



ARTICLE

A Novel Hybrid Sine Cosine-Flower Pollination Algorithm for Optimized Feature Selection

Sumbul Azeem¹, Shazia Javed^{1,*}, Farheen Ibraheem², Uzma Bashir¹, Nazar Waheed³ and Khursheed Aurangzeb⁴

¹Department of Mathematics, Lahore College For Women University (LCWU), Lahore, Pakistan

²Department of Mathematics, Forman Christian College University (FCCU), Lahore, Pakistan

³Faculty of Computer Information Systems, Higher Colleges of Technology, Abu Dhabi, United Arab Emirates

⁴Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

*Corresponding Author: Shazia Javed. Email: shazia.javed@lcwu.edu.pk

Received: 17 August 2025; Accepted: 22 December 2025; Published: 12 March 2026

ABSTRACT: Data serves as the foundation for training and testing machine learning and artificial intelligence models. The most fundamental part of data is its attributes or features. The feature set size changes from one dataset to another. Only the relevant features contribute meaningfully to classification accuracy. The presence of irrelevant features reduces the system's effectiveness. Classification performance often deteriorates on high-dimensional datasets due to the large search space. Thus, one of the significant obstacles affecting the performance of the learning process in the majority of machine learning and data mining techniques is the dimensionality of the datasets. Feature selection (FS) is an effective preprocessing step in classification tasks. The aim of applying FS is to exclude redundant and unrelated features while retaining the most informative ones to optimize classification capability and compress computational complexity. In this paper, a novel hybrid binary metaheuristic algorithm, termed hSC-FPA, is proposed by hybridizing the Flower Pollination Algorithm (FPA) and the Sine Cosine Algorithm (SCA). Hybridization controls the exploration capacity of SCA and the exploitation behavior of FPA to maintain a balanced search process. SCA guides the global search in the early iterations, while FPA's local pollination refines promising solutions in later stages. A binary conversion mechanism using a threshold function is implemented to handle the discrete nature of the feature selection problem. The functionality of the proposed hSC-FPA is authenticated on fourteen standard datasets from the UCI repository using the K-Nearest Neighbors (K-NN) classifier. Experimental results are benchmarked against the standalone SCA and FPA algorithms. The hSC-FPA consistently achieves higher classification accuracy, selects a more compact feature subset, and demonstrates superior convergence behavior. These findings support the stability and outperformance of the hybrid feature selection method presented.

KEYWORDS: Classification algorithms; feature selection process; flower pollination algorithm; hybrid model; metaheuristics; multi-objective optimization; search algorithm; sine cosine algorithm

1 Introduction

With advances in technology, large amounts of data are generated regularly. This data is derived from real-world applications, such as text classification, spam detection, and medical and engineering fields. This massive amount of data contains a feature set of various sizes. Not all of these features are required for discovering and analyzing knowledge. Most of these features might be irrelevant or redundant to the problem. These unwanted features are not useful for handling classification problems. They decrease the

computational simplicity and increase the classification error [1]. One of the significant preprocessing techniques is dimensionality reduction (DR). The purpose of DR is to reduce the number of features while meeting certain criteria and achieving superior performance. Feature Selection (FS) is one of the most substantial tools in DR. It is the process of choosing the most relevant and effective features to decrease classification error and manage the dimensionality of the data set by discarding unnecessary features [2].

There are four main stages of FS [3]. The first stage is the creation of a feature subset; the second stage is the assessment of the feature subset; the third stage is the establishment of termination criteria; and the fourth stage is validation. During the initial phase, various methods are employed to generate a feature subset using a search strategy. In the second stage, an objective function, also known as a fitness function and a classifier, is used to assess the quality of produced solutions. This procedure is repeated until the stopping criteria are reached. A wrapper-based technique is one of the FS approaches. It uses a specific classifier to find optimal or near-optimal solutions. The objective of FS (Feature Selection) is to find the optimum subset of features. It can be considered as a search procedure. Calculating $2^N - 1$ subsets of a dataset with N features is an NP-hard problem, where NP-hard refers to a class of problems for which no efficient solution algorithm is known. The best subset cannot be obtained using an exhaustive search algorithm (which tries all possible combinations).

Nature-inspired algorithms have validated strong performance compared to classical optimization methods, mainly for nonlinear and complex real-world problems. According to the No Free Lunch (NFL) theorem [4], no single optimization algorithm can outperform all others for every problem. This motivates the ongoing development and compositing of efficacy from known procedures. Hybridizing functions allows the incorporation of complementary strengths from multiple algorithms to be better balanced between exploitation and exploration. The Flower Pollination Algorithm (FPA) [5] works fine for global optimization but can suffer from slow convergence due to a stochastic Lévy flight global search. The Sine Cosine Algorithm (SCA) [6] demonstrates durable exploration, but is limited in exploitation when calibrating into a local node search. Thus, combining the exploration capabilities of SCA and the exploitation of FPA can create a more balanced and efficient search procedure. The FPA component adds a layer to enhance local and global pollination dynamics. Together, these two phases sequentially coordinate to have the algorithm converge quickly. Thus, the design enables the proposed model to achieve higher classification accuracy while obtaining fewer but essential features across the multiple benchmark datasets. In FS, the datasets are typically high-dimensional, with idle features, and many complex interactions among features. These characteristics create difficulty for a single algorithm in achieving effective exploratory and efficient exploitative characteristics. Given these practical challenges, the developed hSC-FPA is inspired to find a balance created for a specific problem. The main contributions of this paper are as follows:

1. A new hybridized method, hSC-FPA, supported by the NFL nature of nature-inspired algorithms and tailored for FS problems, is proposed.
2. To avoid the long random jumps of Lévy flight, the global pollination in FPA is replaced by SCA for non-linear oscillations.
3. To merge the advantages of both SCA and FPA to accurately estimate the Pareto optimal solution by achieving an optimal trade-off between classification accuracy rate and feature reduction.
4. Classification performance is evaluated on 14 datasets and compared with existing methods to demonstrate effectiveness.

The rest of the paper is organized as follows: [Section 2](#) presents a literature review of the hybridization of metaheuristic algorithms. [Section 3](#) discusses the preliminaries. In [Section 4](#), the methodology of the hybrid

model is discussed in detail. [Section 5](#) provides insights into the experimental setup. In [Section 6](#), numerical experimentation and performance comparison of the hSC-FPA with other algorithms are presented. Finally, [Section 7](#) concludes the paper and suggests some future directions.

2 Literature Survey

Mafarja and Mirjalili [7] proposed a hybrid method, HBALO (Hybrid Binary Ant Lion Optimizer), integrating the Binary Ant Lion Optimizer with two incremental hill-climbing techniques, QuickReduct and CEBARKCC, to manage the complexity of FS. A novel hybrid FS algorithm called Mayfly-Harmony Search (MA-HS), combining the Mayfly Algorithm with Harmony Search to improve search space exploitation, was introduced by Bhattacharyya et al. [8]. Results showed the superior performance of MA-HS in selecting informative features and improving classification outcomes. Abualigah and Dulaimi [9] introduced the hybrid Sine Cosine Algorithm and Genetic Algorithm. The algorithm was tested on UCI (University of California, Irvine) datasets using metrics such as classification accuracy, fitness values, and the number of selected features. Mohammadzadeh and Gharehchopogh [10] proposed a hybrid firefly singed Sine Cosine Flower Pollination Algorithm called HWOAFPA. In this process, the Whale Optimization Algorithm was enhanced with opposition-based learning, and its final populations were used to initialize the flower pollination algorithm with improved convergence speed and accuracy. Sun et al. [11] presented a Feature Selection technique known as MetaSCA, which is based on Sine Cosine Algorithm (SCA), but enhances the original SCA by introducing features like the golden sine section coefficient and multi-level tuning factor. Abdelhamid et al. [12] introduced a new binary hybrid metaheuristic algorithm for FS called bSCWDTO, which combined the Sine Cosine Algorithm and Dipper Throated Optimization. Ragab [13] proposed a hybrid binary metaheuristic approach that integrated Particle Swarm Optimization and Firefly Algorithm. It was designed to optimize classification accuracy while minimizing the number of features for FS in machine learning classification problems. Hammouri et al. [14] proposed three improved binary versions of the White Shark Optimizer as a method for FS to address challenges with pre-mature convergence and local optima. The binary versions, Binary Distribution-based White Shark Optimizer (BDWSO), Binary Sine Cosine White Shark Optimizer (BSCWSO), and Binary Hybrid Sine Cosine White Shark Optimizer (BHSCWSO), introduced improvements for FS, including distribution-based updates, adaptive functions from sine-cosine algorithms, and hybrid acceleration strategies. Alkhonaini et al. [15] proposed HSCCOFS-DL, a hybrid deep learning-based method for threat recognition in IoT sensor networks. HSCCOFS-DL employed the Hybrid Sine-Cosine Chimp Optimization Algorithm to perform FS, used a symmetrical autoencoder to perform a classification task, and deployed the Sparrow Search Algorithm to tune the hyperparameters.

Kaur et al. [16] looked forward to applying an automated method to determine quality in Golden Delicious apples with an image processing and feature extraction method. Kaur was able to extract a total of 18,654 features from 1256 images and needed a technique for feature selection. For feature selection, Minimum Redundancy Maximum Relevance was used. Kumari et al. [17] addressed the shortcomings of conventional intrusion detection systems by presenting a hybrid model that successfully combined the Flower Pollination Algorithm, Cheetah Optimization Algorithm, and Artificial Neural Networks to increase accuracy and lower false alarms. Azeem et al. [18] presented a new combination of PSO-HHO to promote FS, advancing the strengths of both metaheuristics to eliminate unnecessary features while improving their classifier performance. A wrapper-based approach using K-Nearest Neighbors (K-NN) with Euclidean distance was employed to evaluate candidate solutions. Aly et al. [19] presented a hybrid metaheuristic algorithm HB-GWO (Hybrid Butterfly-Grey Wolf Optimizer), which combined the Butterfly Optimization Algorithm (BOA) and the Grey Wolf Optimizer (GWO) for effective feature selection in high-dimensional datasets. By using an adaptive switching mechanism to balance exploration and exploitation, HB-GWO

achieves superior performance across benchmark datasets when evaluated with a Random Forest classifier. Experimental studies showed that HB-GWO obtained 96.8%, 80.6%, 89.5%, and 72.4% accuracy on breast cancer, colon cancer, and arrhythmia datasets, respectively.

Most existing hybrid metaheuristics emphasize convergence or exploration but overlook the balance between accuracy and feature reduction. To address these gaps, the proposed hSC-FPA enhances the balance between exploration and exploitation, confirming stable convergence and improved accuracy across various datasets.

3 Preliminaries

This sections provides the FS problem design and a summary of the conventional Sine Cosine Algorithm (SCA) and Flower Pollination Algorithm (FPA).

3.1 Feature Selection (FS) Problem Formulation

Feature Selection (FS) aims to identify the most informative subset of features from a dataset of size $I \times F$, where I denotes the number of samples and F the total number of features. The objective is to reduce dimensionality while also maintaining or increasing classification accuracy. FS is viewed as a multi-objective optimization problem consisting of two conflicting objectives: minimizing the number of features selected and maximizing classification performance. The fitness (cost) function applied to assess the candidate subsets is defined as follows:

$$\text{Fitness} = \omega \times \zeta + (1 - \omega) \times \frac{F'}{F} \quad (1)$$

where $\omega \in [0, 1]$ controls the trade-off between classification error (ζ) and the ratio of selected to total features (F'/F). In this study, ω is set to 0.99 to emphasize accuracy while maintaining acceptable feature reduction. Values close to 0.99 (e.g., 0.95–0.97) produced negligible differences, confirming the robustness of this choice. This formulation directly supports the design of the proposed hybrid hSC-FPA algorithm, which optimizes the above objective through an adaptive balance of exploration and exploitation.

3.2 K-Nearest Neighbors (K-NN)

The K-Nearest Neighbors (K-NN) classifier is used as the evaluation model in the wrapper-based FS framework. Each sample is allocated to a class based on the majority vote among its k nearest neighbors. The Euclidean distance, E_d , between feature vectors from the testing and training data is obtained as:

$$E_d = \sqrt{\sum_{i=1}^{F'} (f_i - f'_i)^2} \quad (2)$$

where f_i and f'_i represent corresponding features in the training and testing sets, respectively. The K-NN classifier controls the optimization process in hSC-FPA by providing feedback for each selected feature subset. It enables the algorithm to iteratively converge toward an optimal balance of accuracy and feature reduction.

4 Materials and Methods

The suggested hybrid hSC-FPA for feature selection is presented here along with its three main steps: initialization, binary conversion, and hybrid optimization. In Algorithm 1, step-by-step instructions together with the underlying mathematics for the algorithm are presented.

4.1 Initialization

Let d denote the number of features, N is the number of candidate solutions, t is the current iteration, and t_{\max} is the maximum iteration. Each solution $\mathbf{x}_i(t)$ represents a position in the search space bounded by the lower limit L and upper limit U . The population is initialized randomly using:

$$\mathbf{x}_i(0) = L + (U - L) \cdot r, \tag{3}$$

where $r \in U(0,1)$ is a uniformly distributed random number and $i = 1, 2, \dots, N$.

4.2 Binary Conversion

The feature selection problem is inherently binary, where each bit represents inclusion (1) or exclusion (0) of a feature. The continuous position $x_{i,d}$ is converted into binary form as:

$$x_{i,d}^{\text{bin}} = \begin{cases} 1, & \text{if } x_{i,d} > 0.5, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

here, $x_{i,d}^{\text{bin}}$ denotes the binary feature selection vector used for fitness evaluation [9].

4.3 Hybrid Sine Cosine-Flower Pollination Algorithm (hSC-FPA)

Let $\mathbf{x}_i(t)$ be the position of the i -th solution at iteration t , and $\mathbf{x}_g(t)$ denotes the global best. Let P be the switch probability, α is the control parameter, ϵ is a small random number, and $r_1, r_2, r_3, r_4 \in [0, 1]$ be random coefficients. The update equations are defined as follows:

If $r < P$ (global update via SCA):

$$\mathbf{x}_i(t+1) = \begin{cases} r_1 \cdot \cos(2\pi r_2) \cdot |r_3 \cdot \mathbf{x}_g(t) - \mathbf{x}_i(t)| + \mathbf{x}_i(t), & \text{if } r_4 \geq 0.5, \\ r_1 \cdot \sin(2\pi r_2) \cdot |r_3 \cdot \mathbf{x}_g(t) - \mathbf{x}_i(t)| + \mathbf{x}_i(t), & \text{otherwise,} \end{cases} \tag{5}$$

where

$$r_1 = \alpha - t \cdot \left(\frac{\alpha}{t_{\max}} \right). \tag{6}$$

If $r \geq P$ (local update via FPA):

$$\mathbf{x}_i(t+1) = \epsilon \cdot (\mathbf{x}_j(t) - \mathbf{x}_k(t)) + \mathbf{x}_i(t), \tag{7}$$

where \mathbf{x}_j and \mathbf{x}_k are two randomly chosen solutions.

The updated positions are binarized as:

$$\mathbf{x}_i^{\text{bin}}(t+1) = \begin{cases} 0, & \text{if } \mathbf{x}_i(t+1) < \text{threshold,} \\ 1, & \text{otherwise.} \end{cases} \tag{8}$$

Fitness is calculated using the predefined cost function:

$$f_i(t+1) = \text{Fun}(\mathbf{x}_i^{\text{bin}}(t+1)), \tag{9}$$

and both local and global bests are updated if an improved solution is found.

The proposed FS process using the hybrid Sine Cosine-Flower Pollination Algorithm (hSC-FPA) is presented in Algorithm 1.

Algorithm 1: Feature selection using hybrid SCA-FPA (hSC-FPA)

Input: Dataset with d features, population size N , max iterations t_{max} , switch probability P , threshold $thresh$

Output: Selected feature subset corresponding to the best binary solution

- 1 Initialize N candidate solutions $\mathbf{x}_i(0)$ using Eq. (3);
- 2 **for** $t = 0$ to $t_{max} - 1$ **do**
- 3 **for** $i = 1$ to N **do**
- 4 Generate random values r, r_1, r_2, r_3, r_4 ;
- 5 **if** $P > r$ **then**
- 6 Execute global update using Eq. (5);
- 7 **else**
- 8 Select two random indices $j \neq k$;
- 9 Perform local update using Eq. (7);
- 10 **end**
- 11 Apply binary conversion using Eq. (8);
- 12 Evaluate fitness using Eq. (9);
- 13 Renew personal best using Eq. (7);
- 14 Renew global best using Eq. (5);
- 15 **end**
- 16 **end**
- 17 **return** Binary feature subset corresponding to $\mathbf{x}_g(t_{max})$

Fig. 1 describes the FS process. The dataset is typically preprocessed to handle outliers and missing values. Secondly, the selection of the best features is applied through an iterative procedure. Then, a subset of features is selected and assessed using specific conditions to determine whether to retain or eliminate them. The final output of this iterative procedure is the best feature subset, which can be evaluated using machine learning classifiers.

4.3.1 Distinctive Features of hSC-FPA

To balance accuracy and search efficiency the suggested hSC-FPA blends the investigation strength of SCA with the exploitation capacity of the FPA. Using this hybrid strategy, SCA's update rule is substituted for the global pollination stage in FPA to improve exploratory search and guard against premature convergence. The goal of the hybridization is to obtain an ideal cooperation between feature reduction and categorization accuracy. The previous hybrid sine cosine flower pollination algorithm [20] was mainly designed for continuous optimization and lacked a feature selection formulation. The suggested hSC-FPA introduces several key improvements for wrapper-based binary feature selection. The proposed methodology includes an explicit binary transfer function and threshold mechanism. The dynamic switching mechanism is based on probability and population updates. This adoptive control avoids premature convergence. The algorithm is practical for real-world data mining and biomedical tasks. Furthermore, the proposed procedure includes repeated runs and statistical tests, offering strong evidence of reliability and fairness.

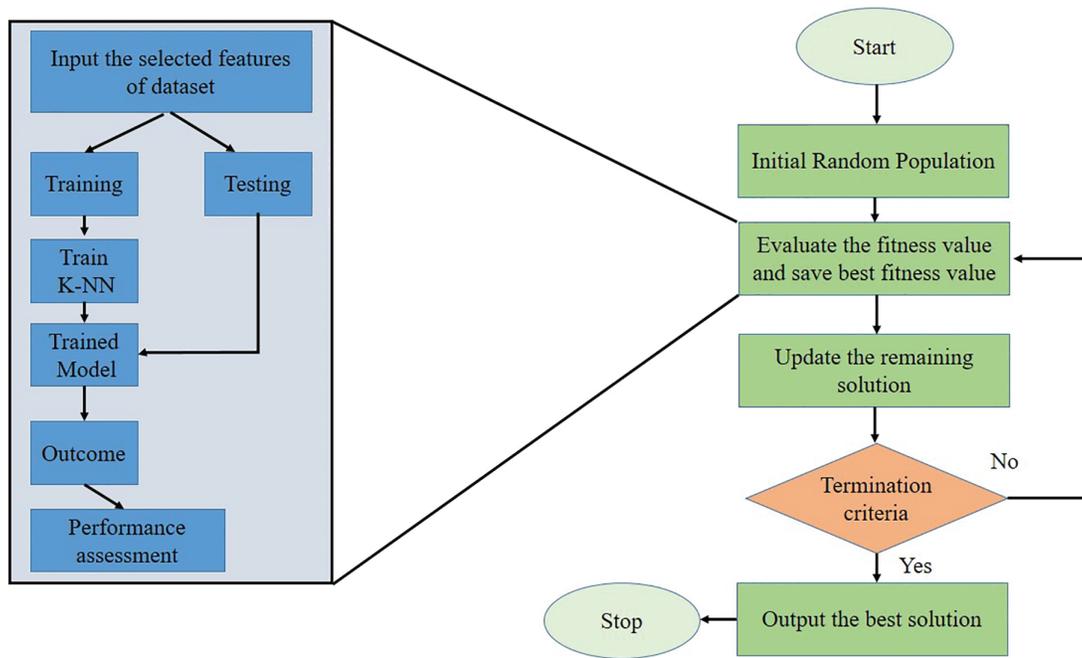


Figure 1: Feature selection procedure of the proposed hSC-FPA model.

4.3.2 Computational Complexity

The computational complexity of the proposed hSC-FPA is mainly determined by the number of search agents N , the problem dimension D (number of features), the maximum number of iterations T , and the cost of evaluating the fitness function. Each iteration per agent includes three chief operations: position update, binary conversion, and fitness evaluation through a wrapper-based K-NN classifier. While the update and binary conversion steps scale linearly with the number of features $O(D)$, the fitness evaluation dominates runtime and scales approximately with the product of the training set size and the number of selected features $O(n_{\text{train}} \cdot D_s)$. Consequently, the per-run complexity of hSC-FPA can be expressed as $O(N \cdot T \cdot (D + n_{\text{train}} \cdot D_s))$. The extra overhead from the hybrid switching mechanism is negligible at $O(N)$ compared to the main terms. This means the algorithm is still similar in performance to standalone SCA and FPA, while providing better search capabilities. When it comes to convergence, a formal math proof is tough for stochastic population-based algorithms because of their randomness. However, empirical analysis shows that hSC-FPA has faster and more stable convergence than the base algorithms.

5 Experimental Setup

5.1 Datasets

Experiments were conducted on fourteen different datasets from various fields to measure the efficiency of the proposed algorithm. The selected datasets were sourced from the UCI repository [21]. Table 1 shows details about the datasets used, including the number of features, sample size, and characteristics.

Table 1: Detail of the utilized datasets.

Dataset	Dataset Type	Class Count	No. of Features	No. of Instances
CNAE	Text classification	9	856	1080
Ecoli	Biological	8	7	336
Ionosphere	Radar signal classification	2	34	351
Iris	Botanical	3	4	150
Libras Movement	Human gesture recognition	15	90	360
Parkinson	Biomedical	2	22	195
Sonar	Physical signals	2	60	208
Soybean Small	Plant disease classification	4	35	47
Spam Base	Text/email spam detection	2	57	4601
SPECT	Biomedical	2	22	267
Tic-Tac-Toe	Game patterns classification	2	9	958
Waveform	Synthetic signals	3	21	5000
Wine	Chemical analysis	3	13	178
Zoo	Zoological	7	16	101

5.2 Parameter Settings

The experimental calculations were done on a Windows 10 Pro PC with an Intel Core i5 CPU running at 2.40 GHz and 8 GB of RAM. We used Python 3.9.12 to carry out the proposed method. To assess the effectiveness of the hSC-FPA approach, we used 14 datasets from the UCI Machine Learning Archive. The data preprocessing phase included cleaning and normalizing to maintain consistent input quality. We handled missing or inconsistent values appropriately. To ensure fair reproducibility, we applied a deterministic random seed based on the run number (`np.random.seed(run)`). This produced distinct but repeatable results across all experimental runs.

We divided the datasets into training and test subsets, each with the same random sizes. During the learning stage, the K-NN classifier was trained using the training subset. The capability of the resultant model was calculated using the testing subset. The experimental setup for hSC-FPA and the competing methodologies are offered in Table 2. The parameter values of P and α used in this study were selected based on settings recommended in [5] and [22], respectively. To ensure consistency and comparability, preliminary sensitivity analysis confirmed that variable key parameters within typical ranges did not significantly change the performance trends. Therefore, the originally recommended values were retained for all experiments.

Table 2: Experimental setup for FPA, SCA, and hSC-FPA algorithms.

Algorithm	Parameter Settings
FPA	Switch Probability (P) = 0.8 [5]
SCA	Scaling Factor (α) = 2 [22]
hSC-FPA	Scaling Factor (α) = 2 Switch Probability (P) = 0.8

(Continued)

Table 2 (continued)

Algorithm	Parameter Settings
Common Parameter Settings	Maximum Number of Iterations (t_{max}) = 100 Inertia Weight (ω) = 0.99 Number of Search Agents (N) = 10 K in K-NN = 5 Training Data = 75% Testing Data = 25% Binary Threshold (θ) = 0.5 Upper Bound (U) = 1 Lower Bound (L) = 0 Initialization = Random Total Number of Runs = 20

5.3 Evaluation Metrics

The attained results were examined based on the criteria given below. These benchmarks were engaged to examine the performance of the suggested FS method.

Let i be the run number, such that $i \in \{1, 2, \dots, 20\}$.

Let $\text{Acc}^{(i)}$ be the classification accuracy obtained in the i^{th} run. Then average accuracy is calculated as:

$$\text{Mean Accuracy} = \frac{1}{20} \sum_{i=1}^{20} \text{Acc}^{(i)} \quad (10)$$

Let $S_F^{(i)}$ denote the number of features selected in the i^{th} run.

$$\text{Mean feature select size} = \frac{1}{20} \sum_{i=1}^{20} S_F^{(i)} \quad (11)$$

6 Results and Discussion

The efficiency of the suggested hSC-FPA algorithm is evaluated using several measurement criteria, including average accuracy and average select size. Classical SCA and FPA are tested along with the proposed hSC-FPA. Furthermore, a comparative analysis was conducted with other hybrid models already proposed in the literature for FS. The proposed optimizer employs a hybrid approach to obtain the optimal solution, which involves maximizing accuracy while maintaining an optimal subset of features.

The following [Table 3](#) presents the classification accuracies of FPA, SCA, and the proposed hybrid hSC-FPA across 14 benchmark datasets. In datasets such as CNAE, Ionosphere, Sonar, and Wine, the hybrid model surpasses FPA and SCA. This shows improved generality and FS capability. On data like Iris, Soybean Small, Ecoli, hSC-FPA performs competitively. The gains are especially pronounced in more complex datasets, such as Sonar and Tic-Tac-Toe, where hSC-FPA shows greater improvements.

[Table 4](#) compares the feature selection sizes from FPA, SCA, and the hybrid hSC-FPA across benchmark datasets. SCA usually generates the fewest features in most datasets. For datasets like Iris, Soybean Small, and Wine, hSC-FPA gives the least number of features. Although hSC-FPA does not always yield the smallest

amount, it often avoids selecting too many features, especially in comparison to FPA. This indicates that the hybrid method is effective for both precision and for getting small, useful feature subsets. Thus, the goal of achieving a Pareto-optimal solution is met, as the proposed hSC-FPA reaches high classification accuracy while balancing accuracy and feature reduction effectively.

Table 3: Classification accuracy comparison.

Dataset	FPA	SCA	hSC-FPA
CNAE	83.28	80.26	86.19
Ecoli	85.24	86.43	85.95
Ionosphere	91.08	94.49	94.94
Iris	98.55	98.68	98.68
Libras Movement	79.61	81.89	82.61
Parkinson	92.04	92.04	92.66
Sonar	88.85	92.02	95.67
Soybean Small	100.00	100.00	100.00
Spam Base	91.72	91.89	93.32
SPECT	79.25	77.54	83.13
Tic-Tac-Toe	81.90	82.71	86.00
Waveform	82.40	81.89	83.77
Wine	96.22	97.22	98.34
Zoo	96.92	96.34	98.46

Note: Bold values indicate the highest classification accuracy achieved for each dataset.

Table 4: Feature count comparison of FPA, SCA, and hybrid hSC-FPA across various datasets.

Dataset	FPA	SCA	hSC-FPA
CNAE	433.25	294.65	325.05
Ecoli	4.45	4.25	5.15
Ionosphere	13.90	3.55	4.45
Iris	1.60	1.90	1.55
Libras Movement	13.45	13.45	21.55
Parkinson	7.95	2.45	2.85
Sonar	27.05	8.35	16.30
Soybean Small	10.05	2.20	2.05
Spam Base	14.85	14.85	22.30
SPECT	10.90	5.00	6.90
Tic-Tac-Toe	6.45	6.95	8.20
Waveform	13.00	11.15	14.00
Wine	4.85	3.65	3.45
Zoo	7.05	4.90	5.40

Note: Bold values represent the lowest number of features selected for each dataset, indicating better feature reduction performance.

The Wilcoxon's rank-sum test was performed at the 0.05 significance level to study the statistical difference between the proposed method and the alternative feature selection methods, FPA and SCA. When

comparing the proposed algorithm's output to other algorithms, this test can help reveal whether or not the results differ significantly. The proposed algorithm's findings differ considerably from those of the compared methods when the p -value is less than 0.05. On the other hand, insignificant results are indicated by a p -value of more than 0.05. The worst p -values in Table 5 are those greater than 0.05, which is the significance level applied. When associating the proposed method with SCA and FPA, the table shows that the p -values obtained with this test are less than 0.05 across datasets such as CNAE, Sonar, Spam base, Tic-tac-toe, and waveform. The results prove the statistical significance of the hSC-FPA method.

Table 5: Wilcoxon signed-rank test p -values comparing hSC-FPA against FPA and SCA across all datasets.

Dataset	FPA	SCA
CNAE	1.06E-07	2.45E-05
Ecoli	6.30E-08	<u>3.37E-01</u>
Ionosphere	<u>4.33E-01</u>	2.87E-06
Iris	<u>3.14E-01</u>	<u>9.52E-01</u>
Libras Movement	<u>5.59E-01</u>	6.45E-03
Parkinson	<u>5.09E-01</u>	<u>6.48E-01</u>
Sonar	6.82E-04	1.75E-04
Soybean Small	0.00E+00	0.00E+00
Spam Base	4.77E-05	3.81E-06
SPECT	2.91E-04	9.06E-04
Tic-Tac-Toe	4.22E-03	4.60E-04
Waveform	3.81E-06	1.91E-06
Wine	<u>7.96E-02</u>	1.68E-02
Zoo	1.93E-02	<u>1.49E-01</u>

Note: $p > 0.05$ are underlined.

The variation in performance of the proposed hSC-FPA algorithm across datasets can be attributed to differences in dataset dimensionality and structure. For high-dimensional datasets such as CNAE and Spam Base, the hybridization mechanism supports effective navigation of a large and sparse feature space. This results in higher accuracy and a substantial reduction in irrelevant features. Conversely, for smaller or low-dimensional datasets such as Iris and Zoo, the search landscape is less complex, and therefore, the advantage of hybrid feature selection is less pronounced. As a result, the performance of hSC-FPA becomes comparable to that of other algorithms, since the reduced dimensionality requires less exploration effort and allows competing methods to achieve similar convergence and classification outcomes. This behavior aligns with the convergence patterns shown in Fig. 2, where high-dimensional datasets exhibit wider initial fluctuations due to exploration, while low-dimensional datasets converge rapidly with minimal variance across iterations.

Table 6 provides a comparative evaluation of the proposed hSC-FPA algorithm relative to other existing hybrid optimization algorithms from the literature.

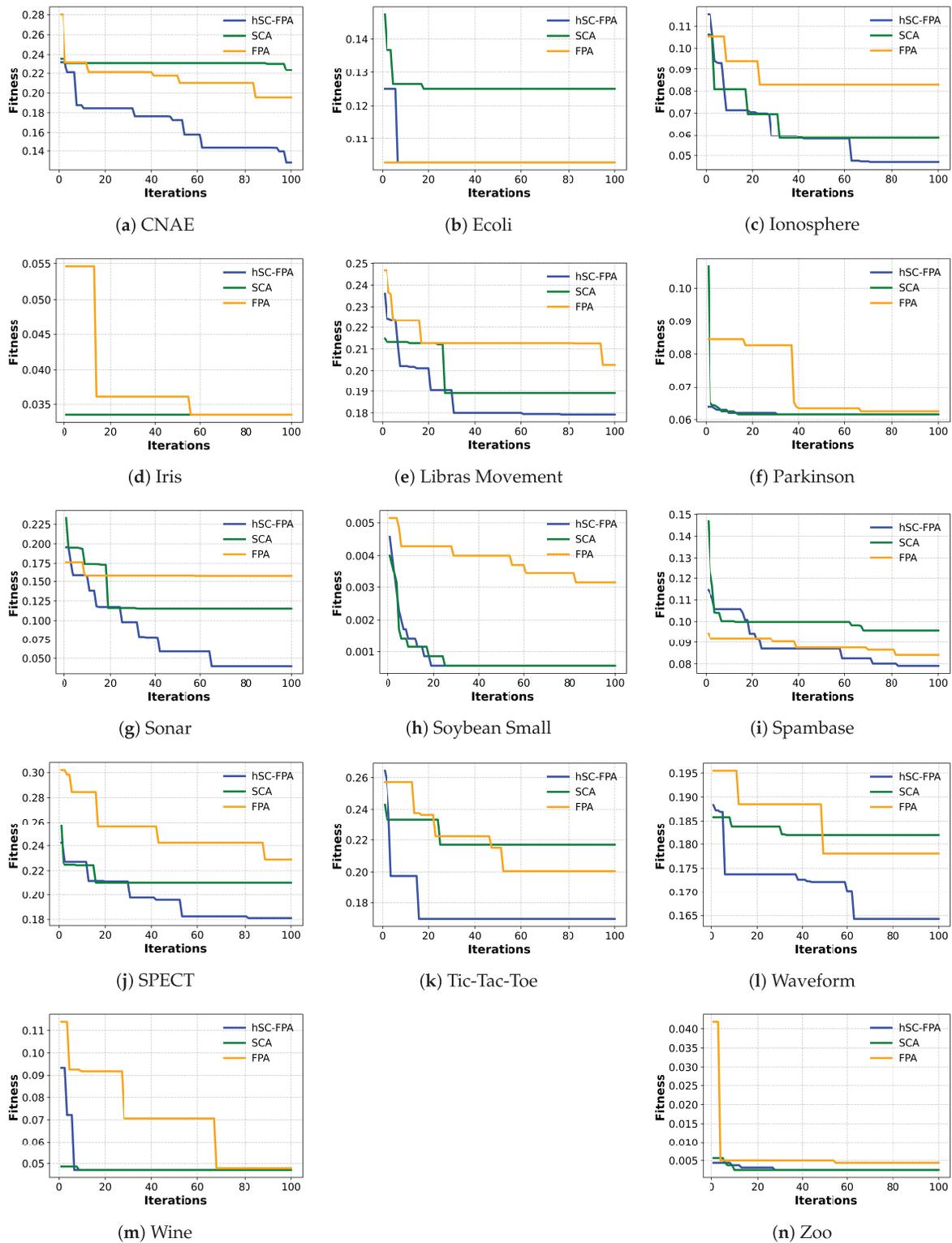


Figure 2: Convergence curves for the FPA, SCA, and hSC-FPA across various datasets. (a)–(n) correspond to individual benchmark datasets, respectively.

Table 6: Classification accuracy (%) comparison of FPA, SCA, hybrid hSC-FPA, and benchmark metaheuristic algorithms across various datasets.

Dataset	hSC-FPA	SCAGA [9]	bSCWDTO [12]	HBWOHH [23]
CNAE	86.19	–	–	–
Ecoli	85.95	–	–	–
Ionosphere	94.94	92.67	44.10	93.00
Iris	98.68	–	–	–
Libras Movement	82.61	–	–	–
Parkinson	92.66	–	59.30	–
Sonar	95.67	92.84	79.10	92.00
Soybean Small	100.00	–	–	–
Spam Base	93.32	–	–	–
SPECT	83.13	84.92	67.60	91.00
Tic-Tac-Toe	86.00	79.59	–	79.00
Waveform	83.77	75.96	49.40	84.00
Wine	98.34	97.92	68.30	100.00
Zoo	98.46	96.75	63.60	97.00

Note: Bold values represent the highest classification accuracy achieved for each dataset. “–” indicates that the result was not reported in the cited study.

7 Conclusion

Implementing feature selection (FS) before the learning stage is crucial for improving classification performance. This approach simplifies the model. This work presents a hybrid method, hSC-FPA, which combines the exploration strengths of SCA with the exploitation abilities of FPA. We use the K-NN classifier for evaluation. In the proposed algorithm, SCA’s global search behavior boosts population diversity, while FPA ensures effective local refinement. The algorithm was tested on fourteen benchmark datasets to evaluate its accuracy and FS effectiveness. Its performance was compared with traditional FPA and SCA. Although the hSC-FPA shows good performance across various benchmark datasets, there are some limitations to consider. As a population-based metaheuristic algorithm, it can sometimes become stuck in local optima, particularly in high-dimensional or complex search spaces. Additionally, hybridizing SCA and FPA adds computational overhead when using single-algorithm strategies, especially with large datasets. Future research will aim to address these limitations by creating adaptive or parallel implementations of hSC-FPA to reduce computation time without losing accuracy. Since the algorithm’s main strength lies in balancing exploration and exploitation, future extensions could explore dynamic parameter control to further enhance this balance. Experiments demonstrated higher classification accuracy with compact feature subsets. As future directions, we suggest utilizing the proposed hybrid optimizer to address other real-world problems, including scheduling issues, engineering optimization problems, and molecular potential energy functions. The presented methodology can be applied to other popular classifiers, such as Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM), and examined to determine whether performance is stable or varies. Another potential future work is to hybridize the FPA with other machine learning and deep learning algorithms.

Acknowledgement: The authors would like to thank the Lahore College for Women University (LCWU), Lahore, Pakistan, for providing computational support and access to resources required for this research.

Funding Statement: This research is supported by a research grant from Lahore College for Women University (LCWU), Lahore, Pakistan.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Sumbul Azeem; methodology, Sumbul Azeem; software, Sumbul Azeem; validation, Sumbul Azeem, Shazia Javed, Farheen Ibraheem and Nazar Waheed; formal analysis, Sumbul Azeem, Farheen Ibraheem and Nazar Waheed; investigation, Sumbul Azeem and Nazar Waheed; resources, Shazia Javed; data curation, Sumbul Azeem; writing—original draft preparation, Sumbul Azeem; writing—review and editing, Sumbul Azeem, Uzma Bashir and Khursheed Aurangzeb; visualization, Uzma Bashir; supervision, Shazia Javed; critical review, Farheen Ibraheem and Khursheed Aurangzeb; project administration, Shazia Javed. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in UCI at [<https://archive.ics.uci.edu/ml>].

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bala I, Karunaratne W, Mitchell L. Optimizing feature selection by enhancing particle swarm optimization with orthogonal initialization and crossover operator. *Comput Mater Contin.* 2025;84(1):727–44. doi:10.32604/cmc.2025.065706.
2. Dokeroglu T, Deniz A, Kiziloz HE. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing.* 2022;494(13):269–96. doi:10.1016/j.neucom.2022.04.083.
3. Liu H, Motoda H. Toward integrating feature selection algorithms for classification and clustering. *IEEE Trans Knowl Data Eng.* 2005;17(4):491–502. doi:10.1109/TKDE.2005.66.
4. Wolpert DH, Macready WG. No free lunch theorems for optimization. *IEEE Trans Evol Comput.* 2002;1(1):67–82. doi:10.1109/4235.585893.
5. Yang XS. Flower pollination algorithm for global optimization. In: *International Conference on Unconventional Computing and Natural Computation.* Berlin/Heidelberg, Germany: Springer; 2012. p. 240–49. doi:10.1007/978-3-642-32894-7_27.
6. Mirjalili S. SCA: a sine cosine algorithm for solving optimization problems. *Knowl-Based Syst.* 2016;96(63):120–33. doi:10.1016/j.knsys.2015.12.022.
7. Mafarja MM, Mirjalili S. Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection. *Soft Comput.* 2019;23(15):6249–65. doi:10.1007/s00500-018-3282-y.
8. Bhattacharyya T, Chatterjee B, Singh PK, Yoon JH, Geem ZW, Sarkar R. Mayfly in harmony: a new hybrid meta-heuristic feature selection algorithm. *IEEE Access.* 2020;8:195929–45. doi:10.1109/ACCESS.2020.3031718.
9. Abualigah L, Dulaimi AJ. A novel feature selection method for data mining tasks using hybrid sine cosine algorithm and genetic algorithm. *Clust Comput.* 2021;24(3):2161–76. doi:10.1007/s10586-021-03254-y.
10. Mohammadzadeh H, Gharehchopogh FS. A novel hybrid whale optimization algorithm with flower pollination algorithm for feature selection: case study email spam detection. *Comput Intell.* 2021;37(1):176–209. doi:10.1111/coin.12397.
11. Sun L, Qin H, Przystupa K, Cui Y, Kochan O, Skowron M, et al. A hybrid feature selection framework using improved sine cosine algorithm with metaheuristic techniques. *Energies.* 2022;15(10):3485. doi:10.3390/en15103485.
12. Abdelhamid AA, El-Kenawy EM, Ibrahim A, Eid MM, Khafaga DS, Alhussan AA, et al. Innovative feature selection method based on hybrid sine cosine and dipper throated optimization algorithms. *IEEE Access.* 2023;11:79750–76. doi:10.1109/ACCESS.2023.3298955.
13. Ragab M. Hybrid firefly particle swarm optimisation algorithm for feature selection problems. *Expert Syst.* 2024;41(7):e13363. doi:10.1111/exsy.13363.

14. Hammouri AI, Braik MS, Al-hiary HH, Abdeen RA. A binary hybrid sine cosine white shark optimizer for feature selection. *Clust Comput.* 2024;27(6):7825–67. doi:10.1007/s10586-024-04361-2.
15. Alkhonaini MA, Al Mazroa A, Aljebreen M, Hassine SBH, Allafi R, Dutta AK, et al. Hybrid sine-cosine chimp optimization based feature selection with deep learning model for threat detection in IoT sensor networks. *Alex Eng J.* 2024;102(1):169–78. doi:10.1016/j.aej.2024.05.051.
16. Kaur S, Sachan MK, Aggarwal AK. Classification of royal delicious apples using hybrid feature selection and feature weighting method based on SVM classifier. *Scalable Comput Pract Exp.* 2025;26(2):940–49. doi:10.12694/scpe.v26i2.3924.
17. Kumari D, Pranav P, Sinha A, Dutta S. A novel approach to intrusion detection system using hybrid flower pollination and cheetah optimization algorithm. *Sci Reports.* 2025;15(1):13071. doi:10.1038/s41598-025-98296-2.
18. Azeem S, Javed S, Naseer I, Ali O, Ghazal TM. A new hybrid PSO-HHO wrapper based optimization for feature selection. *IEEE Access.* 2025;13:87090–9. doi:10.1109/ACCESS.2025.3570901.
19. Aly M, Alotaibi AS. Hybrid butterfly–grey wolf optimization (HB-GWO): a novel metaheuristic approach for feature selection in high-dimensional data. *Stat, Optim & Inform Comput.* 2025;13(6):2575–600. doi:10.19139/soic-2310-5070-2617.
20. Dao TK, Nguyen TT, Nguyen VT, Nguyen TD. A hybridized flower pollination algorithm and its application on microgrid operations planning. *Appl Sci.* 2022;12(13):6487. doi:10.3390/app12136487.
21. Blake CL, Merz CJ. UCI repository of machine learning databases. 1998. [cited 2025 Jul 1]. Available from: <https://archive.ics.uci.edu/ml>.
22. Taghian S, Nadimi-Shahraki MH. Binary sine cosine algorithms for feature selection from medical data. *arXiv:1911.07805.* 2019.
23. Al-Wajih R, Abdulkadir SJ, Aziz N, Al-Tashi Q, Talpur N. Hybrid binary grey wolf with Harris hawks optimizer for feature selection. *IEEE Access.* 2021;9:31662–77. doi:10.1109/ACCESS.2021.3060096.