	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
D1	**0.4167**	0.4807	0.6107	0.4356	0.7057	0.6157	0.5757
D2	**0.21206**	0.34237	0.56661	0.3897	0.40752	0.56358	0.46661
Average fitness							
Dataset	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
D1	**0.30049**	0.33638	0.32279	0.3367	0.3237	0.3305	0.32376
D2	**0.22215**	0.25712	0.28462	0.23	0.2698	0.2753	0.26473
Best fitness							
Dataset	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
D1	**0.20535**	0.22476	0.22476	0.28599	0.2635	0.2635	0.22476
D2	**0.17039**	0.18731	0.22115	0.17711	0.1957	0.1703	0.20423
Worst fitness							
Dataset	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
D1	**0.39988**	0.41888	0.43829	0.36064	0.4382	0.4577	0.39946
D2	**0.32162**	0.33115	0.33962	0.36332	0.3565	0.3734	0.38192
Standard deviation fitness							
Dataset	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
D1	**0.10857**	0.12759	0.13527	0.12759	0.1219	0.1262	0.12604
D2	**0.10867**	0.11822	0.10904	0.15034	0.1180	0.1236	0.12324

On the other hand, the time consumed in feature selection using the proposed approach and other approaches is presented in Table 2. In this table, we can see that the time required by the proposed approach is much lower than the time required by the other methods in selection the significant features necessary for action classification.

Table 2: Average time of the proposed feature selection method with comparison to other methods

Method	Time in seconds for D1	Time in seconds for D2	Average time
SFS-GWOA	31.194	33.612	32.403
bGWO	33.838	35.543	34.6905
bPSO	33.52	35.115	34.3175
bSFS	34.92	34.87	34.895
bWAO	33.327	34.448	33.8875
bFA	34.548	35.132	34.84
bGA	33.794	35.068	34.431

The statistical analysis of the results is presented in Tables 3–8. This analysis is presented in terms of the basic statistical measurements, one-way analysis of variance (ANOVA) and Wilcoxon signed-rank test. The results are compared to the other feature selection methods to confirm the superiority

and stability of the proposed approach. The results presented in these tables confirm the expected findings.

Table 3: Statistical analysis for D1

	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
Number of values	14	14	14	14	14	14	14
Minimum	0.2708	0.2855	0.2768	0.2755	0.2737	0.2806	0.2737
25% Percentile	0.2728	0.2865	0.2768	0.2855	0.2737	0.2806	0.2737
Median	0.2728	0.2865	0.2768	0.2855	0.2737	0.2806	0.2737
75% Percentile	0.2728	0.2865	0.2768	0.2855	0.2737	0.2806	0.2737
Maximum	0.2728	0.2965	0.2968	0.2885	0.2937	0.2906	0.2797
Range	0.002	0.011	0.02	0.013	0.02	0.01	0.006
Mean	0.2725	0.2873	0.2789	0.285	0.2759	0.2818	0.2745
Std. Deviation	0.000578	0.0027	0.00578	0.00284	0.00578	0.00314	0.00200
Std. Error of mean	0.000154	0.0007	0.00154	0.00076	0.00154	0.00084	0.00053
Sum	3.816	4.023	3.905	3.99	3.862	3.945	3.843

Table 4: Statistical analysis for D2

	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
Number of values	14	14	14	14	14	14	14
Minimum	0.2101	0.2201	0.2409	0.2203	0.226	0.2315	0.2208
25% Percentile	0.2101	0.2231	0.2509	0.2303	0.236	0.2415	0.2308
Median	0.2101	0.2231	0.2509	0.2303	0.236	0.2415	0.2308
75% Percentile	0.2101	0.2231	0.2509	0.2303	0.236	0.2415	0.2308
Maximum	0.2121	0.2331	0.2609	0.2403	0.246	0.2615	0.2408
Range	0.002	0.013	0.02	0.02	0.02	0.03	0.02
Mean	0.2104	0.2236	0.2509	0.2303	0.236	0.2429	0.2303
Std. Deviation	0.000578	0.00284	0.00392	0.00392	0.00392	0.00663	0.00484
Std. Error of mean	0.000154	0.00076	0.00104	0.00104	0.00104	0.00177	0.00129
Sum	2.945	3.131	3.513	3.224	3.303	3.401	3.225

Table 5: One-way analysis of variance (ANOVA) Test for D1

ANOVA table	SS	DF	MS	F (DFn, DFd)	*P* value
Treatment (between columns)	0.00256	6	0.00042	F (6, 91) = 30.82	$P < 0.0001$
Residual (within columns)	0.00126	91	1.39E-05		
Total	0.00382	97			

Table 6: One-way analysis of variance (ANOVA) Test for D2

ANOVA table	SS	DF	MS	F (DFn, DFd)	*P* value
Treatment (between columns)	0.01452	6	0.00242	F (6, 91) = 138.8	$P < 0.0001$
Residual (within columns)	0.00158	91	0.0000174		
Total	0.01611	97			

Table 7: Wilcoxon Signed Rank Test for D1

	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
Theoretical median	0	0	0	0	0	0	0
Actual median	0.2728	0.286	0.276	0.285	0.273	0.280	0.273
Number of values	14	14	14	14	14	14	14
Wilcoxon Signed Rank Test							
Sum of signed ranks (W)	105	105	105	105	105	105	105
Sum of positive ranks	105	105	105	105	105	105	105
Sum of negative ranks	0	0	0	0	0	0	0
P value (two tailed)	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact	Exact
P value summary	***	***	***	***	***	***	***
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?							
Discrepancy	0.2728	0.2865	0.276	0.285	0.273	0.280	0.273

Table 8: Wilcoxon Signed Rank Test for D2

	SFS-GWOA	bGWO	bPSO	bSFS	bWAO	bFA	bGA
Theoretical median	0	0	0	0	0	0	0
Actual median	0.2101	0.2231	0.250	0.230	0.236	0.241	0.230
Number of values	14	14	14	14	14	14	14
Wilcoxon signed rank test							
Sum of signed ranks (W)	105	105	105	105	105	105	105
Sum of positive ranks	105	105	105	105	105	105	105
Sum of negative ranks	0	0	0	0	0	0	0
P value (two tailed)	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact	Exact
P value summary	***	***	***	***	***	***	***
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?							
Discrepancy	0.2101	0.2231	0.2509	0.2303	0.236	0.2415	0.2308

Another way to demonstrate the effectiveness of the proposed approach is by visualizing the achieved results. Fig. 5 shows the quartile-quartile (QQ) plot for the achieved results based on the two datasets, D1 and D2. In this figure, it can be noted that the performance of the proposed approach is accurate in classifying the actions.

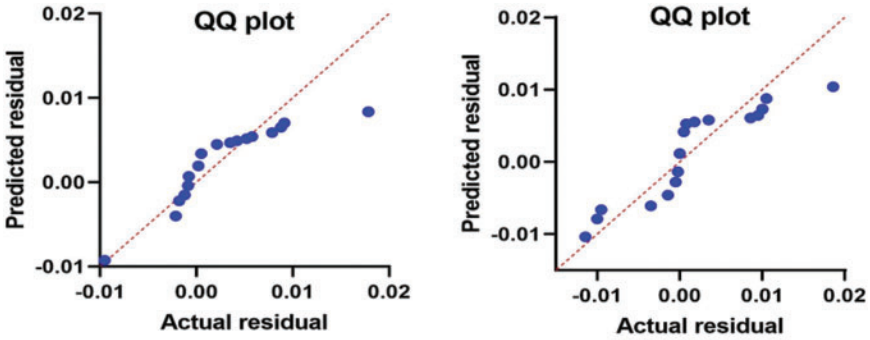

Figure 5: QQ plots. (a) For dataset D1, (b) For Dataset D2

In addition, the plots shown in Fig. 6 depict the values of the objective function for the two datasets using the proposed feature selection algorithm compared to the other six feature selection methods. It can be noted in this figure that the proposed approach is more accurate than the other approaches.

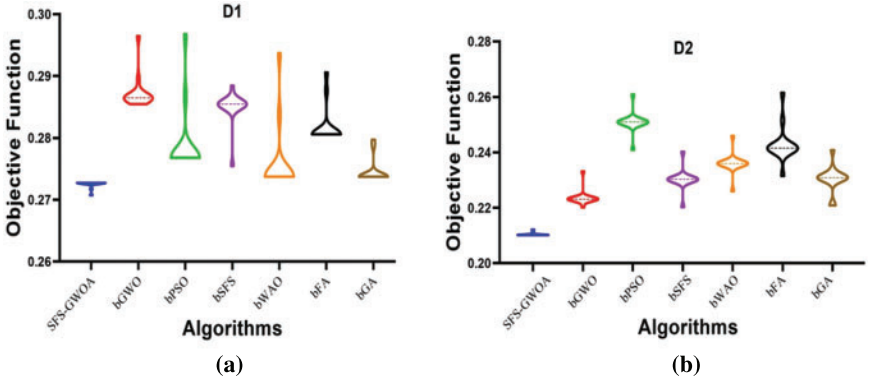

Figure 6: The values of the objective function (a) For dataset D1, (b) For Dataset D2

The histogram of the achieved results is shown in Fig. 7 for the two datasets. This figure is used to study the stability of the proposed approach. This finding is confirmed by the results shown in this figure compared to the other methods for both datasets.

4.3 Action Recognition Results

The action recognition results achieved by the proposed approach are compared to the previous ST-GCN, ST-GDN, and BlazePose methods. The comparison results are presented in Table 9 regarding the Kinetics dataset (D1).

The proposed approach is evaluated using the Cross-Subject (X-Sub) and Cross-View (X-View) criteria suggested by the dataset's developers. Table 10 presents the action recognition results using the proposed approach. In this table, 50% st and 80% st correspond to the results obtained using the

bp-50 and bp-80 subsets (D2). The accuracy achieved by the proposed approach in the case of 50% is 91.33% for X-View and 94.56% for the X-Sub, which are higher than the other methods included in the conducted experiments. In addition, the accuracy of action recognition using the proposed approach is 93.13% and 96.74% for 80% st of the dataset D2. These results confirm the effectiveness of the proposed approach in recognizing human actions.

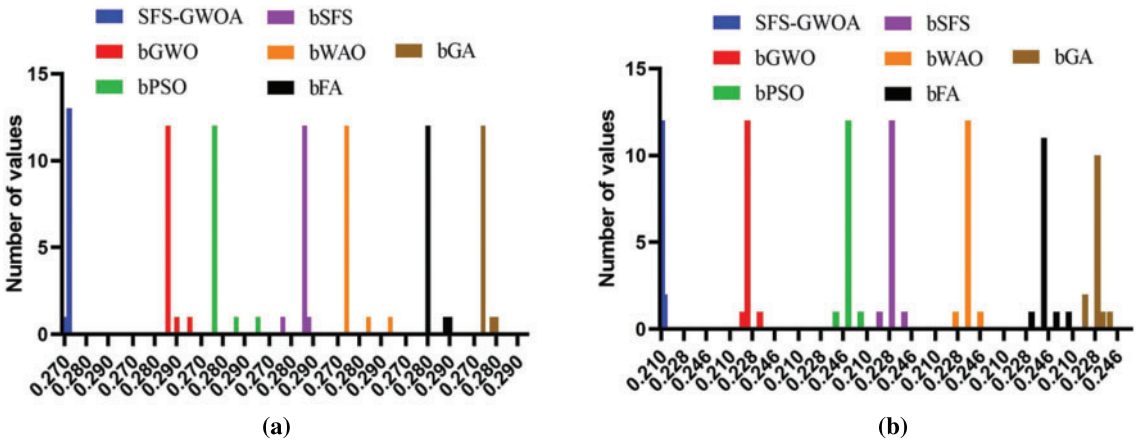

Figure 7: Histogram results. (a) For dataset D1, (b) For Dataset D2

Table 9: Accuracy performance for ST-GCN-based models on the Kinetics dataset (D1). In the table, 50% st and 80% st correspond to the results obtained using the bp-50 and bp-80 subsets, respectively

Method	Top-1	Top-5
ST-GCN	30.70%	52.80%
ST-GDN	37.30%	60.65%
BlazePose, 50% st	36.78%	61.69%
BlazePose, 80% st	37.38%	65.20%
SFS-Guided WOA : BlazePose, 50% st	51.79%	77.13%
SFS-Guided WOA : BlazePose, 80% st	56.87%	81.44%

Table 10: Action recognition accuracy results. In the table, 50% st and 80% st correspond to the results obtained using the bp-50 and bp-80 subsets (D2), respectively

Method	X-view	X-sub
ST-GCN	81.50%	88.30%
ST-GDN	89.70%	95.90%
BlazePose, 50% st	87.30%	90.34%
BlazePose, 80% st	87.62%	91.75%
SFS-Guided WOA : BlazePose, 50% st	91.33%	94.56%
SFS-Guided WOA : BlazePose, 80% st	93.14%	96.74%

5 Conclusion

This study introduces a new action-recognition method by building the BlazePose skeleton topology on top of the ST-GCN architecture and selecting features with SFS-Guided WOA. We have chosen the Kinetics and NTU-RGB+D benchmark datasets to give a reliable basis for comparison with the baseline model in. When the visual data has been acquired in unconstrained contexts, we advocated using alternative skeletal detection criteria to increase the model's performance. We conclude with research that contrasts the suggested strategy with BlazePose, ST-GCN, and ST-GDN. We have demonstrated that BlazePose's topology may be improved by selecting the appropriate features for feet and hands, resulting in more precise data about the motion being captured. In addition, the suggested topology in this research can improve performance even more. The potential drawback of the proposed methodology is the complexity of the proposed feature selection methods. However, this drawback can be verified when applying the proposed methodology on different case studies, which is planned for the future work.

Funding Statement: The authors received no funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Z. Cao, G. Hidalgo, T. Simon, S. Wei and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.

[2] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang *et al.,* "BlazePose: On-device real-time body pose tracking," *arXiv*: 2006.10204, 2020. [Online]. Available: https://arxiv.org/abs/2006.10204

[3] W. Zhang, J. Fang, X. Wang and W. Liu, "EfficientPose: Efficient human pose estimation with neural architecture search," *Computational Visual Media*, vol. 7, no. 3, pp. 335–347, 2021.

[4] L. Shi, Y. Zhang, J. Cheng and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Long Beach, CA, USA, pp. 12026–12035, 2019.

[5] T. Ahsan, S. Khalid, S. Najam, M. A. Khan, Y. J. Kim *et al.,* "Hrneto: Human action recognition using unified deep features optimization framework," *Computers, Materials & Continua*, vol. 75, no. 1, pp. 1089–1105, 2023.

[6] F. Shehzad, M. A. Khan, M. A. E. Yar, M. Sharif, M. Alhaisoni *et al.,* "Two-stream deep learning architecture-based human action recognition," *Computers, Materials & Continua*, vol. 74, no. 3, pp. 5931–5949, 2023.

[7] Y. Liu, R. Ma, H. Li, C. Wang and Y. Tao, "RGB-D human action recognition of deep feature enhancement and fusion using two-stream ConvNet," *Sensors*, vol. 2021, no. 1, pp. 1–10, 2021.

[8] C. Yang, A. Setyoko, H. Tampubolon and K. Hua, "Pairwise adjacency matrix on spatial temporal graph convolution network for skeleton-based two-person interaction recognition," in *2020 IEEE Int. Conf. on Image Processing (ICIP)*, Virtual, pp. 2166–2170, 2020.

[9] M. Alsawadi and M. Rio, "Skeleton-split framework using spatial temporal graph convolutional networks for action recognition," in *2021 4th Int. Conf. Bio-Engineering for Smart Technologies (BioSMART)*, Paris, France, pp. 1–5, 2021.

[10] A. Pauzi, F. Mohd Nazri, S. Sani, A. Bataineh, M. Hisyam *et al.,* "Movement estimation using mediapipe BlazePose," in *7th Int. Visual Informatics Conf.*, Kajang, Malaysia, pp. 562–571, 2021.

[11] I. Nasir, M. Raza, J. Shah, S. Wang, U. Tariq *et al.,* "HAREDNet: A deep learning based architecture for autonomous video surveillance by recognizing human actions," *Computers and Electrical Engineering*, vol. 99, pp. 107805, 2022.
[12] S. Khan, M. Khan, M. Alhaisoni, U. Tariq, H. -S. Yong *et al.,* "Human action recognition: A paradigm of best deep learning features selection and serial based extended fusion," *Sensors*, vol. 21, no. 23, pp. 7941, 2021.
[13] A. Shahroudy, J. Liu, T. Ng and G. Wang, "NTU RGB+D: A large scale dataset for 3D human activity analysis," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 1010–1019, 2016.
[14] M. Skublewska-Paszkowska, P. Powroznik and E. Lukasik, "Learning three dimensional tennis shots using graph convolutional networks," *Sensors*, vol. 20, no. 21, pp. 1–12, 2020.
[15] X. Cao, W. Kudo, C. Ito, M. Shuzo and E. Maeda, "Activity recognition using ST-GCN with 3D motion data," in *Adjunct Proc. of the 2019 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing and Proc. of the 2019 ACM Int. Symp. on Wearable Computers*, London, United Kingdom, pp. 689–692, 2019.
[16] D. Khafaga, A. Alhussan, E. -S. M. El-Kenawy, A. Ibrahim, M. Eid *et al.,* "Solving optimization problems of metamaterial and double T-shape antennas using advanced meta-heuristics algorithms," *IEEE Access*, vol. 10, pp. 74449–74471, 2022.
[17] Y. Galvão, L. Portela, J. Ferreira, P. Barros, O. Fagundes *et al.,* "A framework for anomaly identification applied on fall detection," *IEEE Access*, vol. 9, pp. 77264–77274, 2021.
[18] Y. Jiang, K. Song and J. Wang, "Action recognition based on fusion skeleton of two kinect sensors," in *2020 Int. Conf. on Culture-Oriented Science & Technology (ICCST)*, Beijing, China, pp. 240–244, 2020.
[19] M. Alsawadi and M. Rio, "Skeleton split strategies for spatial temporal graph convolution networks," *Computers, Materials & Continua*, vol. 71, no. 3, pp. 4643–4658, 2022.
[20] D. Sami Khafaga, A. Ali Alhussan, E. M. El-kenawy, A. E. Takieldeen, T. M. Hassan *et al.,* "Meta-heuristics for feature selection and classification in diagnostic breast cancer," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 749–765, 2022.
[21] N. Heidari and A. Iosifidis, "On the spatial attention in spatio-temporal graph convolutional networks for skeleton-based human action recognition," in *2021 Int. Joint Conf. on Neural Networks (IJCNN)*, Virtual, pp. 1–7, 2021.
[22] A. Abdelhamid and S. Alotaibi, "Optimized two-level ensemble model for predicting the parameters of metamaterial antenna," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 917–933, 2022.
[23] U. Bahukhandi and S. Gupta, "Yoga pose detection and classification using machine learning techniques," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 3, no. 12, pp. 186–191, 2021.
[24] D. Neogi, N. Das and S. Deb, "FitNet: A deep neural network driven architecture for real time posture rectification," in *2021 Int. Conf. on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Bahrain, pp. 354–359, 2021.
[25] M. Shams, "Hybrid neural networks in generic biometric system: A survey," *Journal of Artificial Intelligence and Metaheuristics*, vol. 1, no. 1, pp. 20–26, 2022.
[26] A. Takieldeen, E. M. El-kenawy, M. Hadwan and R. M. Zaki, "Dipper throated optimization algorithm for unconstrained function and feature selection," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1465–1481, 2022.
[27] Z. Cao, T. Simon, S. Wei and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, USA, pp. 1302–1310, 2017.
[28] H. Fan, X. Yu, Y. Ding, Y. Yang and M. Kankanhalli, "PSTNET: Point spatio-temporal convolution on point cloud sequences," in *Int. Conf. on Learning Representations*, Virtual, pp. 1–24, 2020.
[29] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran and M. Grundmann, "Blazeface: Sub-millisecond neural face detection on mobile GPUs," arXiv Preprint arXiv1907.05047, 2019.
[30] "MediaPipe," 2022. [Online]. Available: https://mediapipe.dev

[31] A. Abdelhamid and S. R. Alotaibi, "Robust prediction of the bandwidth of metamaterial antenna using deep learning," *Computers, Materials & Continua*, vol. 72, no. 2, pp. 2305–2321, 2022.
[32] A. Ibrahim, S. Mirjalili, M. El-Said, S. M. Ghoneim, M. Al-Harthi *et al.,* "Wind speed ensemble forecasting based on deep learning using adaptive dynamic optimization algorithm," *IEEE Access*, vol. 9, pp. 125787–125804, 2021.
[33] E. El-kenawy, F. Albalawi, S. Ward, S. Ghoneim, M. Eid *et al.,* "Feature selection and classification of transformer faults based on novel meta-heuristic algorithm," *Mathematics*, vol. 10, no. 3144, pp. 1–28, 2022.
[34] D. Khafaga, A. Alhussan, E. El-kenawy, A. Ibrahim, S. Abd Elkhalik *et al.,* "Improved prediction of metamaterial antenna bandwidth using adaptive optimization of LSTM," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 865–881, 2022.
[35] N. Samee, E. El-Kenawy, G. Atteia, M. Jamjoom, A. Ibrahim *et al.,* "Metaheuristic optimization through deep learning classification of COVID-19 in chest X-ray images," *Computers, Materials and Continua*, vol. 73, no. 2, pp. 4193–4210, 2022.
[36] A. Ibrahim, S. Mirjalili, M. El-Said, S. Ghoneim, M. Al-Harthi *et al.,* "Wind speed ensemble forecasting based on deep learning using adaptive dynamic optimization algorithm," *IEEE Access*, vol. 9, pp. 125787–125804, 2021.
[37] G. Yao, X. Liu and T. Lei, "Action recognition with 3D ConvNet-GRU architecture," in *Proc. of the 3rd Int. Conf. on Robotics, Control and Automation*, Chengdu, China, pp. 208–213, 2018.
[38] E. -S. M. El-Kenawy, S. Mirjalili, A. Abdelhamid, A. Ibrahim, N. Khodadadi *et al.,* "Meta-heuristic optimization and keystroke dynamics for authentication of smartphone users," *Mathematics*, vol. 10, no. 16, pp. 1–26, 2022.
[39] O. Nafea, W. Abdul, G. Muhammad and M. Alsulaiman, "Sensor-based human activity recognition with spatio-temporal deep learning," *Sensors*, vol. 21, no. 6, pp. 2141, 2021.
[40] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: Recent updates, datasets, challenges, and applications," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 2259–2322, 2021.
[41] Z. Li, F. Nie, D. Wu, Z. Hu and X. Li, "Unsupervised feature selection with weighted and projected adaptive neighbors," *IEEE Transactions on Cybernetics*, vol. 53, no. 2, pp. 1260–1271, 2023.
[42] F. Nie, Z. Wang, L. Tian, R. Wang and X. Li, "Subspace sparse discriminative feature selection," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 4221–4233, 2022.
[43] M. Khan, Y. Zhang, S. Khan, M. Attique, A. Rehman *et al.,* "A resource conscious human action recognition framework using 26-layered deep convolutional neural network," *Multimedia Tools and Applications*, vol. 80, no. 28, pp. 35827–35849, 2021.
[44] F. Nie, X. Dong, L. Tian, R. Wang and X. Li, "Unsupervised feature selection with constrained $\ell_{2,0}$-norm and optimized graph," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 4, pp. 1702–1713, 2022.
[45] J. Miao, Y. Ping, Z. Chen, X. Jin, P. Li *et al.,* "Unsupervised feature selection by non-convex regularized self-representation," *Expert Systems with Applications*, vol. 173, no. 7, pp. 1–11, 2021.
[46] M. Eid, E. El-Kenawy, N. Khodadadi, S. Mirjalili, E. Khodadadi *et al.,* "Meta-heuristic optimization of LSTM-based deep network for boosting the prediction of monkeypox cases," *Mathematics*, vol. 10, no. 20, pp. 1–20, 2022.