CISTER

Research Centre in
Real-Time & Embedded
Computing Systems

Journal Paper

Feature selection based on dataset variance optimization using Hybrid Sine Cosine 13 Firehawk Algorithm (HSCFHA)

Syed Kumayl Raza Moosavi

Ahsan Saadat

Zainab Abaid

Wei Ni

Kai Li*

Mohsen Guizani

*CISTER Research Centre

CISTER-TR-240203

2024/02/19

Feature selection based on dataset variance optimization using Hybrid Sine Cosine 13 Firehawk Algorithm (HSCFHA)

Syed Kumayl Raza Moosavi, Ahsan Saadat, Zainab Abaid, Wei Ni, Kai Li*, Mohsen Guizani

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: smoosavi.msai21seecs@seecs.edu.pk, ahsan.saadat@seecs.edu.pk, zainab.abaid@nu.edu.pk, Wei.Ni@data61.csiro.au, kai@isep.ipp.pt

https://www.cister-labs.pt

Abstract

Feature selection plays a pivotal role in preprocessing data for machine learning (ML) models. It entails choosing a subset of pertinent features to enhance the model 19s accuracy and minimize overfitting. Wrapper methods based on metaheuristics are one approach to feature selection, leveraging the predictive accuracy of a learning algorithm to form a condensed set of features. Traditionally, this method uses K-Nearest Neighbor (KNN) for maximizing accuracy as its cost function. However, this approach often yields less than optimal results in large sample spaces and demands considerable computational resources. To circumvent the shortcomings of this approach, this work proposes a novel metaheuristic algorithm, termed the Hybrid Sine Cosine Firehawk Algorithm. Furthermore, a novel feature selection technique is designed that uses this hybrid algorithm to eliminate insignificant and redundant features by incorporating the minimization of dataset variance in the cost function. Additionally, the hybridization of multiple metaheuristic algorithms produces the best features of each algorithm to improve the exploration ability. The proposed technique is tested on 22 University of California Irvine datasets containing low, medium and high dimensional datasets and compared to the traditional KNN-based approach. The technique is also compared with other state-of-the-art metaheuristic techniques, namely Particle Swarm Optimizer, Grey Wolf Optimizer, Whale Optimization Algorithm, Hybrid Ant Colony Optimizer and Improved Binary Bat Algorithm. The results show significant improvements over previous techniques in terms of minimal loss in essential data while reducing the size of the raw data in considerably less time, as well as a well-balanced confusion matrix.

Future Generation Computer Systems 155 (2024) 272–286

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Feature selection based on dataset variance optimization using Hybrid Sine Cosine – Firehawk Algorithm (HSCFHA)

Syed Kumayl Raza Moosavi [a], Ahsan Saadat [a,*], Zainab Abaid [b], Wei Ni [c], Kai Li [d], Mohsen Guizani [e]

[a] National University of Sciences and Technology (NUST), Islamabad, Pakistan
[b] FAST School of Computing, National University of Computer and Emerging Sciences, Islamabad, Pakistan
[c] Data61, Commonwealth Scientific and Industrial Research Organisation, Sydney, Australia
[d] CISTER Research Centre, Portugal
[e] Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
Feature selection
Metaheuristic algorithms
Machine learning
Hybrid Sine Cosine Firehawk Algorithm
Optimization algorithms
Classification
Dataset variance optimization

ABSTRACT

Feature selection plays a pivotal role in preprocessing data for machine learning (ML) models. It entails choosing a subset of pertinent features to enhance the model's accuracy and minimize overfitting. Wrapper methods based on metaheuristics are one approach to feature selection, leveraging the predictive accuracy of a learning algorithm to form a condensed set of features. Traditionally, this method uses K-Nearest Neighbor (KNN) for maximizing accuracy as its cost function. However, this approach often yields less than optimal results in large sample spaces and demands considerable computational resources. To circumvent the shortcomings of this approach, this work proposes a novel metaheuristic algorithm, termed the Hybrid Sine Cosine Firehawk Algorithm. Furthermore, a novel feature selection technique is designed that uses this hybrid algorithm to eliminate insignificant and redundant features by incorporating the minimization of dataset variance in the cost function. Additionally, the hybridization of multiple metaheuristic algorithms produces the best features of each algorithm to improve the exploration ability. The proposed technique is tested on 22 University of California Irvine datasets containing low, medium and high dimensional datasets and compared to the traditional KNN-based approach. The technique is also compared with other state-of-the-art metaheuristic techniques, namely Particle Swarm Optimizer, Grey Wolf Optimizer, Whale Optimization Algorithm, Hybrid Ant Colony Optimizer and Improved Binary Bat Algorithm. The results show significant improvements over previous techniques in terms of minimal loss in essential data while reducing the size of the raw data in considerably less time, as well as a well-balanced confusion matrix.

1. Introduction

Technological advancements have generated large amounts of data in various fields, such as healthcare, finance, and social media. With the transformation of manual jobs into automated solutions, approximately 2.5 quintillion bytes of data are being created in cyberspace every day [1]. Thus, big data has become ubiquitous in modern society, providing opportunities for extracting valuable insights from large datasets. However, storing, analyzing, predicting and processing to obtain insightful results from massive datasets with varying and complex structures have become major modern-day challenges. Any solution to these challenges must be computationally inexpensive and should handle high-dimensional data or larger datasets, without compromising the quality of insights achievable from the original data.

Modern machine learning (ML) techniques help transform large, unstructured datasets into smaller, more useful ones. To overcome the challenges mentioned above, feature selection is of paramount importance in the ML, data mining and information retrieval domains. Feature selection techniques simplify the model and mitigate problems such as noise accumulation, spurious correlations, and inadvertent homogeneity, which can lead to reduced computational costs [2]. Thus, feature selection methods are increasingly necessary for various applications, including computer vision, object detection, image processing, image retrieval, speech recognition, data mining, pattern recognition, machine learning, and bioinformatics. However, the applied feature selection techniques must ensure that the quality of insights extracted from the reduced data remains acceptable (see Fig. 1).

* Corresponding author.
E-mail address: ahsan.saadat@seecs.edu.pk (A. Saadat).

https://doi.org/10.1016/j.future.2024.02.017
Received 13 June 2023; Received in revised form 15 February 2024; Accepted 16 February 2024
Available online 19 February 2024
0167-739X/

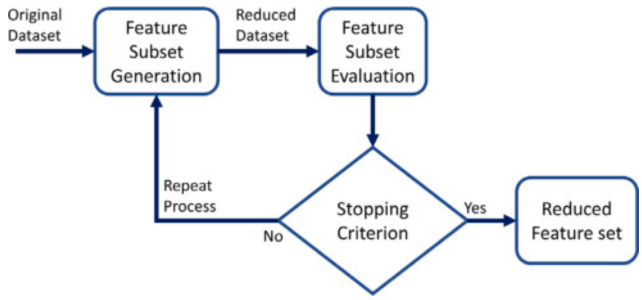


Fig. 1. Process flow diagram for feature reduction: Dataset is reduced to a feature subset via an iterative method of the algorithm, and the final feature set is deduced when a stopping criterion, i.e. no. of iterations or fitness function minimum value, is obtained.

The literature survey reveals various feature selection techniques with advantages and disadvantages. To perform effective feature selection, several factors should be considered, such as relevance, redundancy, number of features, domain knowledge, feature ranking, feature subset selection, and evaluation. Relevant features should be directly related to the problem being solved, and highly correlated features may provide redundant information. The number of features should be reduced to prevent overfitting and reduce computational costs. Domain knowledge can guide the feature selection process, while feature ranking methods can rank the features based on their importance or relevance. Different algorithms, such as wrapper, filter, and embedded methods, can select a subset of features, depending on the specific problem and available resources. The selected features should be evaluated on a separate validation dataset to ensure improved model performance and generalization ability. Therefore, feature selection poses several challenges for ML applications, including overfitting, computational complexity and bias–variance tradeoff, in which the complexity of the feature relationship with the class label needs to be considered.

Metaheuristic algorithms, a type of wrapper-based technique, are a class of optimization algorithms that have been used to solve a variety of optimization problems, including feature selection. These algorithms are particularly useful for solving problems that are computationally intensive and are able to find near-optimal solutions to problems by using heuristics and stochastic processes to explore the search space [3]. One of the main advantages of using metaheuristic algorithms for feature selection is that they can handle high-dimensional data and large datasets. This is because metaheuristic algorithms can explore many possible feature subsets in a relatively short time, allowing them to identify relevant features even when dealing with high-dimensional data [4]. Although many metaheuristic algorithms have been utilized for feature subset representation, they are not immune to inherent drawbacks (e.g., getting stuck in local optima [5]), mostly due to them having a certain amount of dependency on the innate traits of the datasets.

Recently, hybrid models have become popular to overcome the limitations of individual metaheuristic algorithms. For example, [6] is a hybrid metaheuristic algorithm combining Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for feature selection. The proposed hybrid algorithm first uses the PSO algorithm to explore the search space and obtain an initial solution. Then, the GA algorithm is used to refine the solution and improve its performance. In this way, the hybrid algorithm overcame the limitation of getting stuck in the local optima solution of PSO and improved the quality of the selected features. The authors of [7] propose a hybrid of artificial bee colony and whale optimization for feature selection and parameter optimization of an artificial neural network to improve the accuracy of breast cancer diagnosis. The hybrid algorithm overcomes the limitation of class imbalance by using a weighted classification approach that assigns different weights to each class. For example, the study in [8] hybridized game theory and dynamic programming to reduce the size of the ensemble classification technique. The pruning algorithm was tested on 13 UCI datasets and showed improved accuracy while achieving extra diversity in the solution search space. The authors of [9] used a consensus-based combining method to deduce the optimal ensemble classifier that iteratively adjusts its weight, and the final output reaches a consensus. The proposed algorithm was tested on 14 public datasets, and experimental results showed significant improvement in classification accuracy over the average methods. Despite their success, the introduction of hybrid metaheuristic algorithms is continuously ongoing. The No Free Lunch (NFL) theory holds that no method can solve all optimization problems. The choice of the constituent algorithms and their parameters can significantly impact the performance of the hybrid algorithm in different problem domains and datasets. Thus, different algorithm combinations need to be tested on a variety of datasets for continuous innovation in this domain. Secondly, existing hybrid algorithms still suffer from several limitations, such as increased complexity [10], difficulty in balancing the increased number of parameters to tune, and scalability to larger datasets. For example, in [11], the Improved Seeker Optimization Algorithm had two tuning parameters while the firefly algorithm had three tuning parameters. The hybridized version of these two algorithms had four parameters to adjust, and the mathematical model further increased the algorithm's complexity.

To address these gaps, we propose a novel hybrid algorithm that uses a binary variant of the Firehawk metaheuristic technique in conjunction with the Sine Cosine Algorithm (SCA). The Firehawk metaheuristic algorithm is used in our study primarily because it demonstrates capability in dealing with real-size structural frames and provides superior results when tested on numerous benchmark test functions. This validation suggests that the algorithm may apply to high-dimensional and complex problems such as feature selection for large-scale datasets. We enhance the Firehawk technique in its global minima finding capabilities by improving its capacity to adequately move in the solution search space of the underlying problem landscape and produce better results. The proposed novel algorithm is called the Hybrid Sine Cosine – Firehawk Algorithm (HSCFHA). In this proposed technique, the Sine Cosine algorithm plays the role of a low-level team player assisting the Firehawk Algorithm in searching for the global solution optimally, more precisely, and within a smaller time window. Our comparative analysis indicates that the proposed technique is superior to the state-of-the-art in efficiency and effectiveness. The major contributions of our work are summarized as follows:

- A feature selection model is established to create a subset feature dataset that aims to reduce the storage of large amounts of data, remove redundant features and enhance the accuracy of the classification model.
- A hybrid metaheuristic model called HSCFHA is developed using the Firehawk Algorithm and the high-performing Sine Cosine Algorithm, to leverage the exploration and exploitation capabilities of both algorithms and enhance the overall search performance to get better results for the feature selection problem.
- The proposed feature selection technique is evaluated using 22 datasets that contain low, medium and high dimensional datasets using an HSCFHA-based Artificial Neural Network to test its performance.
- A comparative analysis of the performance of the proposed algorithm is performed to evaluate it against state-of-the-art metaheuristic techniques, namely Particle Swarm Optimizer, Grey Wolf Optimizer, Whale Optimization Algorithm, Hybrid Ant Colony Optimizer and Improved Binary Bat Algorithm.

The subsequent section of the paper discusses prior research on feature selection techniques, the accompanying challenges and the use of metaheuristic algorithms in feature selection. Section 3 presents the mathematical models of the vanilla Firehawk Algorithm and the Sine Cosine Algorithm and discusses the hybridized version, considering its cost/fitness function, system update policy and computational complexity. In the experimental results and discussion section, the algorithm is compared with state-of-the-art metaheuristic algorithms on low, medium and high dimensional datasets in terms of accuracy, precision, F1-score, confusion matrix and time complexity. Additionally, the model from which these metrics are drawn from an HSCFHA-based Artificial Neural Network (ANN) that is utilized for the reduced datasets after feature selection has been achieved. Finally, the conclusion and future work section summarizes the work presented and offers insights into future directions.

2. Related work

The related work section first discusses traditional feature selection techniques that have been discussed in the literature. We then introduce metaheuristic algorithms and their use for feature selection.

2.1. Traditional feature selection techniques

A widely used feature selection technique, the filter method, evaluates each feature's significance or relevance independently of the classification model based on a criterion [12]. Commonly used criteria include mutual information, information gain, and correlation-based feature selection (CFS). The mutual information criterion evaluates the relationship between a feature and the class label. The idea behind it is that a feature that is strongly connected to the class label is considered to be more informative. Mutual information is a non-parametric measure that can be used with any type of feature and class label [13]. Information gain is a variation of mutual information used in decision tree algorithms. It measures the reduction in the entropy of the class label after the feature is used to split the data [14]. CFS is based on the idea of measuring the relationship between the feature and the class label. CFS is based on the correlation coefficient, which measures the linear relationship between two variables. While these methods are often simple and efficient, they may not be effective in all cases. For example, they only consider the individual relevance of each feature to the class label. They may not consider complex relationships between features that may also be important for accurate classification [15]. Additionally, these methods do not consider the specific learning algorithm that will be used and may not be optimized for the performance of that algorithm. Another issue with filter-based feature selection is that it is susceptible to the choice of criterion, which can result in varying outcomes based on the chosen criterion. For example, mutual information and information gain are different criteria that can be used to measure the relevance of a feature, but they may give different results [16]. Additionally, these methods can be sensitive to outliers, leading to a bias towards certain features [17].

Another widely used feature selection technique, embedded methods, are based on the optimization of a criterion that is embedded in the classifier's training process [18]. These methods optimize the criterion by adjusting the parameters of the classifier, such as the weights of the features in the linear classifiers or the split parameters in the decision trees [19]. Commonly used embedded methods include LASSO and Random Forest [20]. These techniques require more computational resources than the filter and wrapper approaches, but they can yield better results by considering the relationship between the features and the class label [21]. One of the main challenges with embedded method-based feature selection is that it requires using a specific classifier or algorithm, which can lead to suboptimal feature selection if the classifier or algorithm is not well suited to the dataset. Furthermore, embedded method-based feature selection fails to account for the data distribution, leading to inadequate feature selection by being susceptible to outliers, potentially resulting in a bias towards certain features [22].

Another popular feature selection technique is the wrapper method, which evaluates the performance of a classifier with different subsets of features. The wrapper method is based on a search algorithm that iteratively selects or removes features based on the classifier's performance [23]. The two most frequently employed wrapper methods are sequential forward selection (SFS) and sequential backward selection (SBS) [24]. The wrapper approach requires more computational resources than the filter method, but it can lead to improved results by taking into account the relationship between features and the class label. Accompanied challenges with this area of feature selection include computational expensiveness, as it requires the training of a classifier or algorithm multiple times with different subsets of features. This can pose a significant difficulty when working with large datasets and high-dimensional data. Secondly, the need for a specified algorithm limits its applicability to different datasets and tasks [25].

Recently, deep learning has become increasingly popular in the field of machine learning and is being applied to feature selection. Deep learning-based feature selection methods are able to learn complex representations of the data, which allows them to identify features that are relevant for a specific task or classifier [16]. One of the main advantages of deep learning-based feature selection is that it is able to handle high-dimensional data and large datasets. However, using deep learning for feature selection necessitates the requirement of a substantial amount of data to train deep neural networks, which can pose a problem when dealing with small datasets. These methods can be computationally expensive, as they require the training of large neural networks.

2.2. Metaheuristic algorithms

Metaheuristic algorithms are probabilistic solvers belonging to the family of approximate optimization methods. Fundamentally, metaheuristics offer domain-specific knowledge by employing an upper-level strategy of solution sharing between many nodes being implemented in parallel. These algorithmic functions are segmented into two phases: exploration and exploitation phases. While the exploration phase searches for a global solution around the search space using erratic movements, the exploitation phase converges towards the current iteration's global solution. The trade-off between the two phases is key to an efficient search process, which is dictated by the hyperparameters set for the algorithm. From an inspirational standpoint, we focus on the subdivisions of metaheuristic algorithms such as biology-based, mathematics-based and physics-based [26–28].

Algorithms grouped as biology-based are inspired by the social behavior of birds, fish, animals, and ants and the biological process of evolution. The most prominent of these categories are the PSO algorithm and GA. In [6], the authors propose a feature selection algorithm based on a combination of PSO and a GA. The PSO algorithm is used to generate an initial population of solutions, while the GA is used to perform the selection and crossover operations. The algorithm was tested on several benchmark datasets and was shown to outperform several other feature selection methods. A population-based algorithm mimicking the behavior of cuckoo birds searching for prey was presented in [29] for feature selection. The Ant colony optimization (ACO) algorithm, which was proposed in [30], uses the search strategies of ants to find the optimal feature subset. The authors used the ACO to select features for a breast cancer dataset and showed that it outperformed other popular feature selection methods. A similar case dataset was also presented in [7] with a combined algorithm for Artificial Bee Colony (ABC) Optimization algorithm and Whale Optimization Algorithm (WOA). Alzubi et al. [31] proposed a novel algorithm combining the Harris Hawk Optimizer with Support Vector Machines to detect Android malware. In [32], the authors proposed two models with the hybridization of ACO and GA and compared them with the state-of-the-art on multiple UCI-based datasets. Another study in [33] provided a hybridized version of the GA and Cuckoo search algorithms as a hybridized wrapper-based method to solve the feature selection task. Moradi et al. [34] presented a local exploration model to guide PSO in selecting marginal deducts with respect to their correlation data. In the application of robotic manipulators, a modified PSO was augmented with differential evolution (DE) to provide a general case study focusing on serial and parallel manipulators and the 10-DOF hybrid redundant serial-parallel robots [35]. To the best of our knowledge, the recently introduced Firehawk algorithm (detailed model explained in the later section) has not been applied to the problem of feature selection.

Mathematics-based metaheuristic algorithms use mathematical models and techniques to solve complex optimization problems, including feature selection. Two of the most commonly used mathematics-based metaheuristic algorithms are the Basic Optimization Algorithm (BOA) and the SCA. In the case of SCA, many variants and improvements have been suggested in the literature for the purposes of feature selection. For example, using an Elitism strategy and a new best solution update mechanism, the basic sine-cosine algorithm was improved in [36] and showed efficiency in achieving better classification performance along with fewer features compared to GA, PSO, and basic SCA in selecting the best features for classification. The study in [37] proposes a two-stage framework that combines deep learning and SCA to detect pneumonia in X-ray images. Tagian et al. The authors of [38] propose two binary variants of SCA, S-shaped Binary SCA and V-shaped Binary SCA, for feature selection from medical datasets. The algorithms are compared with four other binary optimization algorithms on five medical datasets. The experimental results show that both SBSCA and VBSCA outperform the other algorithms in terms of classification accuracy on these datasets. In [39], the authors propose an improved SCA algorithm for feature selection in text categorization tasks by combining two positions of the solution to avoid premature convergence. The proposed algorithm is compared with several search algorithms and shows high performance on nine text collections. A new approach was proposed in [40] called MOSCA-FS for feature selection in hyperspectral imagery using a novel discrete Sine Cosine Algorithm. The approach uses a multi-objective optimization framework to balance information preservation and redundancy reduction. The proposed method is validated through experiments on multiple datasets, demonstrating its effectiveness and universality.

In the case of physics-based algorithms, the inspiration for such techniques is drawn from the laws of physics around the world which include physics principles, chemistry, music, dynamic systems and metallurgical processes. In [41], the golden ratio-based equilibrium Optimization algorithm was proposed for feature selection for a classification problem of speech emotion recognition. The technique utilized a hybridized version of the two algorithms. Lenin et al. [42] solved the reactive power problem with the combination of Tabu Search and Simulated Annealing (SA). The algorithm was also applied to a symmetrical travel salesman problem. Mafarja et al. [43] proposed a hybrid feature selection technique by combining SA and GA. The algorithm showed good performance upon testing it on the UCI dataset in terms of the number of selected attributes in comparison with state-of-the-art approaches.

Overall, the outcome of a metaheuristic algorithm depends on the starting point solution and the stopping criterion, which makes it hard to compare the results of different algorithms. Additionally, it is also challenging to determine the optimal parameters of metaheuristic algorithms, which can affect the results of the feature selection process.

The comprehensive literature review highlights the evolving landscape of feature selection techniques, particularly in the context of high-dimensional datasets and complex machine learning challenges. The proposed HSCFHA is a response to the identified gaps and limitations in existing methods. This approach inculcates the advantages of metaheuristic algorithms with the adaptive capabilities of ANNs. By doing so, it can refine feature selection processes significantly, enhancing both the accuracy and efficiency of model training and prediction. This research not only offers a novel solution but also sets the stage for further advancements in feature selection methodologies, potentially revolutionizing approaches in data-intensive domains.

3. Proposed hybrid Sine Cosine - Firehawk algorithm

In this section, we provide background details and present a novel hybrid algorithm based on a combination of two metaheuristic algorithms. This section first introduces the concepts of the vanilla Firehawk Optimization algorithm and the Sine Cosine algorithm, and then provides details of the mathematical model and complexity analysis of the proposed hybrid algorithm.

3.1. Background

3.1.1. Firehawk algorithm

A metaheuristic algorithm called the Firehawk Optimization Algorithm has been developed using **sout** inspiration drawn from the prey-hunting strategy of the firehawk birds found in nature [44]. There are some challenges and limitations accompanying the algorithm that hinder the performance of the technique. Firstly, the capability of the agent to explore the search space is limited by the movement allowed due to the Gaussian randomization process. Secondly, due to the complexity of the algorithm, the agents get stuck in local optima and fail to explore the search space effectively, leading to suboptimal solutions in case of multifaceted problems.

3.1.2. Sine-Cosine algorithm

The Sine Cosine Algorithm comes under the family of mathematics-based metaheuristic algorithms and the goal of algorithms is to find the optimal solution by exploring the solution space in a balanced way, considering both global and local information about the objective function. The SCA initiates multiple initial candidate solutions and guides them towards the optimal solution using a mathematical model based on trigonometric sine and cosine functions [45]. The performance of the standard SCA is satisfactory for solving unimodal benchmark function problems. However, when confronted with complex multimodal functions, the algorithm often exhibits a tendency to quickly converge to local minima, potentially missing out on better solutions [46]. Although the SCA is capable of finding the optimum given sufficient computational resources, its reliance on random walks hinders its ability to guarantee fast convergence. Consequently, there is a possibility to enhance the SCA's performance by integrating it with other complementary techniques. The algorithmic details of the SCA are provided in Algorithm 1.

Algorithm 1 Steps for SCA

```
Initialize population size, dimensions d, population size and max_iter

Initialize population fitness values
while iter ≤ max_iter do
    Evaluate parameter ρ
    for i = 1: population_size do
        for j = 1 : dimension (d) do
            Evaluate candidate solution using the exploration and
            exploitation phases
        end for
    end for
    Evaluate fitness values of population
    Choose the global best value
    iter = iter + 1
end while
return Global Best
```

3.2. Mathematical model of the hybrid Sine Cosine - Firehawk algorithm

The initial conditions for the proposed hybrid algorithm are the same as those for the FHA. The FHA posits the mathematical model of the hunting mechanism of firehawks, where the objective is to set and spread fires and consequently catch prey. First, a solution candidate

(X) matrix is determined, set initially as random values within a search space, representing the initial positional vectors of the firehawk and its prey.

$$PR = \begin{bmatrix} PR_1 \\ PR_2 \\ \vdots \\ PR_N \end{bmatrix} = \begin{bmatrix} PR_1^1 & PR_1^2 & \dots & PR_1^d \\ PR_2^1 & PR_2^2 & \dots & PR_2^d \\ \vdots & \vdots & \dots & \vdots \\ PR_N^1 & PR_N^1 & \dots & PR_N^d \end{bmatrix}, \quad (1)$$

where PR represents the candidate solution of the prey in the search space, d is the number of dimensions of the current problem, and N is defined as the number of candidates utilized for the metaheuristic algorithm.

Eq. (2) represents the search space boundary conditions, where PR_{min} is the minimum value that the current solution of PR at iteration t can achieve, while PR_{max} is the maximum possible attainable value that can be achieved. These values are set according to the problem at the start of the execution of the algorithm.

$$PR_j^i(t) = Range \;\; \text{---} \;\; [PR_{min} \text{ - } PR_{max}] \; ; \quad \text{where } i = 1, 2, \dots, d; j = 1, 2, \dots, N. \quad (2)$$

To determine the initial position of the firehawks in the search space, the cost function of each candidate PR is evaluated. Some of the best candidates are chosen from the solution search space to become the firehawks. From the firehawks, the one with the best search space is considered the main fire, which can also be considered the global best position. The main fire is then spread by the firehawks. The consequent movement of the hunter (firehawk) and prey determines the mathematical model of the algorithm. The distribution of the prey candidates is divided as per (3) and (4), where m is the number of firehawks and n is the number of prey in the investigation domain. Here, $m + n = N$.

$$PR_{initial} = \begin{bmatrix} PR_1^1 & PR_2^1 & PR_d^1 \\ PR_1^2 & PR_2^2 & PR_d^2 \\ \vdots & \vdots & \vdots \\ PR_1^n & PR_2^n & PR_d^n \end{bmatrix}; \quad (3)$$

$$FH_{initial} = \begin{bmatrix} FH_1^1 & FH_2^1 & FH_d^1 \\ FH_1^2 & FH_2^2 & FH_d^2 \\ \vdots & \vdots & \vdots \\ FH_1^m & FH_2^m & FH_d^m \end{bmatrix}. \quad (4)$$

The initial search area can be graphically represented as shown in Fig. 2. When the hunting process of the firehawks begins, they spread fires in surrounding areas to entrap the prey. To avoid collisions in each other's area of attack, the firehawks avoid the territory of another firehawk's fire circle. The distribution of each firehawk with their respective number of prey can be mathematically evaluated using a Cartesian distance function, as per Eq. (5). The prey are assigned a firehawk territory according to the closest firehawk around them. The initial population of PR_N is assigned a random (x, y) coordinate in the search space.

$$Dist_F^P = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \quad \text{where } P = 1, 2, \dots, n; F = 1, 2, \dots, m. \quad (5)$$

The $Dist_F^P$ represents the distance of the P*th* prey to the F*th* firehawk; and the (x_1, y_1) and (x_2, y_2) represent the prey's and firehawk's coordinates respectively. It is worth noting that the best prey is assigned to the best firehawk in the search space. The next firehawk takes on the next nearest prey in the search space, and so on. This ensures that the firehawk with the best objective function performs better in its hunting strategy than the weaker birds. Once this procedure is completed, the algorithm's initial conditions are configured with each prey assigned to a firehawk territory.

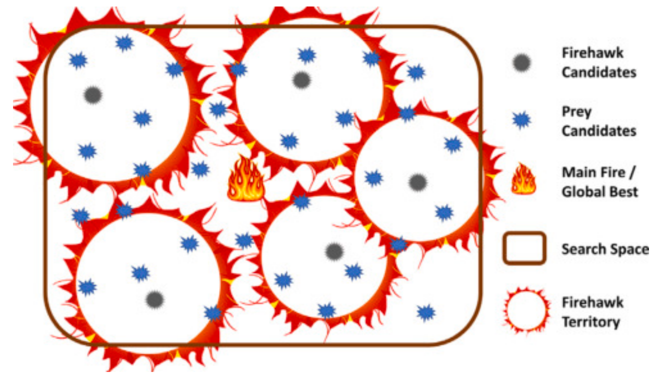


Fig. 2. Firehawk model search space schematic.

Following that, the algorithm is executed for a user-defined number of iterations, set at the start, in which the firehawk position update procedure is conducted against the objective fitness function. The methodology mainly contains the usage of the exploration and exploitation phases to conduct the mechanism of updating the firehawk to become the main fire/global best or move its territory towards another position in the search space. The movement of the algorithm towards the next position in the search space is determined by random and adaptive variables, which introduces a level of unpredictability. Consequently, the algorithm cannot guarantee finding a satisfactory solution consistently. While the FHA demonstrates promising results on certain benchmark functions by utilizing population knowledge effectively, it may not always produce optimal solutions for complex problems.

To further improve the exploitation capabilities of the FHA, a specialized version of the Sine Cosine Algorithm is integrated into the FHA, resulting in a novel algorithm called Hybrid SCFHA. The Hybrid SCFHA algorithm utilizes the sine and cosine functions to generate new positions within the search space. When solving minimization problems, if the newly generated solution possesses a lower fitness value, it replaces the existing solution, effectively serving as the solution for the subsequent iteration.

At every iteration, the new firehawk candidates and the main fire are determined for N solution candidates. This is done using burning sticks from the main fire and spreading the fire across other places, creating firehawk territories. The schematic representation shows a circular pattern of the firehawk territory that can sometimes move away from the search space. This violation is also catered for by the boundary conditions set in Eq. (2). The firehawk updating policy with Sine Cosine mechanism infused is given by

$$FH(iter) = \begin{cases} FH_m^{new\_iter} = FH_m + r_1(\sin(r_1)) \times \\ \quad MainFire - r_2(\sin(r_2)) \cdot FH_{nr} \quad \text{if } r_a < 0.5; \\ FH_m^{new\_iter} = FH_m + r_1(\cos(r_1)) \times \\ \quad MainFire - r_2(\cos(r_2)) \cdot FH_{nr} \quad \text{if } r_a \geq 0.5, \end{cases} \quad (6)$$

where $FH_m^{new\_iter}$ is the updated firehawk position, FH_m^{iter} is the previous firehawk position, FH_{nr} is the position of one of the other firehawks and r_1 and r_2 are two Gaussian distributed random values within the range $[0, 1]$. The movement of the prey is also considered a key aspect of animal behavior. The prey either runs away, hides, or, in distress,

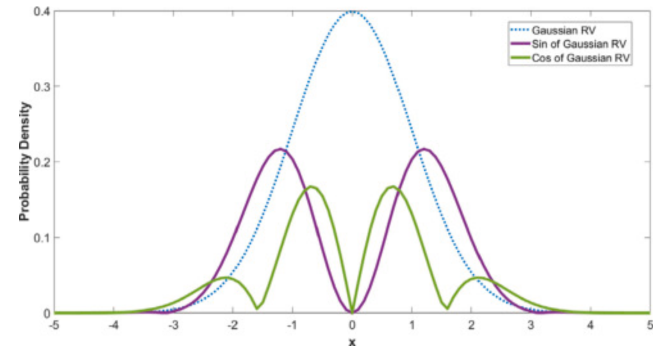


Fig. 3. PDF of Random Variables (RV).

moves towards the firehawk territory by mistake. The prey updation process of running away from fires is given by

$$PR(iter)_{exploitation} = \begin{cases} PR_n^{new\_iter} = PR_n + r_3(sin(r_3))\times \\ \quad FH_m - r_4(sin(r_4)) \cdot SP_m & \text{if } r_b < 0.5; \\ PR_n^{new\_iter} = PR_n + r_3(cos(r_3))\times \\ \quad FH_m - r_4(cos(r_4)) \cdot SP_m & \text{if } r_b \geq 0.5, \end{cases} \tag{7}$$

Similarly, the movement of the prey that results in the prey getting closer to another firehawk $FH_{another}$ is given by

$$PR(iter)_{exploration} = \begin{cases} PR_n^{new\_iter} = PR_n + r_5(sin(r_5))\times \\ \quad FH_{another} - r_6(sin(r_6)) \cdot SP_{all} & \text{if } r_b < 0.5; \\ PR_n^{new\_iter} = PR_n + r_5(cos(r_5))\times \\ \quad FH_{another} - r_6(cos(r_6)) \cdot SP_{all} & \text{if } r_b \geq 0.5, \end{cases} \tag{8}$$

where $PR_n^{new\_iter}$ is the new prey $n's$ position, r_3, r_4, r_5 and r_6 are random functions in the range [0,1] and SP_m is considered as a safe place away from the fires that the prey of the m_{th} firehawk territory may find while running away from that firehawk, represented in Eq. (9).

$$SP_m = \frac{\sum_{q=1}^{n_m}(PR_q)}{TP_m}, \tag{9}$$

where TP_m is total no. of prey in m_{th} firehawk territory and PR_q represents the q_{th} prey in m_{th} firehawk territory.

Let SP_{all} represent the safe place for all prey from each firehawk in the search space depicted with Eq. (10).

$$SP_{all} = \frac{\sum_{k=1}^{n}(PR_k)}{TP}, \tag{10}$$

where TP is the total prey population. It is noted that for each search iteration, the number of firehawks is dynamic and is determined as the difference between the total particles in the search space and the number of prey that are calculated randomly with a Gaussian distribution. r_a and r_b are Gaussian random values within the range [0,1]. Over a series of iterations, the position of every individual in the FHA is updated using Eqs. (6), (7) and (8). These hybridized equations are then employed in the process of feature selection. Algorithm 2 illustrates the pseudo-code for the complete algorithm.

Fig. 3 approves the theory of optimized exploration as the Sine Cosine-based random variable has a more spread-out probability density function (PDF) than the simple Gaussian random variable from the vanilla FHA algorithm. This means that the hybrid SCFHA exploration capability is enhanced as the SCA-based random variable will explore

Algorithm 2 Steps for HSCFHA

Set the number of max_iter, population size, dimension d and α for the fitness function
Initialize the prey and firehawk population according to Eqs. (3) and (4).
Evaluate the initial candidate solution of the population
Determine Global Best (GB) / Main Fire position from the entire population
while iter ≤ max_iter **do**
  Determine new firehawk and prey population distribution at each iter
  for m = 1 : firehawk_pop. **do**
    Sort firehawk / prey population and determine prey position in each firehawk territory using Eq. (5)
  end for
  Calculate SP for preys in complete search space using Eq. (10)
  for m = 1 : firehawk_pop. **do**
    for d = 1 : Total_dimensions **do**
      Determine FH_{nr}, $FH_{another}$ and calculate new position of firehawk using Eq. (6)
    end for
    Evaluate SP_m using Eq. (9)
    for n = 1 : Prey_pop. **do**
      for d = 1 : Total_dimensions **do**
        if rand() ≤ 0.5 **then**
          Determine new prey position using Eq. (7)
        else
          Determine new prey position using Eq. (8)
        end if
      end for
    end for
    Evaluate boundary conditions using Eq. (2)
    Calculate fitness of prey population
  end for
  Calculate firehawk population fitness and sort population
  Determine the GB position from the population
  iter = iter + 1
end while
return GB position

new values in the search space before moving towards a more promising area. Following the exploration phase, the Hybrid SCFHA algorithm incorporates the SC algorithm to conduct a local search with smaller steps, ultimately obtaining the best solution. Within the framework of Hybrid SCFHA, the FHA emphasizes diversification during the initial stages of the search by taking larger steps to thoroughly explore the search space and avoid getting stuck in local optima. In contrast, the later stage of the optimization process focuses on intensification, where the SC algorithm guides individuals towards the best solution. This combined approach effectively addresses both global and local search objectives.

3.3. Complexity analysis of the hybrid Sine Cosine - Firehawk algorithm

To analyze the computational complexity of a metaheuristic algorithm, it is common practice to employ the Big O notation, a widely recognized mathematical notation in the field of computer science. The computational complexity of the proposed HCSFHA and the standard FHA remains the same, as the solution update mechanism in both algorithms is $O(iter_{max} \cdot m \cdot n) + O(iter_{max} \cdot m \cdot n \cdot d)$, where $iter_{max}$ denotes maximum iterations, m the firehawk population size, n the prey population size, and d the problem dimension. This is because the proposed HSCFHA does not supplement any additional processes;

Table 1
Comparison of metaheuristic algorithms with HSCFHA for $\vec{T}$ time complexity analysis.

Fun.	HSCFHA	SCA	FHA	IBBA	HACO	GWO	PSO	WOA
CEC01	335.14	323.71	329.90	1316.47	1317.40	347.00	364.89	404.96
CEC02	14.14	13.15	13.41	157.82	142.92	15.27	12.62	14.33
CEC03	18.96	18.50	18.83	110.60	110.33	20.58	19.56	20.90
CEC04	12.31	11.83	12.12	323.88	325.72	13.89	14.30	13.76
CEC05	14.77	13.55	13.74	214.40	212.36	14.97	15.31	16.95
CEC06	120.59	114.43	115.17	3182.05	3110.40	159.38	111.73	165.83
CEC07	15.07	13.37	14.99	64.37	65.24	16.40	15.09	16.98
CEC08	13.52	12.17	12.08	114.51	124.25	14.39	14.83	15.48
CEC09	13.78	12.48	12.95	33.90	34.29	14.07	13.19	15.13

it simply replaces the exploration or exploitation processes in the standard FHA with the hybridized version that includes the SCA equation, which makes no difference in terms of computational cost. Therefore, we argue that the proposed HSCFHA performs better than the original FHA with no additional cost.

A time-based computational cost analysis of the HSCFHA algorithm is evaluated using CEC-2020 [47] protocol benchmark test functions. Four specific computational times, T_0, T_1, T_2 and $\hat{T}_2$ are determined, which establish the complexity of the algorithm. T_0 is the runtime of a specific mathematical algorithm, T_1 is the computational time of the CEC function which has been run 10,000 times, T_2 is the computational time of the technique used to solve the minimization problem of the CEC function under 10,000 iterations, and $\hat{T}_2$ is the mean value of 5 repetitions of the T_2 time analysis.

$$\vec{T} = \frac{\hat{T}_2 - T_1}{T_0}. \tag{11}$$

The proposed algorithm of HSCFHA is tested on these computational times and is compared with other similar metaheuristic algorithms on each CEC benchmark test function. The result of the experiment is shown in Table 1. The value of T_0 was determined to be 0.1162 seconds. Since the algorithms have a multi-agent solution-finding technique, a population size of 10 was used for each algorithm in comparison. The machine used for the complete analysis is a Core i7 9750 h with 16 GB RAM.

It is crucial to acknowledge the significance of the dimensions in multi-agent algorithms, particularly the balance between exploration and exploitation phases, aligned with the rate of convergence. This factor plays a vital role in assessing the innovative HSCFHA algorithm. The comprehensive tests performed in this study demonstrate that the HSCFHA algorithm's optimization process is inclined to focus on a global search space, leading to more effective outcomes.

4. Feature selection by hybrid Sine Cosine - Firehawk algorithm

The goal of feature selection is ultimately to choose adequate features from the complete dataset. This means that the algorithm will output features that are selected and features that are not, i.e., a binary output of 1s and 0s. Therefore, this work uses a one-dimensional vector to represent a solution, with the length of the vector depending on the number of attributes in the original dataset. "1" or "0" indicates each value in a vector (cell). The associated attribute is picked if the value is "1"; otherwise, it is set to "0". The values are chosen depending on the fitness function of feature selection that the algorithm uses to determine the global best position. The objective of the technique is to minimize the fitness function utilized, as given by

$$\begin{aligned} F.F &= \alpha \cdot Variance(Agent_{Dataset}) + \\ &(1-\alpha) \cdot \frac{Tot\_feat. - unselected\_feat.}{Tot\_feat.} \end{aligned} \tag{12}$$

where the parameter α represents the significance of the variance quality and subset length, the value of which is decided before the start of the algorithm. For the purposes of this study, the value of α is chosen to be 0.5. In contrast, the traditional KNN-based fitness function tries to maximize the accuracy of the KNN model and the percentage of unselected features from the current iteration simultaneously. The feature selection process removes the requirement for a classification algorithm in its iterative process by employing the variance-based technique. This drastically reduces the time the algorithm takes to produce optimal features. The flowchart of the technique employed is shown in Fig. 4.

4.1. Model verification using HSCFHA-based artificial neural network

Artificial Neural Networks (ANN) belong to a category of machine learning algorithms that draw inspiration from the structure and functioning of the human brain. These networks find extensive application in diverse domains such as image recognition, natural language processing, financial forecasting, etc. A neural network consists of interconnected nodes, called neurons, that process and transmit information using mathematical operations. Metaheuristic-based neural networks combine the strengths of neural networks and metaheuristic algorithms to improve the performance of NN in terms of accuracy, convergence speed, and computational efficiency. These methods use metaheuristic algorithms to optimize the parameters of the neural network, such as the weights and biases. A literature survey of recent studies on metaheuristic-based neural networks shows that these methods consistently produce better results compared to traditional neural networks, with less computational time and fewer iterations [48–50]. For instance, Movassagh et al. [51] integrated invasive weed optimization with differential evolution to enhance the perceptron neural network precision.

By utilizing the strengths of both neural networks and metaheuristic algorithms, the proposed model, shown in Fig. 5, is expected to produce better results with less computational time and fewer iterations.

5. Experimental results

We now describe the experimental evaluation of our proposed technique. The extensive literature review suggests that metaheuristic algorithms exhibit suitable and balanced behavior to solve problems with non-exact solutions. For this reason, our study utilized the use of such unique algorithms. To achieve this objective, it is crucial to present a thorough and accurate depiction of the problem at hand. While metaheuristic algorithms do not guarantee optimal solutions, their purpose is to provide solutions that approach optimality while minimizing CPU usage. Various metaheuristic techniques proposed in existing literature have their strengths and limitations. This study aims to develop a more comprehensive and precise strategy by considering a broader range of parameters. Many datasets will provide a comprehensive analysis, and using multi-dimensional datasets and multi-class problems will help illustrate the accuracy and precision of the algorithm under study.

The experimental results showed that our proposed strategy, which focuses on a larger set of parameters, was able to solve the multi-objective optimization problem related to various sizes of datasets effectively and efficiently. The results demonstrate that our strategy outperformed the baseline approaches and was able to find solutions

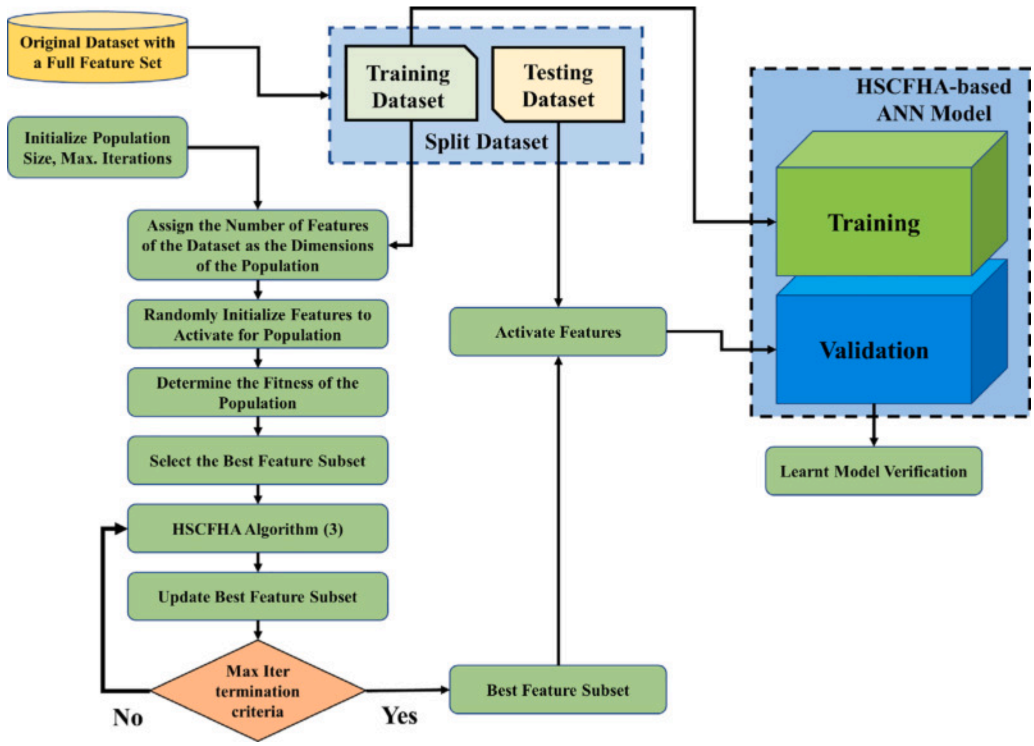


Fig. 4. HSCFHA-based feature selection flowchart.

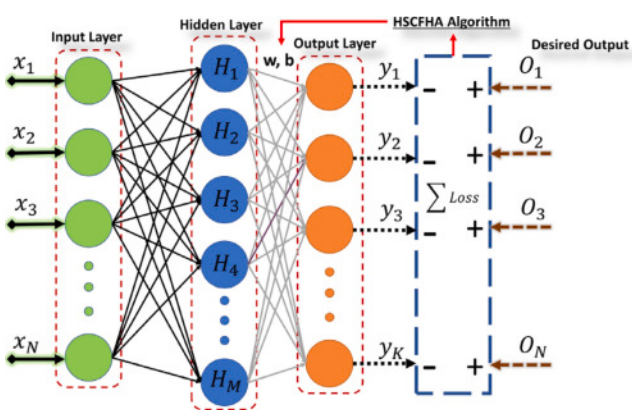


Fig. 5. HSCFHA based ANN Model.

that were closer to the optimal solution, while also taking into account the practicality and realism of the solutions. In particular, variance-based FS proves to be a far more effective strategy for feature selection than the traditional KNN-based approach.

5.1. Dataset preprocessing

For the purposes of this study, the proposed feature selection technique is employed on 22 publicly available datasets from the UCI database that include low, medium and high dimensions with multi-class and binary-class problems. Table 2 provides the details of each dataset in terms of attributes, instances and the number of classes. Figure 11 in the appendix section depicts the data spread of the datasets used in the experimental results section through boxplots. It is evident that the datasets are an amalgamation of categorical and numerical data, and some features depicted do not contribute to the overall class specification of the dataset. For example, for the seismic dataset, features 14, 15 and 16 do not have a well-rounded spread and, hence, will generally be removed after the feature selection process. Furthermore, the number of instances in the dataset contribute to the overall computational time for the operation of feature selection.

The datasets used in this research contained missing values, which can significantly impact the accuracy of the analysis. To address this issue, first, the instances that had missing values were identified and removed from the dataset. This was achieved by identifying the entries in each dataset with "NA" and "?" values. Then, the subsequent rows were deleted from the dataset. This ensured that the remaining entries were complete and suitable for analysis. Another important step in dataset preprocessing is data cleaning, which involves identifying and correcting any errors or inconsistencies in the data. In our study, we performed various data cleaning techniques such as removing duplicate entries, correcting misspelled words and converting categorical data to numerical entries.

5.2. Experimental setup

The dataset is split into two sections: a training set and a testing set with an 80/20 split. An Intel Core i7 9th GEN with 16 GB RAM was used for the experiment for each algorithm and feature selection technique to draw out a fair comparison. The proposed HSCFHA method is compared with state-of-the-art metaheuristic-based feature selection techniques, including FHA, SCA, GWO [53], PSO [54], Improved Binary Bat Algorithm (IBBA) [55], Hybrid Ant Colony Optimizer (HACO) [56] and WOA [57]. The experimental setup is described below:

- Each dataset is run on variance-based and KNN-based feature selection methods with 10 iterations and a population size of 50. To achieve a well-balanced score on metrics, the algorithm is run 30 times to calculate the best scores and the average scores.
- Number of features and the time taken for each algorithm to choose a subset of features are determined from the training dataset.
- Using the chosen features, the testing dataset is utilized to ascertain the accuracy, precision, confusion matrices and F1-Score of the classification results.

Table 2
Dataset details.

S.No.	Datasets [52]	# of Inst.	# of Attr.	Classes
1	Ad	625	1558	2
2	Adult	32 561	14	2
3	Car	1728	6	4
4	Cardiotocography	2126	36	3
5	Voting	232	16	2
6	Gestures processed	9873	32	5
7	Gestures raw	9901	18	5
8	HCV	589	12	5
9	KrVKp	3196	36	2
10	Nursery	12 960	8	5
11	Obesity	2111	16	7
12	Promoters	106	57	2
13	Biodeg	1055	41	2
14	Room	10 129	16	4
15	Seismic	2584	18	2
16	Splice	3190	60	3
17	Drug	1885	12	7
18	Tumor	62	2000	2
19	BLDCL	35	1657	3
20	Arcene	100	10 000	2
21	Amazon	1500	10 000	50
22	Religious	590	8266	7

- The HSCFHA-based NN uses a 3-layered network with a single input layer, a hidden layer of 10 neurons and ReLu-based activation function, and a single output layer.
- The hyperparameter values chosen for the metaheuristic algorithm in comparison are the same as those suggested in the original papers.

5.3. Results

The experimental results are presented and discussed in this section. In the comparison of the overall performance based on the metrics as a whole, the HSCFHA-based FS method got better results. Secondly, the comparison of the two FS methods, namely variance and KNN, also shows that the variance-based FS method performs much better, especially in terms of time spent to get optimal features and the number of features chosen.

Fig. 7 illustrates the average of all the datasets provided in the experiment. In terms of accuracy, precision and FI-score, the HSCFHA algorithm outperforms, showing the highest values compared to the original algorithms, FHA and SCA, and the other compared algorithms. Furthermore, the HSCFHA-based FS method for variance and KNN show that both show the lowest recorded number of features and the time spent to calculate and produce them is also minimal. Compared to KNN, variance-based methods perform better and faster across the board. Similarly, of the 30 runs accomplished in the experiment, the best results were also chosen to be the focus of this study.

Fig. 8 illustrates the best metrics values from the 22 datasets for each algorithm for comparative analysis. In the same way, as the average results were seen, the best results also show that the HSCFHA algorithm produces the best result from the 30 runs in terms of all the metrics, and the variance-based FS is also comparably much better at producing good results than the KNN based FS method. The bar charts show that the proposed algorithm can outperform other metaheuristic algorithms and prove that the algorithm can produce better results because the exploration phase enhancement of the hybrid algorithm finds a better solution in less computational time, and the Sine Cosine randomization technique also produces a lower fitness function value by which less number of features are used to produce better accuracy and precision scores.

To gauge the performance of each algorithm against each dataset in the experimental database, the values for each metric have been tabulated in the appendix section (Tables 3–12). The highest value has been highlighted in each comparison for easier analysis and comparison of the metric scores obtained from each algorithm on each dataset.

In terms of accuracy, a comparison of accuracy results between HSCFHA and other algorithms that are assessed under the same conditions are depicted in Tables 3 and 4 for average accuracy and best accuracy, respectively. Of the 22 datasets, Table 3 shows that 18 datasets have the highest accuracy for the proposed HSCFHA, while FHA and SCA follow a close second and third in most cases. Within each dataset feature selection method, variance-based FS is seen to find the higher optimal value for all 22 datasets compared to the KNN method. The results also show that for a few datasets, the highly cited and experimented conventional GWO algorithm finds high accuracy values. In the case of best accuracy, 19 datasets show the highest accuracy with HSCFHA, and as expected, the accuracy values of variance are higher than KNN for all datasets.

Table 4 also depicts that for the datasets Voting, Tumor and DLBCL, all algorithms were able to find the global optimal solution to the feature selection problem in both FS methods. In such cases, the impact of metaheuristics on feature selection may be negligible. Consequently, previous studies have often employed various metaheuristic algorithms for feature selection without significant distinctions, particularly when dealing with problems of relatively low dimensions (below 100). This is because the expected outcomes of feature selection can be achieved regardless of the specific choice of metaheuristic algorithm.

In terms of precision, the average of the 30 runs and the best results from the 30 runs are tabulated in Tables 5 and 6, respectively. The former shows that the HSCFHA algorithm performs better on 18 out of the 22 datasets, and the latter highlights 19 datasets with HSCFHA as the best algorithm from all the metaheuristic techniques. The best precision table shows that similar to the results of the accuracy, voting and tumor datasets are also able to find the global optima solution across all metaheuristic algorithms, thereby removing the effect of metaheuristic technique in feature selection for the said datasets.

Table 7 presents the average F1 score for the 22 datasets. The result reveals that 16 datasets have the highest optimized values for HSCFHA, while the FHA and SCA algorithms have close values to the optimal solution. Conversely, Table 8 reveals that 7 out of 22 datasets have HSCFHA with optimal values, while the majority of global solutions for the case of F1 scores are determined by IBBA and WOA algorithms.

The results of the features selected for the experiment suggest that the performance of the HSCFHA algorithm outperformed the other algorithms in terms of bringing the lowest average number of features for 12 out of 22 datasets in the experiment, as shown in Table 9, and the best number of features for 15 out of 22 datasets, as illustrated in Table 10. This is a promising result and suggests that HSCFHA may be an effective method for feature selection. The number of features also depicts that the variance-based FS method is able to find lower features in every dataset case against the KNN-based technique for the average and best number of features. In some cases, such as Adult, Car and Cardiotocography, it can be seen that other algorithms calculated the lowest number of features in comparison, but to review the disparity in the superiority claim of the HSCFHA algorithm, other metrics also need to be discussed in congruence. The holistic comparison of the results is further elaborated in the discussion section.

As suggested by the complexity analysis section of the paper, the results of convergence time calculation show that the speed to determine features of optimal value for HSCFHA is similar to the one in FHA. In most cases, HSCFHA is even better at finding optimal solutions quickly because the use of the SCA technique in the proposed hybrid can converge with the exploitation phase far better. Table 11 depicts the average computational time taken from the 30 runs, while Table 12 reveals the best time from the 30 runs to go through the feature selection process.

A comparison between the average of the best metric values is drawn for each algorithm and depicted in Fig. 6. Furthermore, the bar charts also illustrate the values of variance-based and KNN-based FS

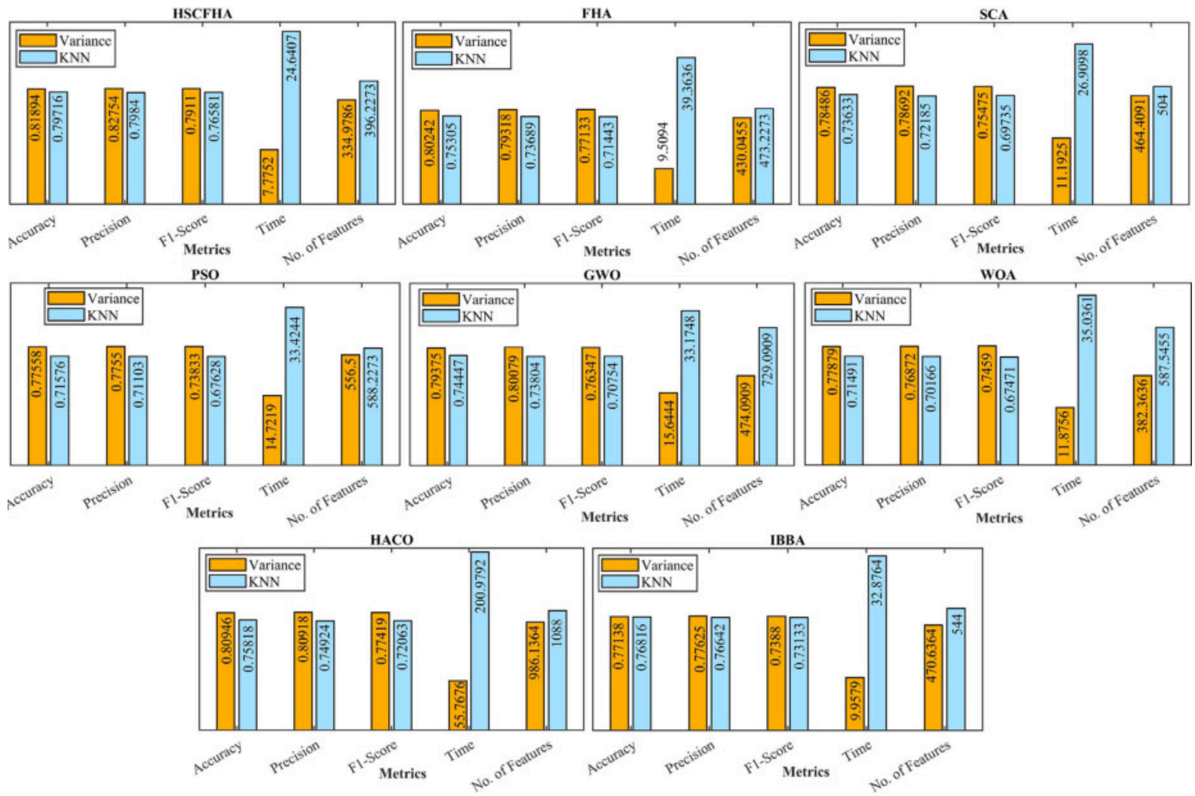


Fig. 6. Comparison between variance and KNN-based FS for HSCFHA and other metaheuristic algorithms on best metrics.

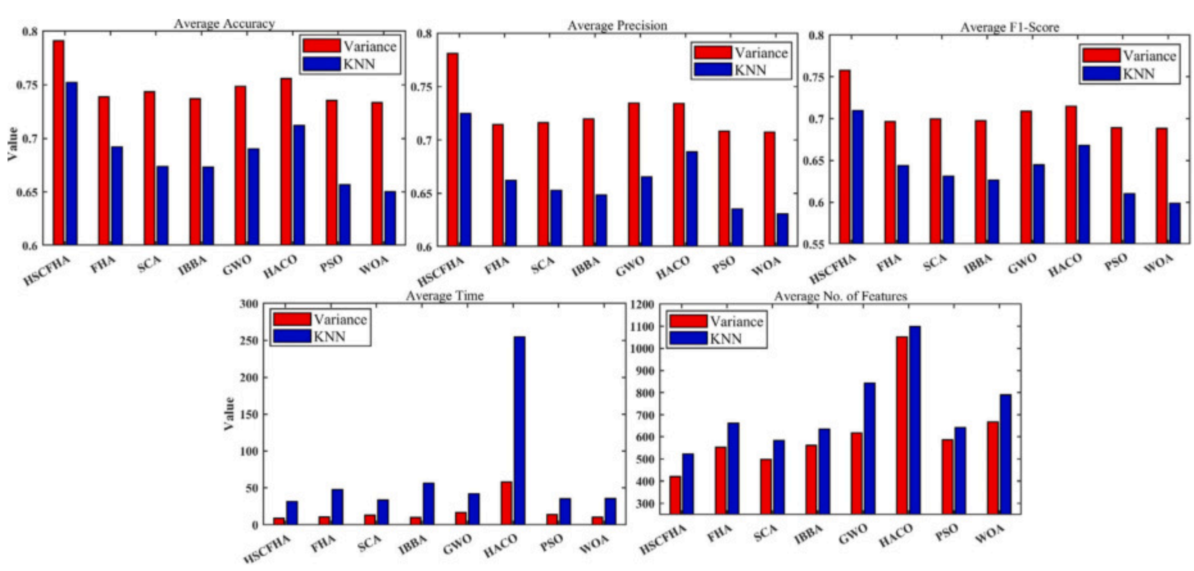


Fig. 7. Comparative Analysis of algorithms in terms of the average metrics from the 30 runs.

for each algorithm. To depict the values in a single bar chart, that data is normalized from 0 to 1 to fit onto the bar chart and the final actual resultant values are shown on the bar charts for each metric on each algorithm bar chart. It can be seen that the proposed hybrid technique can find high values of accuracy, precision and f1-score while getting an overall low computational time and number of features when compared with other techniques. Moreover, the variance-based approach has a huge difference from the KNN-based approach in terms of computational time, around three times lower and can find better optima solutions, i.e., subset features of all datasets.

Graphically, to exhibit the class imbalance on datasets used for the experiment, the confusion matrix for HSCFHA, FHA and SCA on the first few dataset results are shown in Fig. 10 and the remaining dataset confusion matrix are added in the appendix section in Figure 12. The confusion matrix offers a graphical depiction of an algorithm's performance on a specific dataset. It consists of four categories: true positives, true negatives, false positives, and false negatives. True positives and true negatives represent correctly classified instances, whereas false positives and false negatives represent instances that were classified incorrectly. By examining the values within the confusion matrix, we can calculate several metrics, including accuracy, precision, recall, and F1-score. These metrics provide valuable insights into the algorithm's performance and its ability to correctly classify instances in the dataset.

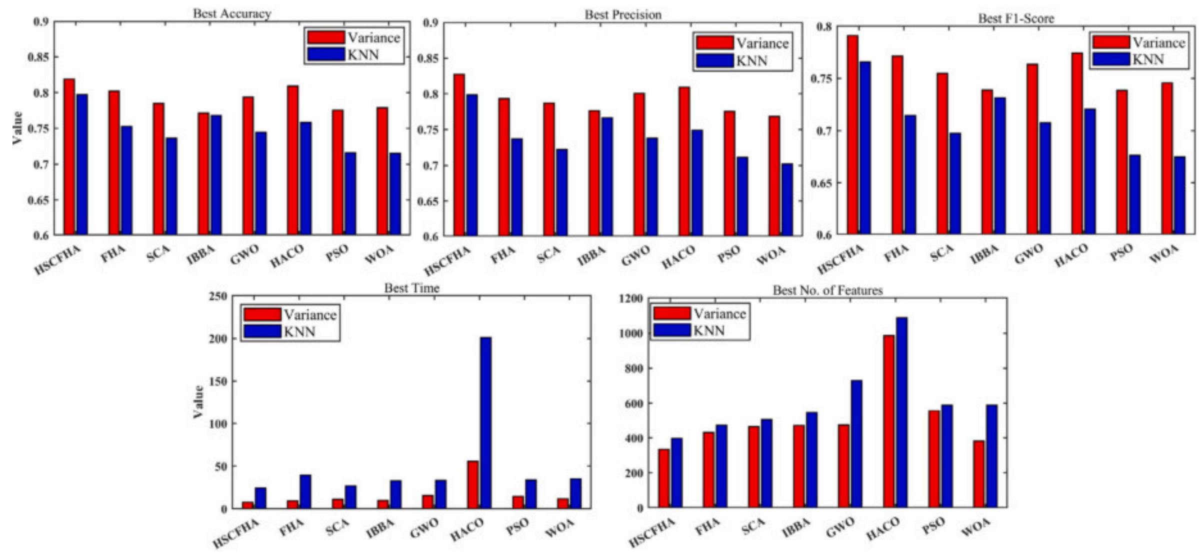


Fig. 8. Comparative analysis of algorithms in terms of the best metrics from the 30 runs.

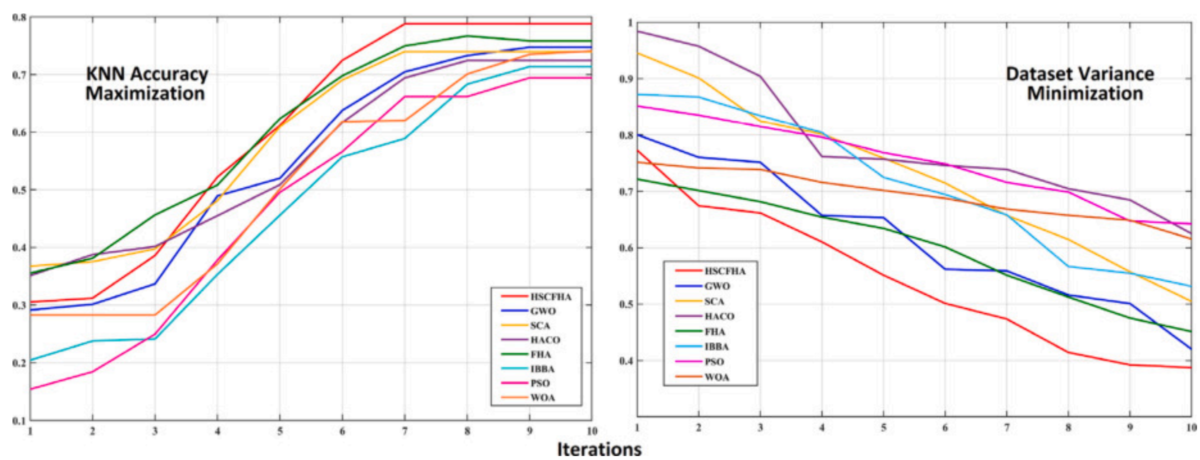


Fig. 9. Average convergence curves from the best values of each dataset for each metaheuristic algorithm.

We further conducted the Wilcoxon Ranked and Friedman statistical tests on the HSCFHA algorithm, compared to other metaheuristic algorithms. The results are detailed in Tables 13 and 14 in the appendix section.

The analysis of convergence curves across 10 iterations reveals a noteworthy trend. HSCFHA stands out in both scenarios, asserting its supremacy in efficiently navigating the feature selection landscape. Fig. 9(b), centered on the task of optimal feature selection for variance minimization, sees HSCFHA's convergence curve steeply descending, a clear indicator of its robust capability in pinpointing the most relevant features while reducing redundancy swiftly. FHA and GWO closely follow, exhibiting a decent convergence rate, though not as rapid as HSCFHA, suggesting their competent but slightly less efficient approach in filtering out optimal features for variance reduction.

Fig. 9(a), on the other hand, shows the trend for the KNN maximization technique. Here again, HSCFHA's performance is better, with its curve ascending prominently, reflecting its prowess in selecting features that enhance the KNN criterion effectively. SCA and WOA demonstrate moderate performance, whereas PSO, IBBA, and HACO trail behind. The latter group's underperformance can be attributed to PSO's simpler strategy, which might be less effective in the complex feature space, and the intricate designs of HACO and IBBA, which do not align well with the specific demands of the feature selection task.

The accuracy of a classification model represents the ratio of correctly classified instances to the total number of instances. On the other hand, precision measures the proportion of instances correctly classified as positive out of all instances classified as positive. It focuses on the correctness of positive predictions. The F1-score is a metric that combines precision and recall by taking their harmonic mean. It provides a balanced assessment of the model's performance, considering precision and recall.

6. Discussion

The datasets utilized for the experiment provide a fair comparison for the analysis of the proposed technique because, between the 22 datasets, the varying size of the dimensions of the data samples and the usage of multi-class output allow for a comprehensive evaluation of the effectiveness of the proposed technique across diverse data types and complexities.

The key advantage that feature selection provides is the removal of redundant features from a dataset in an attempt to reduce the size of the dataset while also keeping the information that the resultant classes represent. From our experimental evaluation, it can be determined that the HACO algorithm takes the most time to gather subset features. This is primarily because the hybridization of the algorithm causes computational complexity to increase while providing better optimal solutions.

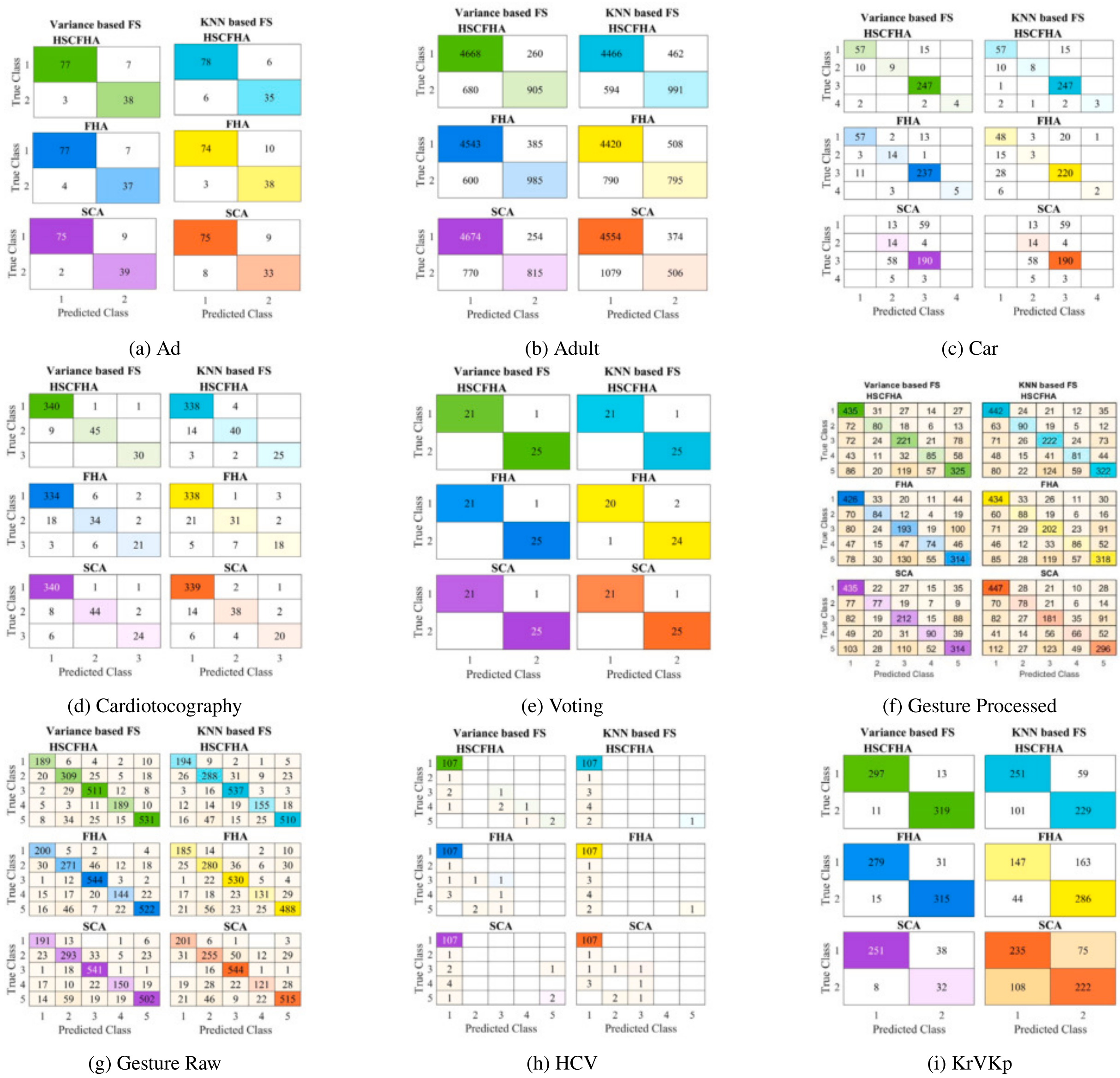


Fig. 10. Comparison of Confusion matrix for dataset (1–9) for HSCFHA, FHA and SCA. Confusion matrix for the remaining datasets are added in the appendix section.

Hybrid metaheuristic algorithms are beneficial because they combine different optimization techniques to overcome the limitations of individual algorithms. In other words, they can leverage the strengths of multiple algorithms while minimizing their weaknesses. The use of multiple algorithms in a hybrid approach can help to improve solution quality, convergence speed, and robustness. By integrating different algorithms, the hybrid approach can explore a larger solution space and find a better solution in less time than using a single algorithm. HACO is further unable to find the global optima for the 22 datasets because when used as an application for feature selection, the exploratory and exploitative phases of the algorithm are diminished because the upper and lower bounds, i.e. 1 and 0, of the methodology remove the erratic movement of the algorithm. While the literature suggests the HACO algorithm is beneficial for constraint optimization problems and np-hard benchmark test functions, it is counterproductive for the application of feature selection due to the added computational complexity.

Moreover, the results of the experiment also reveal that the conventional algorithms i.e., WOA and GWO, outperform the IBBA and HACO algorithms for feature selection. For example, in the case of Cardiotocography, the average accuracy of GWO and WOA is 0.910798 and 0.900067 while the accuracy of HACO and IBBA is lower. Similarly, the average precision for the Obesity dataset is 0.8366256 for the GWO algorithm but is 0.826886 and 0.796044 for IBBA and HACO, respectively, which is lower.

Overall, it can be seen in some of the datasets that HSCFHA did not achieve the highest accuracy, however, it still offered a far lower number of features. For example, the Amazon dataset shows that the HSCFHA, on average, scores an accuracy of 0.261905, while the highest average accuracy recorded was from WOA of 0.275714. Conversely, the number of features selected by HSCFHA was 2847.286, and the WOA gave 5084.571. This shows that even though HSCFHA accuracy is slightly lower than the optimal value, the difference in the number of features selected proves that the proposed algorithm is far better overall for the purpose of feature selection while maintaining the value of the data for class representation.

The HSCFHA, in its theoretical construct, integrates the strengths of both the Firehawk Algorithm and the Sine Cosine Algorithm, aiming for a balance between exploration and exploitation in feature selection. In comparison, some recent hybrid metaheuristic algorithms, such as BIWSO3 [58], IBSCA3 [59] and EOSSA [60] have certain limitations. BIWSO3 primarily focuses on intensifying local search, which might lead to quicker convergence but at the risk of premature optimization. IBSCA3, on the other hand, emphasizes a broader search space exploration, potentially beneficial in avoiding local minima but could result in slower convergence for complex, high-dimensional datasets. EOSSA, with its emphasis on opposition-based optimization, offers a unique approach to maintaining a steady search process. However, it is not as dynamic in adapting to the specific contours of a dataset as HSCFHA, and its complex algorithm architecture might lead to higher computational time requirements, particularly in large-scale feature selection tasks. Each of these algorithms brings distinct approaches to feature selection. By contrast, HSCFHA aims to strike an optimal balance between depth (local optimization) and breadth of search (global exploration), which is critical in handling diverse and complex datasets efficiently.

Additionally, for all datasets, variance-based FS produces more satisfactory accuracy, precision, F1-score, time and number of feature values and a well-balanced confusion matrix as compared to the traditional KNN-based approach. This proves that the proposed feature selection technique is superior, and the hybridized version of FHA and SCA algorithm is more advantageous for the feature selection problem. Hybrid metaheuristic algorithms are beneficial because they leverage the strengths of different algorithms to overcome their limitations and improve solution quality, computational efficiency, and customization to specific optimization problems.

7. Conclusion and future work

The critical role of feature selection, coupled with its NP-hard nature, has led to the increasing popularity of metaheuristic-based methods for feature selection. Our research signified the effectiveness of metaheuristic algorithms in addressing feature selection problems across low-dimensional, medium-dimensional, and high-dimensional datasets. This study introduced a novel approach called HSCFHA, which improves upon the performance of the FHA by integrating the SCA. The hunting mechanism of Firehawks to spread fires and catch prey is utilized in which the movement search space of firehawks is enhanced by the sin cosine movement. By leveraging the strong exploratory capabilities of SCA, the performance of FHA is enhanced, leading to improved feature selection outcomes. To evaluate the efficacy of the proposed HSCFHA method, we conducted experiments using 22 benchmark datasets from the University of California Irvine (UCI) repository and compared its performance against other competitive metaheuristic algorithms, including the original FHA, SCA, GWO, PSO, IBBA, and HACO. The results demonstrated that the HSCFHA method outperforms these algorithms across various datasets. Notably, as the dimensionality of the problems increases, the role of metaheuristics becomes crucial in feature selection, and the performance gaps among different metaheuristic algorithms become more pronounced. This highlights the importance of carefully selecting an appropriate metaheuristic for feature selection in practical problems. Additionally, our study introduced a novel variance minimization-based feature selection approach, as opposed to the traditional KNN accuracy maximization methodology, that improves overall accuracy performance and reduces computational time complexity three-fold while selecting a lower number of features. Furthermore, we have conducted the Wilcoxon and Friedman statistical tests to elaborate on the usefulness of the proposed model. In our future work, we plan to explore the application of the proposed HSCFHA method in different problem domains, such as multi-objective problems, engineering constraint optimization, ML hyperparameter optimization, and multilevel threshold segmentation. Particularly, computer vision applications seem to be a good starting point, where the HSCFHA method can be applied to select features efficiently and enhance object recognition and image classification. While the proposed HSCFHA method demonstrated promising results in feature selection, there are some limitations to be acknowledged. Firstly, the proposed method may not necessarily be the best fit for every dataset and problem, as different datasets may require different feature selection approaches. Secondly, since the algorithm is an iterative ML-based technique, datasets with value of millions of entries may take more than the applicable time to pick out relevant features.

CRediT authorship contribution statement

Syed Kumayl Raza Moosavi: Writing – original draft, Software, Methodology, Conceptualization. **Ahsan Saadat:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Zainab Abaid:** Writing – review & editing, Investigation, Formal analysis. **Wei Ni:** Writing – review & editing, Visualization, Validation, Investigation. **Kai Li:** Visualization, Validation. **Mohsen Guizani:** Visualization, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported in part by the National University of Sciences and Technology (NUST), the CISTER Research Unit (UIDP/UIDB/04234/2020) and project ADANET (PTDC/EEICOM/3362/2021), financed by National Funds through FCT/MCTES (Portuguese Foundation for Science and Technology).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.future.2024.02.017.

References

[1] K. Madhura, R. Mahalakhsmi, End2end unstructured data processing, confidential data structuring & storage using image processing, nlp, machine learning, and blockchain, J. Theoret. Appl. Inf. Technol. 100 (13) (2022).
[2] G. Zhao, D.X. Jiang, et al., Data mining for fault diagnosis and machine learning for rotating machinery, in: Key Engineering Materials, vol. 293, Trans Tech Publications, Ltd., 2005, pp. 175–182.
[3] A.E. Eiben, J.E. Smith, Introduction to Evolutionary Computing, Springer, 2015.
[4] Y. Saeys, I. Inza, P. Larranaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.
[5] N.M. Hijazi, et al., A parallel metaheuristic approach for ensemble feature selection based on multi-core architectures, Expert Syst. Appl. 182 (2021) 115290.
[6] X. Yu, A. Aouari, R.F. Mansour, et al., A hybrid algorithm based on PSO and GA for feature selection, J. Cybersecurity 3 (2) (2021) 117.
[7] P. Stephan, T. Stephan, R. Kannan, et al., A hybrid artificial bee colony with whale optimization algorithm for improved breast cancer diagnosis, Neural Comput. Appl. 33 (20) (2021) 13667–13691.
[8] O. Alzubi, J. Alzubi, M. Alweshah, I. Qiqieh, S. Al-Shami, R. Manikandan, An optimal pruning algorithm of classifier ensembles: dynamic programming approach, Neural Comput. Appl. 32 (2020) 1–17.
[9] O.A. Alzubi, J.A.A. Alzubi, S. Tedmori, H. Rashaideh, O. Almomani, Consensus-based combining method for classifier ensembles, Int. Arab J. Inf. Technol. 15 (1) (2018) 76–86.
[10] G. Kalyani, S. Chaudhari, Data privacy preservation in MAC aware internet of things with optimized key generation, J. King Saud Univ.-Comput. and Inf. Sci. 34 (5) (2022) 2062–2071.
[11] M. Tuba, N. Bacanin, Improved seeker optimization algorithm hybridized with firefly algorithm for constrained optimization problems, Neurocomputing 143 (2014) 197–207.
[12] L. Yu, H. Liu, Feature selection for high-dimensional data: A fast correlation-based filter solution, in: Proceedings of the 20th International Conference on Machine Learning, ICML-03, 2003, pp. 856–863.

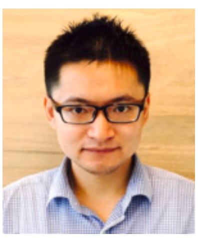
[13] R.T. T. Hastie, et al., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2, Springer, 2009.
[14] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1–2) (1997) 273–324.
[15] Y. Masoudi-Sobhanzadeh, et al., FeatureSelect: A software for feature selection based on machine learning approaches, BMC Bioinform. 20 (1) (2019) 1–17.
[16] Y. Li, C.-Y. Chen, W.W. Wasserman, Deep feature selection: Theory and application to identify enhancers and promoters, J. Comput. Biol. 23 (5) (2016) 322–336.
[17] Y. Bengio, et al., Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.
[18] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[19] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[20] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (1) (2014) 16–28.
[21] T. Hasanin, et al., Investigating random undersampling and feature selection on bioinformatics big data, in: 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications, IEEE, 2019, pp. 346–356.
[22] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[23] L. Gao, J. Song, X. Liu, J. Shao, et al., Learning in high-dimensional multimedia data: The state of the art, Multimedia Syst. 23 (2017) 303–313.
[24] R.E. Neapolitan, et al., Learning Bayesian Networks, vol. 38, Pearson Prentice Hall Upper Saddle River, 2004.
[25] Y. Fan, J. Liu, P. Liu, Y. Du, et al., Manifold learning with structured subspace for multi-label feature selection, Pattern Recognit. 120 (2021) 108169.
[26] M. Gendreau, J.-Y. Potvin, Metaheuristics in combinatorial optimization, Ann. Oper. Res. 140 (2005) 189–213.
[27] I.F. Jr., X.-S. Yang, I. Fister, et al., A brief review of nature-inspired algorithms for optimization, 2013, arXiv preprint arXiv:1307.4186.
[28] A.K. Sangaiah, Z. Zhang, M. Sheng, Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications, Academic Press, 2018.
[29] L. Pereira, D. Rodrigues, et al., A binary cuckoo search and its application for feature selection, in: Cuckoo Search and Firefly Algorithm: Theory and Applications, Springer, 2014, pp. 141–154.
[30] L. Wen, Q. Yin, P. Guo, Ant colony optimization algorithm for feature selection and classification of multispectral remote sensing image, in: IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium, vol. 2, IEEE, 2008, pp. II–923.
[31] O.A. Alzubi, et al., An efficient malware detection approach with feature weighting based on Harris Hawks optimization, Cluster Comput. 25 (2022) 2369–2387.
[32] Y. Wan, M. Wang, Z. Ye, X. Lai, A feature selection method based on modified binary coded ant colony optimization algorithm, Appl. Soft Comput. 49 (2016) 248–258.
[33] J. Jona, N. Nagaveni, Ant-cuckoo colony optimization for feature selection in digital mammogram, Pakistan J. Biol. Sci.: PJBS 17 (2) (2014) 266–271.
[34] P. Moradi, M. Gholampour, A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy, Appl. Soft Comput. 43 (2016) 117–130.
[35] B. Mao, Z. Xie, Y. Wang, H. Handroos, et al., A hybrid differential evolution and particle swarm optimization algorithm for numerical kinematics solution of remote maintenance manipulators, Fusion Eng. Des. 124 (2017) 587–590.
[36] R. Sindhu, et al., Sine–cosine algorithm for feature selection with elitism strategy and new updating mechanism, Neural Comput. Appl. 28 (2017) 2947–2958.
[37] S. Chattopadhyay, et al., Pneumonia detection from lung X-ray images using local search aided sine cosine algorithm based deep feature selection method, Int. J. Intell. Syst. 37 (7) (2022) 3777–3814.
[38] S. Taghian, M.H. Nadimi-Shahraki, Binary sine cosine algorithms for feature selection from medical data, 2019, arXiv preprint arXiv:1911.07805.
[39] M. Belazzoug, et al., An improved sine cosine algorithm to select features for text categorization, J. King Saud Univ.-Comput. Inf. Sci. 32 (4) (2020) 454–464.
[40] Y. Wan, et al., Multiobjective hyperspectral feature selection based on discrete sine cosine algorithm, IEEE Trans. Geosci. Remote Sens. 58 (5) (2020) 3601–3618.
[41] A. Dey, S. Chattopadhyay, et al., A hybrid meta-heuristic feature selection method using golden ratio and equilibrium optimization algorithms for speech emotion recognition, IEEE Access 8 (2020) 200953–200970.
[42] K. Lenin, B.R. Reddy, M. Suryakalavathi, Hybrid Tabu search-simulated annealing method to solve optimal reactive power problem, Int. J. Electr. Power Energy Syst. 82 (2016) 87–91.
[43] M. Mafarja, S. Abdullah, Investigating memetic algorithm in solving rough set attribute reduction, Int. J. Comput. Appl. Technol. 48 (3) (2013) 195–202.
[44] M. Azizi, S. Talatahari, A.H. Gandomi, Fire Hawk optimizer: A novel metaheuristic algorithm, Artif. Intell. Rev. (2022) 1–77.
[45] S. Mirjalili, SCA: A sine cosine algorithm for solving optimization problems, Knowl.-Based Syst. 96 (2016) 120–133.
[46] R. Kommadath, J. Dondeti, P. Kotecha, Benchmarking Jaya and sine cosine algorithm on real parameter bound constrained single objective optimization problems, in: Proceedings of the 2017 International Conference on Intelligent Systems, cec2016, 2017, pp. 31–34.
[47] C. Yue, et al., Problem Definitions and Evaluation Criteria for the CEC 2020 Special Session and Competition on Single Objective Bound Constrained Numerical Optimization, Technical Report 201911, Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China, 2019.
[48] M.H. Zafar, N.M. Khan, S.K.R. Moosavi, et al., Artificial Neural Network (ANN) trained by a novel arithmetic optimization algorithm (AOA) for short term forecasting of wind power, in: Intelligent Technologies and Applications: 4th International Conference, INTAP 2021, Springer, 2022, pp. 197–209.
[49] S.K.R. Moosavi, M.H. Zafar, M.N. Akhter, et al., A novel Artificial Neural Network (ANN) using the mayfly algorithm for classification, in: 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), IEEE, 2021, pp. 1–6.
[50] S.K.R. Moosavi, H.B. Younis, M.H. Zafar, et al., A novel group teaching optimization algorithm based artificial neural network for classification, in: Intelligent Technologies and Applications: 4th International Conference, INTAP 2021, Springer, 2022, pp. 58–70.
[51] A.A. Movassagh, J.A. Alzubi, M. Gheisari, et al., Artificial Neural Networks training algorithm integrating invasive weed optimization with differential evolutionary model, J. Ambient Intell. Humaniz. Comput. 14 (2023) 6017–6025.
[52] UCI archives, 2023, https://archive-beta.ics.uci.edu/. (Accessed 29 March 2023).
[53] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey Wolf Optimizer, Adv. Eng. Softw. 69 (2014) 46–61.
[54] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95-International Conference on Neural Networks, vol. 4, IEEE, 1995, pp. 1942–1948.
[55] S. Mirjalili, et al., Binary bat algorithm, Neural Comput. Appl. 25 (2014) 663–681.
[56] W. Ma, et al., A two-stage hybrid ant colony optimization for high-dimensional feature selection, Pattern Recognit. 116 (2021) 107933.
[57] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67.
[58] N.A. Alawad, B.H. Abed-alguni, M.A. Al-Betar, et al., Binary improved white shark algorithm for intrusion detection systems, Neural Comput. Appl. 35 (2023) 19427–19451.
[59] K.S. Reddy, et al., A new binary variant of Sine–cosine algorithm: Development and application to solve profit-based unit commitment problem, Arab. J. Sci. Eng. 43 (2018) 4041–4056.
[60] S. Pathak, et al., A new salp swarm algorithm for the numerical optimization problems based on an elite opposition-based learning, in: 2021 Asian Conference on Innovation in Technology, ASIANCON, 2021, pp. 1–6.

Syed Kumayl Raza Moosavi received his bachelor's degree in Mechatronics engineering in 2016 and his master's degree in Artificial Intelligence in 2023 from the National University of Sciences and Technology (NUST), Islamabad, Pakistan. He will be pursuing a Ph.D. program in the field of Robotics and AI. His research interests include Robotics, Artificial Intelligence and Swarm Optimization.

Ahsan Saadat is currently working as an assistant professor at School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad, Pakistan. Dr. Ahsan completed his Ph.D. from Macquarie University Sydney Australia in 2017 and has worked in multiple Australian universities. Prior to Ph.D., he completed his MS in Electrical Engineering from NUST in 2013 and BS Computer Engineering from COMSATS in 2009 where he secured two gold medals. His research interests include applied machine learning and game theory.

Zainab Abaid is currently working as an Assistant Professor at the FAST School of Computing, National University of Computer and Emerging Sciences, Islamabad, Pakistan. She completed her Ph.D. from the University of New South Wales (UNSW), Australia, and has worked as a postdoctoral researcher in CSIRO Australia as well as UNSW. Her research interests are adversarial machine learning, applying machine learning to cybersecurity applications, and various domains of cybersecurity such as web security, malware detection and network security.

Wei Ni (Fellow, IEEE) received the B.E. and Ph.D. degrees in Communications Engineering from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He is a Principal Research Scientist at CSIRO, Australia, a Conjoint Professor at the School of Computer Science and Engineering, UNSW, an Adjunct Professor at the School of Electrical and Data Engineering, University of Technology Sydney, and an Honorary Professor at the School of Computing, Macquarie University. He was a Research Scientist and Deputy Project Manager at Alcatel/Alcatel-Lucent Bell Labs (2005–2008) and a Senior Researcher at Nokia (2008–2009). His research interests include machine learning, stochastic optimization, and their applications to system security and integrity. Dr. Ni serves as an Editor for the IEEE Transactions on Wireless Communications and IEEE Transactions on Vehicular Technology. He served as the Secretary, Vice-Chair, and Chair of the IEEE Vehicular Technology Society (VTS) New South Wales Chapter from 2015 to 2022, Track Chair for VTCSpring 2017, Track Co-chair for IEEE VTC-Spring 2016, and Publication Chair for BodyNet 2015. He also served as Student Travel Grant Chair for WPMC 2014, and a Program Committee Member of CHINACOM 2014, IEEE ICC'14, ICCC'15, EICE'14, and WCNC'10.



Kai Li (S'09–M'14–SM'20) received the B.E. degree from Shandong University, Jinan, China, in 2009, the M.S. degree from The Hong Kong University of Science and Technology, Hong Kong, in 2010, and the Ph.D. degree in computer science from The University of New South Wales, Sydney, NSW, Australia, in 2014. He is currently a Senior Research Scientist with CISTER Research Centre, Porto, Portugal. From 2021 to 2022, he was a CMU-Portugal Research Fellow, which is jointly supported by Carnegie Mellon University (CMU), Pittsburgh, PA, USA, and The Foundation for Science and Technology (FCT), Lisbon, Portugal. In June 2022, he was a Visiting Research Scholar with CyLab Security and Privacy Institute at CMU. Prior to this, he was a Postdoctoral Research Fellow with the SUTD-MIT International Design Centre, The Singapore University of Technology and Design, Singapore, from 2014 to 2016. He was a Visiting Research Assistant with ICT Centre, CSIRO, Australia, from 2012 to 2013. From 2010 to 2011, he was a Research Assistant with Mobile Technologies Centre, The Chinese University of Hong Kong, Hong Kong. His research interests include machine learning, vehicular communications and security, resource allocation optimization, cyber–physical systems, Internet of Things, and UAV networks.

Mohsen Guizani (Fellow, IEEE) received the BS (with distinction), MS and Ph.D. degrees in Electrical and Computer engineering from Syracuse University, Syracuse, NY, USA in 1985, 1987 and 1990, respectively. He is currently a Professor of Machine Learning at the Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE. Previously, he worked in different institutions in the USA. His research interests include applied machine learning and artificial intelligence, smart city, Internet of Things (IoT), intelligent autonomous systems, and cybersecurity. He became an IEEE Fellow in 2009 and was listed as a Clarivate Analytics Highly Cited Researcher in Computer Science in 2019, 2020, 2021 and 2022. Dr. Guizani has won several research awards including the "2015 IEEE Communications Society Best Survey Paper Award", the Best ComSoc Journal Paper Award in 2021 as well 5 Best Paper Awards from ICC and Globecom Conferences. He is the author of 11 books, more than 1000 publications and several US patents. He is also the recipient of the 2017 IEEE Communications Society Wireless Technical Committee (WTC) Recognition Award, the 2018 AdHoc Technical Committee Recognition Award, and the 2019 IEEE Communications and Information Security Technical Recognition (CISTC) Award. He served as the Editor-in-Chief of IEEE Network and is currently serving on the Editorial Boards of many IEEE Transactions and Magazines. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker and is currently the IEEE ComSoc Distinguished Lecturer.