Journal of
Information Systems Engineering
and Business Intelligence

Vol.11, No.2, June 2025
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Boosting Multiverse Optimizer by Simulated Annealing for Dimensionality Reduction

Wamidh K. Mutlag [1], Wamidh Jalil Mazher [2], Hadeel Tariq Ibrahim[3] *, Osman Nuri Ucan[4]

[1] *Southern Technical University, Basra, Iraq*
[1] wamid.almuhaysen@stu.edu.iq

[2] *Departement of Electrical Engineering, Southern Technical University, Basra, Iraq*
[2] wamidh.mazher@stu.edu.iq

[3] *Information Technology Division, College of Education for Women, Shatrah University, Thi-Qar, Iraq*
[3] hadeel.tariq@shu.edu.iq

[4] *Departement of Electric and Computer, The Graduate School of Science and Engineering, Altinbas University, Istanbul, Turkey*
[4] Osman.ucan@altinbas.edu.tr

Abstract

Background: Because of The Multi-Verse Optimizer (MVO) has gained popularity in feature selection due to its strong global and local search capabilities. However, its effectiveness diminishes when tackling high-dimensional datasets due to the exponential growth of the search space and a tendency for premature convergence.
Objective: This study aims to enhance MVO's performance by integrating it with the Simulated Annealing Algorithm (SAA), creating a hybrid model that improves search convergence and optimizes feature selection efficiency.
Methods: A High-level Relay Hybrid (HRH) architecture is proposed, where MVO identifies promising regions of the feature space and passes them to SAA for local refinement. The resulting MVOSA-FS model was evaluated on ten high-dimensional benchmark datasets from the Arizona State University (ASU) repository. Support Vector Machine (SVM) classifiers were used to assess the classification accuracy. MVOSA-FS achieved superior performance compared to six state-of-the-art feature selection algorithms: Atom Search Optimization (ASO), Equilibrium Optimizer (EO), Emperor Penguin Optimizer (EPO), Monarch Butterfly Optimization (MBO), Satin Bowerbird Optimizer (SBO), and Sine Cosine Algorithm (SCA).
Results: The proposed model yielded the lowest average classification error rate (1.45%), smallest standard deviation (0.008), and most compact feature subset (0.91%). The hybrid MVOSA-FS model effectively balances exploration and exploitation, delivering robust and scalable performance in feature selection for high-dimensional data.
Conclusion: This hybridization approach demonstrates improved classification accuracy and reduced computational burden.

Keywords: Multiverse Optimizer, Simulated Annealing, Feature Selection, High-Dimensional Data, Metaheuristics, Hybrid Optimization

Article history: Received 5 September 2024, first decision 3 February 2025, accepted 24 June 2025, available online 22 July 2025

I. Introduction

The dimensionality of input data is a critical factor influencing the performance of machine learning (ML) models in data mining applications. With the rapid advancement of data acquisition technologies, future systems are expected to generate and access increasingly large and high-dimensional datasets [1]. However, processing such voluminous data imposes significant computational demands. Additionally, the inclusion of irrelevant or redundant features—commonly referred to as noisy data—can considerably degrade model performance. These features may mislead the learning process, resulting in reduced accuracy and increased complexity [2].

To address this issue, Feature Selection (FS) has emerged as a vital preprocessing technique. FS aims to identify a subset of the most informative and representative features, thereby enhancing model interpretability and reducing computational burden [3]. Broadly, FS methods are categorized into filter and wrapper approaches. Filter methods rank features based on statistical measures such as distance metrics and mutual information, independently of any

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.11.2.254-266

learning algorithm [4]. Conversely, wrapper methods evaluate subsets of features by assessing their impact on the performance of a specific ML model. Although computationally more intensive, wrapper techniques generally yield superior performance compared to filter methods due to their model-specific optimization [5].

The FS process is thought to be an NP-hard combinatorial optimization problem. A dataset with N features requires the examination of a total of 2^N feature subsets in order to select the best subset [6]. Consequently, searching through every possible subset in quest of the perfect feature subset is not practicable for the high-dimensional dataset.

The purpose of this work is to propose a new hybrid wrapper approach to increase the feature selection task's efficiency in high dimensional datasets by boosting the Multi-Verse Optimizer (MVO) with the SAA. Based on the order in which their functions are used, metaheuristic hybridization paradigms are classified as high or low level [7]. While any metaheuristic function can be replaced by any other algorithm function at the low level, at the high level there is no direct relationship between the inner workings of coupled algorithms (self-contained). There are two processes that can be applied at both low and high levels: relay and teamwork hybridization.

While several cooperating agents boost simultaneously in a collaboration style, relay hybridization applies a set of metaheuristics in a workflow style where the second algorithm uses (inputs) the output of the first [7]. Each agent search in a solutions space. Our goal in this work is to solve feature selection problems using the High-level Relay Hybrid (HRH) layout. In this instance, the population-based MVO algorithm and the single solution-based SAA algorithm will be hybridized, with the SAA algorithm boosting the exploitation in the MVO algorithm. It goes without saying that metaheuristic algorithms, such as MVO, are inappropriate for fine-tuning structures that are extremely close to optimal solutions, in contrast to local search algorithms, such as SAA.

Multi-Verse Optimizer (MVO) is presented by [8]. A new evolving metaheuristic algorithm called MVO imitates the principles of a multiverse theory. This algorithm's primary source of inspiration is the hypothesis of several universes existing and interacting through wormholes, black holes, and white holes. The global optimum for optimization problems with a set of solutions is approximated using this population-based stochastic approach [9]. MVO strength comes from its good convergence speed and good for extensive search space but it suffers from early convergence and several parameters need to setup.

The primary constraints and limitations stem from the no-free-lunch theorem in the search and optimization domain (NFL theorem), which asserts that there isn't a viable optimization algorithm for every type of optimization issue. This implies that while solving optimization problems in the actual world, the multiverse optimizer method can require alterations, corrections, and modifications. As it takes into account that altering parameters (WEP and TDR) tend to approach the near optimal solution, the main disadvantage of the multi-verse optimizer algorithm is its limited ability to handle the complexities of multimodal search methods [10]. Accordingly, it is highly recommended to combine MVO with other algorithms like SAA.

Recently, there has been a lot of attention in the topic of combinatorial optimization for hybrid metaheuristics. Numerous hybrid strategies have been put forth in the literature. For high-dimensional feature selection, [11] provides a two-stage hybrid Ant Colony Optimization (ACO) method (TSHFS-ACO). It calculates the optimal feature subset (OFS) size for the subsequent OFS search using the interval technique. The step of assessing the partial feature number endpoints' performance beforehand helps to lower the algorithm's complexity and prevent it from reaching a local optimum, in contrast to the conventional one-stage methods that calculate the size of OFS and search for OFS concurrently. A combination of two approaches that offers the benefits of excellent standard data accuracy in classification, reasonable feature selection from an extremely large set, and good learning from fewer examples. Ant Lion Optimization (ALO), Grey Wolf Optimization (GWO), and an ALO-GWO combo are the techniques employed. They evaluate their effectiveness using datasets with fewer than 200 instances and nearly 50,000 features [12].

In [13], a hybrid optimization approach that blends Sine-Cosine Algorithm (SCA) in Harris Hawks Optimization (HHO) for computational optimization and feature selection was proposed. In addition to improving exploitation by dynamic modifying candidate solutions to prevent solution stagnation in HHO, the purpose of SCA integration is to address poor exploration in HHO. The tournament selection approach in [14] was used to broaden the starting population's individual variety. The accuracy of classification is assessed using the KNN classifier. Results from experiments on thirteen open medical datasets demonstrate that the suggested BCROSAT works better than other cutting-edge techniques. The goal of [15] was to create a hybrid algorithm based on the Grey Wolf Optimizer (GWO) and Simulated Annealing Algorithm (SAA) that may be used to choose features for biological data. Here, they showed two feature selection techniques (BGWO1-SA and BGWO2-SA). In the final stage of the two aforementioned approaches, the SAA algorithm receives the updated position of the wolves as input, hence intensifying the suggested algorithm even further.

In order to determine if support vector machines (SVMs) are useful for categorizing emails as spam or not, Drucker et al. compared SVMs against three other classification algorithms: boosting decision trees, Rocchio, and Ripper [16]. Two distinct data sets were used to test these four algorithms: one set had features limited to the top 1000 features, and the other set had dimensionality exceeding 7000. SVMs functioned best when binary features were used. In terms of accuracy and speed, boosting trees and SVMs performed satisfactorily on the tests for both data sets [16]. Next, a method to rank individual components based on how much of an impact they have on class assignments was suggested by Hermes & Buhmann [17]. A suitable subset of the features is chosen using this rating. The original feature set is replaced without a discernible drop in classification accuracy. Frequently, the classifier's capacity to generalize even improves because of the implicit regularization that feature selection provide [17].

In this study, six innovative strategies were also modified to validate the effectiveness of the proposed methodology. The main contribution of this paper is proposing a hybrid MVOSA-FS approach to solve high-dimensional FS problems. It is compared to six famous swarm intelligence algorithms that have been modified as ASO [18], EO [19], EPO [20], MBO [21], SBO [22] and SCA [23] for FS applied on 10 high-dimensional datasets using the standard deviation, average FS, and error rates evaluation measures. Support Vector Machines (SVM) [24] is used to realize the effect of the proposed MVOSA-FS based classifier kind.

There are six sections in this article. The details of the suggested method MVOSA-FS and how it is applied to FS in high dimensional datasets are provided in Section 3. The experimental findings are presented in Section 4. Section 5 concludes by summarizing and discussing the research findings.

II. METHODS

A. Multi-Verse Optimizer (MVO)

The Multi-Verse Optimizer (MVO), a unique algorithm motivated by nature, was introduced by [8] This method is primarily inspired by three cosmological concepts: wormholes, black holes, and white holes. These three ideas have mathematical models that are designed to carry out local searching, exploitation, and exploration, in that order. The authors used a roulette wheel mechanism to mathematically describe the white/black hole tunnels and swap the objects of universes. Every time, they use the roulette wheel to select a universe with a white hole by sorting the universes according to their rates of inflation. For this, the actions listed below are taken:

$$UN = \begin{bmatrix} a_1^1 & a_1^2 \dots & a_1^d \\ a_2^1 & a_2^2 \dots & a_2^d \\ a_n^1 & a_n^2 \dots & a_n^d \end{bmatrix} \quad (1)$$

The number of universes (possible solutions) is denoted by n, and the number of parameters (variables) by d:

$$a_x^y = \begin{cases} a_z^y & ir < MI(UNi) \\ a_x^y & ir \geq MI(UNi) \end{cases} \quad (2)$$

Where a_x^y denotes the y[th] variable of x[th] universe, UNi shows the x[th] universe, MI((UNi) is inflation rate normalization for the x[th] universe, $ir1$ is an arbitrary number in [0,1], and a_z^y point to the y[th] parameter of z[th] universe chosen by mechanism of roulette wheel selection. Authors supposed that wormhole pipes are constantly constructed between a universe and the best universe generated up to now in order to achieve the high chance of enhancing the inflation rate via wormholes and to give local modifications for each world. This mechanism is formulated as follows:

$$a_x^y = \begin{cases} \begin{cases} A_y + TDR \times ((ul_y - ll_y \times ir4 + ll_y)\, ir3 < 0.5\,, & ir2 < WEP \\ A_y - TDR \times ((ul_y - ll_y \times ir4 + ll_y)\, ir3 \geq 0.5\,, & ir2 \geq WEP \end{cases} \\ a_x^y \end{cases} \quad (3)$$

Where A_y denotes the y[th] parameter of finest universe designed till now, TDR and WEP are two coefficients, ll_y shows the lower limit of y[th] variable, ul_y is the upper limit of y[th] variable, a_x^y denotes the y[th] parameter of x[th] universe, and $ir4$, $ir3$, $ir4$ are arbitrary numbers in [0,1].

The mathematical formulation suggests that the two primary coefficients involved are the travelling distance rate (TDR) and the wormhole existence probability (WEP). The former coefficient is used to determine the likelihood that wormholes exist in different universes. To highlight exploitation to be the optimization phase moves forward, it must rise linearly throughout the iterations. The rate at which an object can be transported by a wormhole around the most optimal universe discovered to date is determined by its traveling distance, which is also a determining factor. TDR is raised across the iterations in comparison with WEP in order to ensure more accurate exploitation and local search around the optimal produced universe. The following is the adaptive formula for each of the two coefficients:

$$WEP = low + current_{it} \times \left(\frac{high - low}{IT_{high}}\right) \quad (4)$$

Where low is the lowest (0.2 here), $high$ is the highest (1 here), $current_{it}$ shows the current iteration, and IT_{high} denotes the highest iterations.

$$TDR = 1 - \frac{{current_{it}}^{1/xp}}{{IT_{high}}^{1/xp}} \quad (5)$$

Where the exploitation accuracy over the iterations is defined by xp, which in this study equals More xp indicates faster and more precise local search and exploitation. The following observations may help understanding how the suggested algorithm might hypothetically be able to tackle optimization problems, namely the high inflation rate worlds are more likely to form white holes, which can transport items to other universes and help them achieve higher rates of inflation. In addition, low inflationary rates universes are also more probable to contain black holes, which increases the likelihood that these universes will encounter items from other universes. For the universes with low inflation rates, this raises the possibility of improving inflation rates once more. Throughout a number of iterations, the overall/average inflation rate of all universes improves because white/black hole tubes tend to transfer things from universes with high inflation rates to those with low inflation rates. Fig. 1.

B. Simulated Annealing Algorithm (SAA)

A well-researched local search metaheuristic for discrete and, to a lesser extent, continuous optimization issues is called Simulated Annealing Algorithm (SAA). It was introduced by [25]. The main benefit of simulated annealing is that it offers a way to avoid local optima by permitting steps that degrade the value of the objective function, or hill-climbing motions, in the expectation of locating a global optimum. In order to get around the issue of local optima stagnation, SAA uses a particular likelihood to accept a subpar solution.

A randomly produced solution (the initial solution) is the starting point of the algorithm. For every iteration that follows, a neighbor solution to the best solution to date is generated based on a preset neighborhood structure and assessed using a fitness function. A worse neighbor is approved with a probability determined by the Boltzmann probability, $P = e^{-\theta/T}$, where θ is the variance between the fitness of the generated neighbor (TrialSol) and the best solution (BestSol). The improving move, where the neighbor is more fit than the original solution, is always approved. Furthermore, during the search procedure, T is a parameter known as the temperature that varies based on a cooling plan.

C. The Proposed MVOSA-FS Algorithm

The goal of current study is to develop an MVO-based feature selection algorithm that is more potent. One of MVO's issues is that it suffers from early convergence, which makes it difficult to search locally [10]. Thus, the MVO method is paired with SAA algorithm, which is a local search algorithm, in this instance [26]. Because of its emphasis on local searches, SAA algorithm typically discovers workable answers. Additionally, given the directed random process (the low acceptance rate for non-optimal answers), it is capable of passing the local optimal. Consequently, it can be said that our algorithm incorporates the finest aspects of MVO and SAA algorithms. That is, the MVO algorithm's capacity for global search as well as the SAA algorithm's capacity for local search. The SAA algorithm is given the candidate solution findings from MVO in the suggested fashion, allowing it to do a local search around the found solutions. Consequently, we suggest a new method known as MVOSA-FS. MVO is the first step of MVOSA-FS, and SAA is initialized using the solution that MVO has discovered as shown in Fig.1 which represents the 'High level relay hybridization' HRH [7].

A binary optimization problem using just binary 0 and 1 values as solutions is feature selection. Prior to applying the MVO method to the feature selection problem, a binary version of the algorithm is developed. The number of features in the high dimensional datasets determines the overall length of the vector, which is used in this study to represent the outcome as a one-dimensional vector. For each value in the vector, "1" or "0" is used. The value is set to "0" if the right attribute is not selected, and so on. The number "1" indicates that the right attribute was chosen. The number of selected features and classification accuracy of each solution are assessed using the proposed fitness-function, which is based on the SVM classifier. The steps to build MVOSA-FS system are listed below:

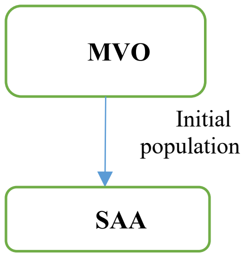

Fig. 1 HRH schema for MVO and SAA

1) Features Normalization

First, we need to normalize the inputted features using a vector of real values. Based on Min-Max normalization [27] features are randomly mapped onto the interval [0,1] using eq. (5). The variable is scaled to a percentage of the whole range of the original dataset by this division. The adjusted value thus lies between 0 and 1. As a result, if the component value is higher or equal to 0.5, it will be replaced by 1 and the feature is selected; if not, the value is calculated to be 0 and the feature is not selected.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

2) Fitness Function

The ideal feature subset is the one with the fewest selected features and the lowest classification error rate. The fitness function is used in MVOSA-FS to evaluate individual searching agents, as seen in the following equation:

$$Fit_{min} = ac_L(DS) + b\frac{|L|}{|O|} \quad (7)$$

Where $ac_L(DS)$ is the error rate which related to the decision on selection DS. The chosen features subset's length is denoted by L, where O is the overall count of datasets' features. The variables a and b are equivalent to the classification quality significance and chosen subset's feature length regarding that $a\epsilon\,[0,1]\; and\; b = 1 - a$ as approved in [28].

3) System Architecture

The MVOSA-FS architecture is shortened in Fig. 2, which reveals the relatives amongst the key portions of the system. As seen, MVOSA-FS begins by collecting datasets and then, as previously mentioned, applies normalization to such datasets. This paper proposes the MVOSA-FS algorithm, which is based on the High-level relay hybridization (HRH) model [7]. The HRH model states that after developing a binary MVO and applying it to find the best solution, the output of MVO is sent into SAA in order to enhance the best feature that was chosen. The MVOSA-FS algorithm's flow diagram was shown in Fig. 2. Such an algorithm's primary stages are divided into three stages.

Stage-0 (Pre-processing) which consists of normalizing data, building training and testing set, and choosing subset of features. Normalizing Data means the earlier feature selection processing activity is made available. Features are normalized to be constrained at the [0,1] phase. Building of training and testing sets includes separation of each dataset into two categories: training and testing. In the binary MVO approach, the training set made up 80% of the overall dataset, while the testing set employed the remaining 20%. To build the model, we ran the training and testing sets through an SVM classifier. Choosing a subset of features means that in this instance, the features with values of 1 from the training set have been chosen.

Stage-1 starts with Fitness assessment which is used to determine the classification performance after the SVM classifier trained using the vectors from the selected training set. The Ending condition, the process has been ended altogether by determining the top iteration. The maximum repeat was really set at 100. The next step, binary MVO algorithm is executed.

Stage-2 involves processing the subset of characteristics in this phase using the SAA method. When SA and MVO are combined by MVOSA-FS, MVO is able to break out of the local optimum and is better able to explore and search during the last stage of development. We receive the last optimum subset of features at the end of this stage.

4) Datasets Specifications

Ten high dimensional benchmark datasets from the ASU repository were employed in this study. These dataset's specifics are shown in TABLE 1. Every test was done with the setting listed in TABLE 2.

TABLE 1
EMPLOYED ASU DATASETS [29]

No	Dataset	No. of features (attributes)	No. of instance
1	CLL-SUB-111	111	11340
2	20newsgroups	171	5748
3	GLA-BRA-180	180	49151
4	GLI-85	85	22283
5	orlraws10P	100	10304
6	pixraw10P	100	10000
7	SMK-CAN-187	187	19993
8	TOX-171	171	5748
9	AR-10P	130	2400
10	PIE10P	210	2420

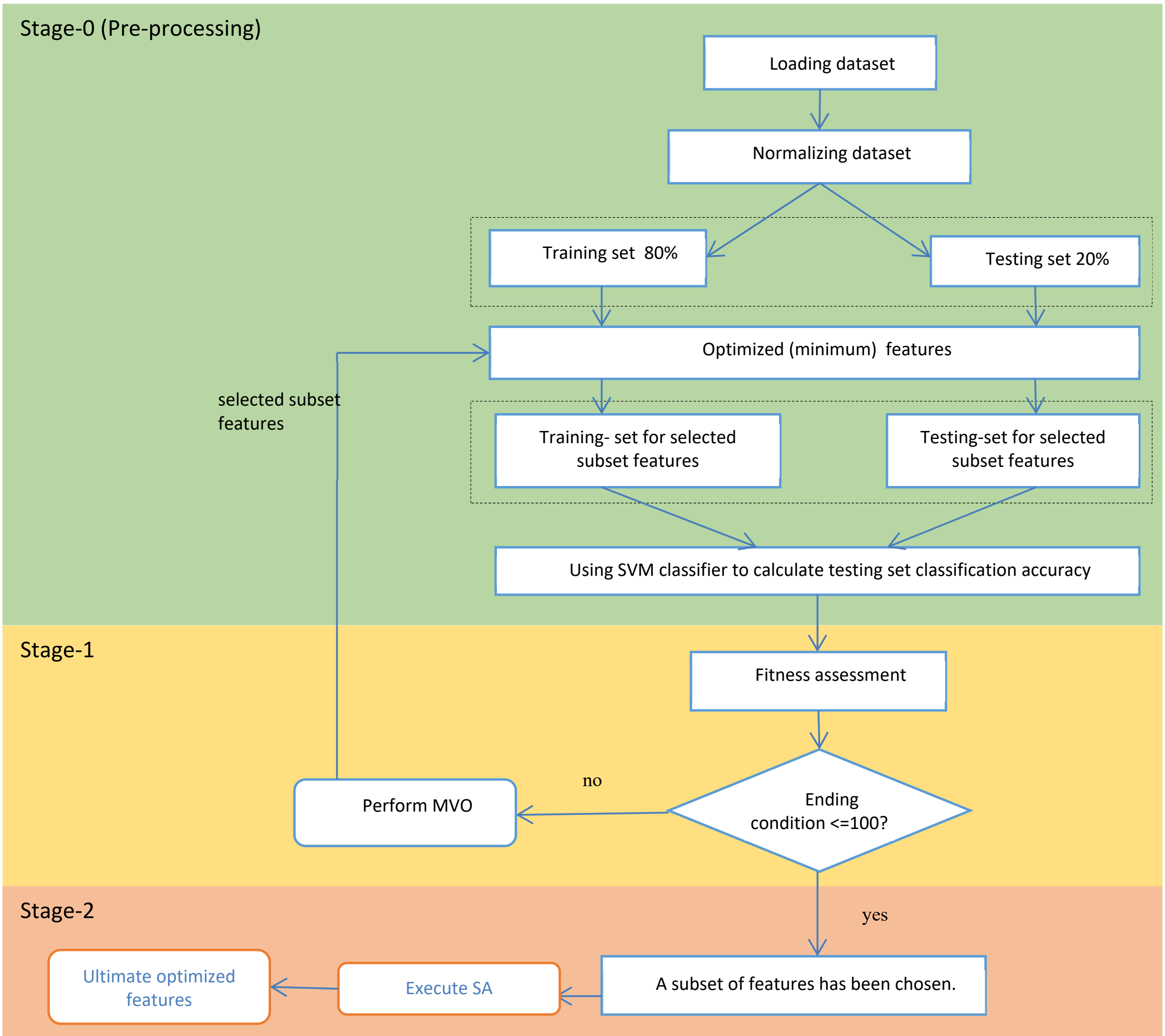

Fig. 2 Proposed MVOSA-FS architecture

D. The Parameters Settings for Algorithms and Experiments

In order to verify and assess the efficacy of the suggested MVOSA-FS algorithm, MVOSA-FS was contrasted with six well-known and contemporary optimization algorithms, such as ASO [18], EO [19], EPO [20], MBO [21], SBO [22] and SCA [23]. Every experiment was run using 10 high dimensional benchmark datasets that were taken from Arizona State University (ASU) repository [29]. The following stages present the used datasets and complete experiment details:

The suggested algorithm's fit performance was verified in each experiment using an SVM classifier based on the wrapper technique. TABLE 2. and TABLE 3. also display the used PC descriptions and the parameter settings for the additional baseline optimization methods, which are MVO, ASO, EO, EPO, MBO, SBO, and SCA, respectively. Furthermore, the population size for each method was set to 10, and the highest number of iterations permitted was 100.

TABLE 2
PC DESCRIPTIONS

Name	Descriptions
CPU	Intel(R) Core(TM) i7-5500U
RAM	2.40 GHz, 8 GB RAM
OS	Windows 10
APPLICATION	MATLAB R2015a

TABLE 3
OPTIMIZATION ALGORITHMS' PARAMETER SET IN USE

Algorithm name	Parameters setting
MVO [8]	$WEP_{minimum}$ = 0.2 $WEP_{minimum}$ = 1 $TDR_{minimum}$ = 0 $TDR_{maximum}$ = 0.6
ASO [18]	α= 50 β=0.4
EO [19]	a1=3 (constant 1) a2=2 (constant 2) GP=0.7 (Generation probability)
EPO [20]	M = 4 (movement parameter)
MBO [21]	Per = 1.2 (migration period) p = 5/12 (ratio) Smax= 1 (maximum step) BAR= 5/12 (butterfly adjusting rate) N1= 4 (number of butterflies in land1)
SBO [22]	α= 0.95 (constant) z= 0.01 (constant) MR = 0.04 (mutation rate)
SCA [23]	α= 2 (constant)

E. Evaluation Metrics

All algorithm's ultimate accuracy in classification is assessed using the well-known Support Vector Machine (SVM) classifier [30], together with wrapper-based feature selection. A three of evaluation metrics is employed here to evaluate various facets of performance:

Classification error rate: to calculate a classifier's error rate using test data is to divide the number of erroneously categorized objects the by total number of items [31].

Standard Deviation (STD): the variable "std" represents the variance of the best results that were achieved after a random optimizer was executed for a number of runs. Std is used to measure the resilience and stability of optimization; lesser values of Std indicate that the algorithm always ends to the same solution, whereas bigger values of Std indicate significantly more irregular performance [32].

Average selected features percentage: the secondary goal of the fitness function that is being employed is average selected features percentage, which is the average ratio of the features that have been chosen to the entire number of features multiplied by 100.

III. RESULTS

Here, we provide an overview and the findings from each experiment. The MVOSA-FS technique is proposed to handle the feature selection problem in high dimensional datasets through three key experiments, as mentioned in last subsection. The suggested adaptive algorithm, or MVOSA-FS, was put into practice on a PC, and its specifications are given in TABLE 2. The algorithms tested on ten publicly available high dimensional datasets included Atom Search Optimization (ASO), Equilibrium Optimizer (EO), Emperor Penguin Optimizer (EPO), Monarch Butterfly Optimization (MBO), Satin Bowerbird Optimizer (SBO), and Sine Cosine Algorithm (SCA).

TABLE 4
COMPARING THE SUGGESTED METHODS DEPENDING ON THE RATE OF CLASSIFICATION ERROR

Datasets	MVOSA-SA	ASO	EO	EPO	MBO	SBO	**SCA**
CLL-SUB-111 (111*11340)	**3.12**	22.728	18.182	13.637	22.728	31.819	9.091
20newsgroups (171*5748)	**2.32**	8.824	8.824	14.706	11.765	17.648	8.824
GLA-BRA-180 (180*49151)	**1.01**	13.889	16.667	13.889	30.556	19.445	25.000
GLI-85 (85*22283)	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	17.648	5.883
orlraws10P (100*10304)	**0.000**	10.000	**0.000**	**0.000**	5.000	10.000	**0.000**
pixraw10P (100*10000)	**0.000**	5.000	**0.000**	**0.000**	**0.000**	5.000	**0.000**
SMK-CAN-187(187*19993)	**1.25**	10.811	18.919	13.514	21.622	16.217	21.622
TOX-171 (171*5748)	**2.54**	11.765	2.942	14.706	11.765	17.648	5.883
AR10P (130*2400)	**4.26**	42.308	30.770	19.231	42.308	30.770	11.539
PIE10P (210*2420)	**0.000**	4.762	7.143	2.381	2.381	9.524	**0.000**
Average error rate	**1.45**	13.01	10.34	9.21	14.81	17.57	8.78

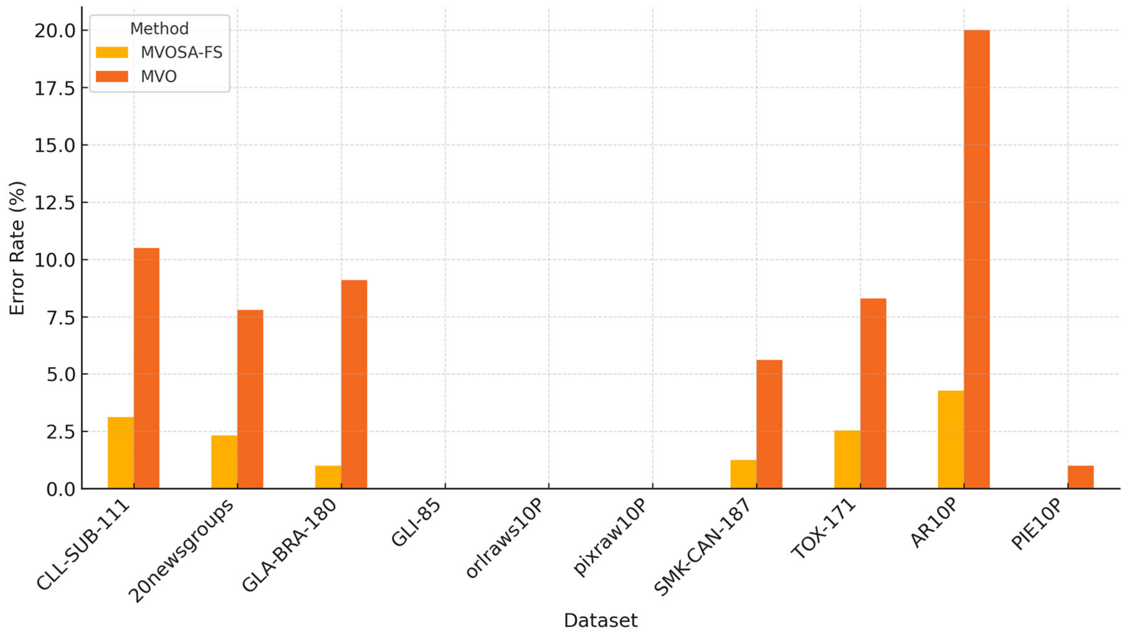

Fig. 3 Comparing the Suggested MVOSA-FS Depending on The Rate of Classification Error

Table 3 lists the parameters that are set in each algorithm. The experiments repeated for 100-iteration and ten search agents in all employed algorithms. Based on the rate of classification error, Table 4 compares the performance of all methodologies. The results in Table 4 show that the proposed MVOSA-FS achieved the best (less) classification error rate in all datasets and lowest average error rate (1.45) while EO, EPO and SCA come next with (0.000) error rates in three datasets, MBO in two datasets and ASO in single dataset. Table 4. compares the classification error rates of the proposed MVOSA-SA model against six benchmark optimization algorithms across ten high-dimensional datasets. MVOSA-SA consistently achieved the lowest error rates in most datasets, including perfect (0%) error in four datasets. The average error rate of MVOSA-SA is the lowest at 1.45%, demonstrating its superior classification accuracy and effectiveness in selecting informative features. Table 4. is visualized in Fig. 3.

The second table assesses the robustness and stability of each algorithm using the standard deviation (STD) of the results. MVOSA-FS again outperforms the others, achieving the lowest average STD value (0.008), which indicates high consistency across multiple runs. Other algorithms show larger variations, especially SBO and ASO, reflecting instability in their performance.

According to the second metric, the standard deviation (STD), MVOSA-FS achieved the lowest STD in 70% of all datasets and less STD rate (0.008), as listed in TABLE 5. and visualized in **Fig. 4**. These outcomes show that the suggested method can manage high-dimensional data collections. Furthermore, for the majority of the data sets, MVOSA-FS displays decreased Std values, confirming the algorithm's robustness.

TABLE 5
COMPARING THE SUGGESTED METHODS DEPENDING ON STANDARD DEVIATION

Datasets	MVOSA-FS	ASO	EO	EPO	MBO	SBO	SCA
CLL-SUB-111 (111*11340)	**0.002**	0.028	0.037	0.044	0.013	0.011	0.073
20newsgroups (171*5748)	**0.007**	0.019	0.061	0.009	0.028	0.010	0.037
GLA-BRA-180 (180*49151)	0.005	0.006	0.025	0.080	0.008	**0.002**	0.038
GLI-85 (85*22283)	0.031	0.033	0.033	0.016	**0.011**	0.730	0.013
orlraws10P (100*10304)	**0.000**	**0.000**	0.001	0.007	0.080	0.047	0.036
pixraw10P (100*10000)	0.011	0.450	0.015	**0.006**	0.770	0.077	0.023
SMK-CAN-187(187*19993)	**0.002**	0.020	0.042	0.034	0.003	0.009	0.015
TOX-171 (171*5748)	**0.011**	0.026	0.030	0.018	0.021	0.024	0.028
AR10P (130*2400)	**0.009**	0.044	0.024	0.061	0.011	0.011	0.047
PIE10P (210*2420)	**0.000**	**0.000**	0.015	0.060	0.002	**0.000**	0.014
Average standard deviation rate	**0.008** ↓	0.063	0.028	0.034	0.095	0.092	0.032

Finally, by employing the third metric, average selected features percentage, we found that MVOSA-FS achieved the lowest FS percentage over 90% all 10 datasets in comparison with the other six state of art algorithms in addition to achieving 0.91% as average FS. The results of current metric are listed in TABLE 6. and visualized in **Fig. 5**. Table 6 displays the number of selected features that each approach obtained through evaluation. The MVOSA-FS approach yields a minimum number of meaningful selected features for all datasets, indicating its high efficiency and suitability for the FS process especially for high dimensional datasets. Suppose a dataset contains **200 features**. After applying the feature selection algorithm, only **20 features** were selected. The percentage of feature selection is calculated as:

$$\textbf{Feature Selection (\%)} = \text{SelectedFeatures} \div \text{TotalFeatures} \times 100 \qquad (7)$$

Equation 7 is used to quantify the effectiveness of a feature selection process. It measures the percentage of original features that have been eliminated after applying a dimensionality reduction or feature selection technique. Here, TotalFeatures refers to the number of features before selection, while SelectedFeatures is the number of features retained. By dividing the selected features by the total, we determine how many were removed. Multiplying this by 100 converts the result into a percentage, indicating the extent of feature reduction achieved. This metric is useful for evaluating how efficiently irrelevant or redundant features have been removed from a dataset.

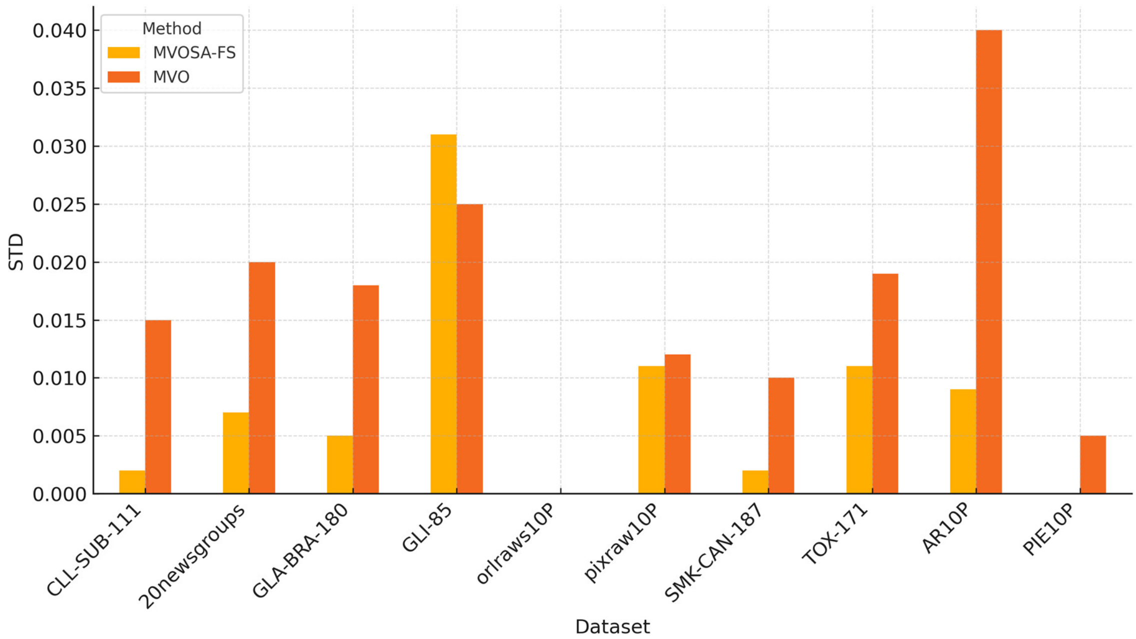

Fig. 4 Comparing the Suggested Methods Depending on The STD

Table 6 measures the percentage of selected features, reflecting how well each method reduces dimensionality. MVOSA-FS exhibits the strongest feature reduction capability, selecting only 0.91% of the features on average—far fewer than any other algorithm. This suggests that MVOSA-FS efficiently identifies the most relevant features, which contributes to both its high accuracy and low computational burden.

TABLE 6
COMPARING THE SUGGESTED METHODS DEPENDING ON THE AVERAGE SELECTED FEATURES PERCENTAGE

Datasets	MVOSA-FS	ASO	EO	EPO	MBO	SBO	SCA
CLL-SUB-111 (111*11340)	**0.40%**	48.40%	11.00%	0.50%	43.30%	48.30%	2.00%
20newsgroups (171*5748)	**1.00%**	48.00%	14.30%	1.30%	44.20%	47.90%	6.00%
GLA-BRA-180 (180*49151)	**1.20%**	48.70%	2.70%	1.80%	43.30%	48.70%	2.10%
GLI-85 (85*22283)	1.77%	49.70%	5.10%	2.00%	43.40%	48.30%	**1.70%**
orlraws10P (100*10304)	**0.10%**	46.80%	0.20%	5.00%	39.30%	46.70%	0.50%
pixraw10P (100*10000)	**0.07%**	46.00%	0.08%	3.00%	39.10%	47.20%	0.20%
SMK-CAN 187(187*19993)	**0.18%**	49.90%	5.60%	0.20%	44.00%	44.00%	2.70%
TOX-171 (171*5748)	**1.01%**	48.20%	17.30%	1.30%	43.90%	48.20%	4.40%
AR10P (130*2400)	**1.80%**	45.70%	6.70%	2.20%	37.70%	46.90%	4.80%
PIE10P (210*2420)	**1.52%**	44.10%	3.10%	6.00%	37.20%	44.70%	1.70%
Average FS percentage	**0.91%** ↓	47.55%	6.61%	2.33%	41.54%	47.09%	2.61%

IV. DISCUSSION

The experimental results clearly demonstrate the superiority of the proposed MVOSA-FS model in handling high-dimensional feature selection tasks. Across ten benchmark datasets, MVOSA-FS consistently achieved the lowest classification error rates, with an average of **1.45%**, outperforming all six baseline algorithms. This indicates the model's strong classification capability and its effective exploitation-exploration balance enabled by combining the global search strength of MVO with the local refinement power of SAA.

In terms of stability, MVOSA-FS recorded the lowest average standard deviation (**0.008**), suggesting that its performance is not only accurate but also consistent across multiple runs. This robustness is crucial in real-world applications where reliability and repeatability are essential, especially in domains like healthcare and finance. Algorithms such as SBO and ASO, while competitive in some instances, showed higher variability, which may limit their dependability in practice.

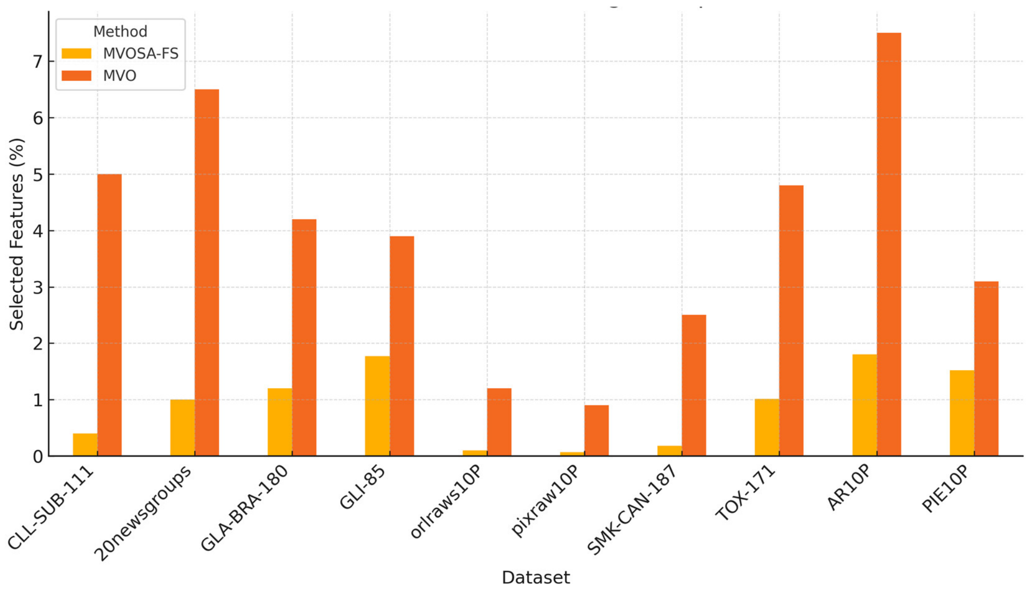

Fig. 5 Comparing the Suggested Methods Depending on Average Selected Features Percentage

Furthermore, MVOSA-FS achieved the **most aggressive feature reduction**, selecting only **0.91%** of the features on average. This significant selection did not compromise classification accuracy, confirming that the selected features were highly informative. The ability to minimize feature sets while maintaining or improving predictive performance is particularly valuable in reducing overfitting, enhancing model interpretability, and lowering computational costs. These results validate the effectiveness of the high-level relay hybridization strategy and position MVOSA-FS as a promising solution for scalable and reliable feature selection in complex, high-dimensional datasets.

The proposed MVOSA-FS demonstrates remarkable superiority in balancing classification accuracy, stability, and dimensionality reduction across diverse datasets. Unlike traditional MVO which suffers from early convergence, the hybrid approach leverages SAA's local search capability to improve exploitation. Compared to related works such as GWO-SA [15] and HHO-SCA [13], MVOSA-FS yields better classification performance with fewer selected features. However, some limitations include parameter sensitivity and lack of real-world time-series dataset validation, which may impact generalizability. Future work could explore adaptive parameter tuning or test on dynamic datasets.

V. Conclusions

One of the most important elements in improving the classifier's performance in the classification problem is feature selection. The suggested method combines the MVO global search with the SA algorithm. Under the high-level relay hybrid model (HRH), SA was used in the suggested methodology. After every MVO iteration, SA was utilized to look for solutions in the vicinity of the best one. Three evaluation criteria are employed to examine various aspects of the performance of comparison algorithms, and the experiments are done on ten high dimensional benchmark datasets from ASU datasets to investigate the performance of the suggested MVOSA-FS technique. The experimental findings demonstrated that the suggested MVOSA-FS technique outperformed the six well-known meta-heuristic algorithms ASO, EO, EPO, MBO, SBO, SCA from current literature in terms of results. The findings demonstrated that the MVOSA-FS produced the lowest error rate with the less classification STD and minimum FS percentage for the majority of datasets when used with SVM as the classifiers. The MVOSA-FS proved to be much more advantageous for comparatively large datasets. We get to the conclusion that the suggested MVOSA-FS technique reduced the number of important features chosen while achieving excellent performance in comparison to the other tested methods. In health care with large datasets, this can facilitate faster and more accurate illness diagnosis and treatment development by doctors, improving the efficiency and effectiveness of a delicate and complex process and ultimately benefiting patients.

Author Contributions: *Wamidh K. Mutlag*: Conceptualization, Formal Analysis and Investigation, Writing – Review & Editing. *Wamidh Jalil Mazher*: Conceptualization, Formal Analysis and Investigation, Writing – Review & Editing. *Hadeel Tariq Ibrahim*: Methodology, Software and Validation, Formal Analysis and Investigation, Writing – Original Draft, Writing – Review & Editing. *Osman Nuri Ucan*: Methodology, Formal Analysis and Investigation, Supervision, Writing – Review & Editing.

All authors have read and agreed to the published version of the manuscript.

Funding: The authors did not receive support from any organization for the submitted work.

Conflicts of Interest: The authors have no conflicts of interest to declare that are relevant to the content of this article.

Data Availability: The data that support the findings of this study are available from ASU [29].

Informed Consent: There were no human subjects.

Institutional Review Board Statement: Not applicable.

Animal Subjects: There were no animal subjects.

ORCID:
Wamidh K. Mutlag: https://orcid.org/0000-0002-6533-3098
Wamidh Jalil Mazher: https://orcid.org/0000-0003-2092-3745
Hadeel Tariq Ibrahim: https://orcid.org/0000-0001-9749-4024
Osman Nuri Ucan: -

References

[1] N. Dupin and E.-G. Talbi, "Machine Learning-Guided Dual Heuristics and New Lower Bounds for the Refueling and Maintenance Planning Problem of Nuclear Power Plants," *Algorithms*, vol. 13, no. 8, p. 185, Jul. 2020, doi: 10.3390/a13080185.

[2] U. Moorthy and U. D. Gandhi, "RETRACTED ARTICLE: A novel optimal feature selection technique for medical data classification using ANOVA based whale optimization," *J Ambient Intell Humaniz Comput*, vol. 12, no. 3, pp. 3527–3538, Mar. 2021, doi: 10.1007/s12652-020-02592-w.

[3] M. Opoku Agyeman, A. F. Guerrero, and Q.-T. Vien, "Classification Techniques for Arrhythmia Patterns Using Convolutional Neural Networks and Internet of Things (IoT) Devices," *IEEE Access*, vol. 10, pp. 87387–87403, Jun. 2022, doi: 10.1109/ACCESS.2022.3192390.

[4] K. Jha and S. Saha, "Incorporation of multimodal multiobjective optimization in designing a filter based feature selection technique," *Appl Soft Comput*, vol. 98, p. 106823, Jan. 2021, doi: 10.1016/j.asoc.2020.106823.

[5] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif Intell*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997, doi: 10.1016/S0004-3702(97)00043-X.

[6] E. Emary and H. M. Zawbaa, "Feature selection via Lèvy Antlion optimization," *Pattern Analysis and Applications*, vol. 22, no. 3, pp. 857–876, Aug. 2019, doi: 10.1007/s10044-018-0695-2.

[7] E.-G. Talbi, "A Taxonomy of Hybrid Metaheuristics," *Journal of Heuristics*, vol. 8, no. 5, pp. 541–564, Sep. 2002, doi: 10.1023/A:1016540724870.

[8] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, "Multi-Verse Optimizer: a nature-inspired algorithm for global optimization," *Neural Comput Appl*, vol. 27, no. 2, pp. 495–513, Feb. 2016, doi: 10.1007/s00521-015-1870-7.

[9] H. Faris, M. A. Hassonah, A. M. Al-Zoubi, S. Mirjalili, and I. Aljarah, "A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture," *Neural Comput Appl*, vol. 30, no. 8, pp. 2355–2369, Oct. 2018, doi: 10.1007/s00521-016-2818-2.

[10] L. Abualigah, "Multi-verse optimizer algorithm: a comprehensive survey of its results, variants, and applications," *Neural Comput Appl*, vol. 32, no. 16, pp. 12381–12401, Aug. 2020, doi: 10.1007/s00521-020-04839-1.

[11] W. Ma, X. Zhou, H. Zhu, L. Li, and L. Jiao, "A two-stage hybrid ant colony optimization for high-dimensional feature selection," *Pattern Recognit*, vol. 116, p. 107933, Aug. 2021, doi: 10.1016/j.patcog.2021.107933.

[12] H. M. Zawbaa, E. Emary, C. Grosan, and V. Snasel, "Large-dimensionality small-instance set feature selection: A hybrid bio-inspired heuristic approach," *Swarm Evol Comput*, vol. 42, pp. 29–42, Oct. 2018, doi: 10.1016/j.swevo.2018.02.021.

[13] K. Hussain, N. Neggaz, W. Zhu, and E. H. Houssein, "An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection," *Expert Syst Appl*, vol. 176, p. 114778, Aug. 2021, doi: 10.1016/j.eswa.2021.114778.

[14] C. Yan, J. Ma, H. Luo, and A. Patel, "Hybrid binary Coral Reefs Optimization algorithm with Simulated Annealing for Feature Selection in high-dimensional biomedical datasets," *Chemometrics and Intelligent Laboratory Systems*, vol. 184, pp. 102–111, Jan. 2019, doi: 10.1016/j.chemolab.2018.11.010.
[15] F. Moeini and S. J. Mousavirad, "An Evolutionary Hybrid Feature Selection Approach for Biomedical Data Classification," in *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, IEEE, Oct. 2020, pp. 623–628. doi: 10.1109/ICCKE50421.2020.9303648.
[16] H. Drucker, Donghui Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *IEEE Trans Neural Netw*, vol. 10, no. 5, pp. 1048–1054, 1999, doi: 10.1109/72.788645.
[17] L. Hermes and J. M. Buhmann, "Feature selection for support vector machines," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, IEEE Comput. Soc, 2000, pp. 712–715. doi: 10.1109/ICPR.2000.906174.
[18] W. Zhao, L. Wang, and Z. Zhang, "Atom search optimization and its application to solve a hydrogeologic parameter estimation problem," *Knowl Based Syst*, vol. 163, pp. 283–304, Jan. 2019, doi: 10.1016/j.knosys.2018.08.030.
[19] A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, "Equilibrium optimizer: A novel optimization algorithm," *Knowl Based Syst*, vol. 191, Jul. 2019, doi: 10.1016/j.knosys.2019.105190.
[20] G. Dhiman and V. Kumar, "Emperor penguin optimizer: A bio-inspired algorithm for engineering problems," *Knowl Based Syst*, vol. 159, pp. 20–50, Nov. 2018, doi: 10.1016/j.knosys.2018.06.001.
[21] G.-G. Wang, S. Deb, and Z. Cui, "Monarch butterfly optimization," *Neural Comput Appl*, vol. 31, no. 7, pp. 1995–2014, Jul. 2019, doi: 10.1007/s00521-015-1923-y.
[22] S. H. Samareh Moosavi and V. Khatibi Bardsiri, "Satin bowerbird optimizer: A new optimization algorithm to optimize ANFIS for software development effort estimation," *Eng Appl Artif Intell*, vol. 60, pp. 1–15, Apr. 2017, doi: 10.1016/j.engappai.2017.01.006.
[23] S. Mirjalili, "SCA: A Sine Cosine Algorithm for solving optimization problems," *Knowl Based Syst*, vol. 96, pp. 120–133, Mar. 2016, doi: 10.1016/j.knosys.2015.12.022.
[24] M. Turchi and N. Cristianini, "Support vector machines," *Wiley Interdiscip Rev Comput Stat*, vol. 1, pp. 283–289, Jul. 2009, doi: 10.1002/wics.49.
[25] S. Kirkpatrick, ; C D Gelatt, and ; M P Vecchi, "Optimization by Simulated Annealing," 1983.
[26] E. Başbüyük, S. P. Brooks, and B. J. T. Morgan, "Optimization Using Simulated Annealing," 1995.
[27] S. Gopal, K. Patro, and K. Kumar Sahu, "Normalization: A Preprocessing Stage." [Online]. Available: www.kiplinger.com,
[28] M. M. Mafarja and S. Mirjalili, "Hybrid Whale Optimization Algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017, doi: 10.1016/j.neucom.2017.04.053.
[29] Dataset Repository, "Datasets | Feature Selection @ ASU."
[30] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature selection for SVMs," in *Advances in Neural Information Processing Systems*, Jul. 2000, pp. 668–674.
[31] Z. Xu, G. Huang, K. Q. Weinberger, and A. X. Zheng, "Gradient boosted feature selection," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 2014, pp. 522–531. doi: 10.1145/2623330.2623635.
[32] L. Jia, W. Gong, and H. Wu, "An Improved Self-adaptive Control Parameter of Differential Evolution for Global Optimization."

Publisher's Note: Publisher stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.