




Multiclass feature selection with metaheuristic optimization algorithms: a review

Olatunji O. Akinola¹ · Absalom E. Ezugwu¹ · Jeffrey O. Agushaka¹ · Raed Abu Zitar² · Laith Abualigah^{3,4} 

Received: 19 January 2022 / Accepted: 2 August 2022 / Published online: 30 August 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Selecting relevant feature subsets is vital in machine learning, and multiclass feature selection is harder to perform since most classifications are binary. The feature selection problem aims at reducing the feature set dimension while maintaining the performance model accuracy. Datasets can be classified using various methods. Nevertheless, metaheuristic algorithms attract substantial attention to solving different problems in optimization. For this reason, this paper presents a systematic survey of literature for solving multiclass feature selection problems utilizing metaheuristic algorithms that can assist classifiers select optima or near optima features faster and more accurately. Metaheuristic algorithms have also been presented in four primary behavior-based categories, i.e., evolutionary-based, swarm-intelligence-based, physics-based, and human-based, even though some literature works presented more categorization. Further, lists of metaheuristic algorithms were introduced in the categories mentioned. In finding the solution to issues related to multiclass feature selection, only articles on metaheuristic algorithms used for multiclass feature selection problems from the year 2000 to 2022 were reviewed about their different categories and detailed descriptions. We considered some application areas for some of the metaheuristic algorithms applied for multiclass feature selection with their variations. Popular multiclass classifiers for feature selection were also examined. Moreover, we also presented the challenges of metaheuristic algorithms for feature selection, and we identified gaps for further research studies.

Keywords Feature selection · Classifier · Machine learning · Optimization · Metaheuristic algorithm · Multiclass

✉ Absalom E. Ezugwu
ezugwua@ukzn.ac.za

✉ Latih Abualigah
Aligah.2020@gmail.com

Olatunji O. Akinola
221069617@stu.ukzn.ac.za

Jeffrey O. Agushaka
218088307@stu.ukzn.ac.za

Raed Abu Zitar
raed.zitar@sorbonne.ae

³ Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman 19328, Jordan

⁴ Faculty of Information Technology, Middle East University, Amman 11831, Jordan

¹ School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, King Edward Avenue, Pietermaritzburg Campus, Pietermaritzburg 3201, KwaZulu-Natal, South Africa

² Sorbonne Center of Artificial Intelligence, Sorbonne University-Abu Dhabi, 38044 Abu Dhabi, United Arab Emirates

1 Introduction

Our world is witnessing a vast increase in digital data generation daily. These massive data come in various forms, for example, text, images, numbers, audio, videos, and graphs, which can help humans advance knowledge. Such data are generated in various fields of human endeavors like engineering, healthcare, science, industries, education, and others which can be grouped for insight, prediction, and knowledge purposes. Unfortunately, a significant part of these data are raw, irrelevant, unusable, redundant, and therefore need to be transformed to be suitable for modeling. Before modeling, the data are pre-processed and classified into relevant datasets for the model's training to get insights, make predictions, and contribute to knowledge.

Feature selection problems have become a major real-world problem of reducing the dimension of large datasets available in various scientific endeavors, yet maintaining performance accuracy is paramount. Typically, there is often an insufficient number of objects to sufficiently represent the distribution of the high-dimensional feature spaces [21]. Therefore, reducing dimensionality is a critical issue in numerous scientific fields. Researchers have proposed different approaches to feature reduction in the literature [77, 111]. Feature reduction was classified into two main categories: feature extraction or construction and feature selection [8]. Feature extraction builds some new set of subsets from the original dataset, while feature selection chooses the appropriate subset from the original dataset without any transformation.

Feature selection has posed a significant challenge in machine learning. Based on the increasing time required to locate the best features in a high-dimensional dataset, feature selection is considered an NP-hard problem [279]. To find the best subset, the authors proposed several techniques such as exhaustive search, greedy search, and random search [8]. Many of these techniques are plagued with the problem of converging prematurely, immense complexity, and high computational cost. Hence, metaheuristic algorithms have become more prevalent in dealing with these challenges. Metaheuristic algorithms are, therefore, very efficient and effective in locating the best subset of a dataset and are still able to maintain the model's accuracy. Due to its strength, this study focuses on feature selection problems with metaheuristic algorithms.

In recent years, multiple researchers have proposed various metaheuristic algorithms to solve optimization problems. This study reviews modern metaheuristic algorithms that solve feature selection for multiclass classification problems. Some review works were found in the literature, but none was conducted for the multiclass

feature selection problem to the best of our knowledge. Agrawal et al. [8] conducted a complete review of metaheuristic algorithms for a decade between 2009 and 2019 to solve binary feature selection problems. However, real-world problems are not always dichotomous. A comprehensive survey on evolutionary computation for feature selection was conducted by Xue et al. [261, 266]. Their survey enumerated the state of the art of approach with a focus on genetic algorithm (GA), particle swarm optimization (PSO), ant colony optimization (ACO), and genetic programming.

Also, Jović et al. [119] reviewed feature selection methods with applications. The review focused on feature selection in four application domains: text mining, image processing, computer vision, industrial application, and bioinformatics. However, the study considered only one research work in each domain mentioned, which is not comprehensive. A comprehensive literature review on feature selection algorithms in swarm intelligence was conducted by Brezočnik et al. [32]. The study categorized the sixty-four algorithms examined into eight different taxonomy categories and presented the most common application areas. Similarly, Rostami et al. [203] introduced a comparative analysis of other swarm intelligence-based feature selection methods, considering the strength and weaknesses of this method. Moreover, Abiodun et al. [5] conducted an organized review of evolving feature selection methods for text classification optimization tasks. The scope of their work was from 2015 to 2021, reviewing over 200 articles concerning metaheuristic and hyperheuristic procedures. To this end, the research question below is expressed to achieve this study:

What articles employed metaheuristic algorithms to solve multiclass feature selection between the years 2000 to 2022?

The following questions were formulated to answer the question above.

- What metaheuristic algorithms have been used to solve the multiclass feature selection problems?
- What are the major feature selection methods used by the algorithms identified in (a) above?
- What are the key factors of the methods identified in (b) above are used by the metaheuristic algorithms?
- What are the issues or challenges of metaheuristic algorithms in multiclass feature selection?
- What are the research gaps and future directions in multiclass feature selection problems?

A limited related study has been done on metaheuristic feature selection on the multiclass problem [50, 261, 266], which focuses on evolutionary computation approaches to feature selection and high dimensional feature selection on microarray data, respectively. Therefore, this study reviews

various categories of metaheuristic algorithms in solving multiclass problems within two decades (i.e., 2000–2022), e.g., human-based, physics-based, evolutionary-based, and swarm intelligence-based approaches. The motivation of this study is the lack of literature reviewed to explore the multiclass feature selection problem holistically and since many real-world classification problems are not just binary. Therefore, this study is unique because it presents a new holistic approach to reviewing multiclass feature selection with metaheuristic algorithms. Due to this, we make the following contributions:

1. Presentation of the existing body of knowledge found in the literature for feature selection was examined and compiled, followed by a presentation of classification and lists of metaheuristic algorithms in each category.
2. Systematic literature of multiclass wrapper-based metaheuristic algorithms for feature selection problems is presented. Here, the variations of each examined metaheuristic algorithm are discussed.
3. A presentation of the key factors of the wrapper-based approach like classifier description, datasets used, and their respective evaluation metrics utilized.
4. Issues and challenges of metaheuristic algorithms for multiclass feature selection were presented, and various approaches to solving the challenges were mentioned.
5. Lastly, we presented research gaps and explained future research areas as deduced from the literature to assist the researchers in this domain.

The remaining aspects of this paper are structured as follows: In Sect. 2, the background for feature selection and metaheuristic algorithm is presented. Section 3 presents the methodology used to collect papers for this study. Section 4 provides the various categories of metaheuristic algorithms for solving multiclass feature selection difficulties. After that, Sect. 5 presents various hybrid techniques found in the literature for solving multiclass problems. The study discussed the problems and challenges of metaheuristic algorithms in Sect. 6; Sect. 7 outlines the application areas where multiclass feature selection problems were solved. The review discussion with future directions and other research areas in metaheuristic algorithms are presented in Sect. 8, while Sect. 9, which is the last section, presents this paper's conclusion.

2 Background

Feature selection is a necessary data preprocessing or preparation procedure to illustrate the best relevant, applicable, and essential feature space(s). This approach entails choosing a subset with the utmost discrete and

appropriate feature(s) from a huge class of features for record representation in a dataset for predictive modeling [62]. It is an aspect of feature engineering where the dataset attribute is employed to reduce the dimension of the problem to be tackled and, as a result, ease the classification process phase. The main goal of feature selection is to minimize measurement in large dimensional datasets [5]. Conversely, metaheuristic algorithms contribute substantially to optimizing the challenge of selecting the best or near best optima solution. Therefore, a comprehensive description of the definition concepts with the problem of feature selection and categorization of metaheuristic algorithms will be considered in this section.

2.1 Feature selection

Feature selection has drawn the research community's attention in the past years. It aimed at removing the best features from an original dataset. It was adopted to advance the quality of feature sets in many machine learning responsibilities, i.e., classification, regression, clustering, and prediction of time-series activities [261, 266]. In machine learning activities, feature selection is a crucial and noticeable task. There are different application areas where feature selection was applied in several areas of specialization to solve classification problems, e.g., text mining, image analysis, and biomedical problems. Where significant features exist, there is learning to overfit, which results in their performance degeneration. A central branch of data mining and machine learning research is dimensional reduction techniques, which have been studied and widely adopted strategy for dimensionality reduction among various practitioners, which is aimed at selecting the best subset of relevant features from an original dataset using specific relevant evaluation criteria that produces better performance in learning like better interpretability of model, lower the cost of computational and higher learning accuracy [7]. Figures 1 and 2 show a general concept of feature selection where an original feature set was presented and underwent a selection process. Finally, some relevant or best feature sets are chosen.

Generally, various feature selection methods have been developed to get the optima subsets. Depending on whether the training set is a labeled one or not, we can group feature selection algorithms into supervised [7, 227], unsupervised [58, 160], and semi-supervised [258, 260, 285]. We can further classify supervised feature selection into three: filter, wrapper, and embedded methods. The filter methods isolate feature selection from the classifier learning such that it removes any bias of the learning algorithm from interfering with the feature selection's algorithm [7]. It usually concentrates on the overall characteristics of the data [258, 260].

Fig. 1 Generalization of the concept of feature selection

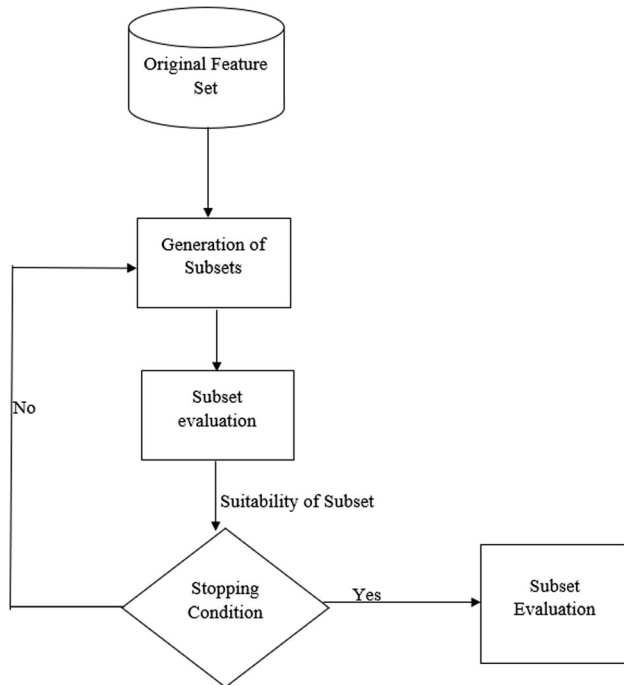
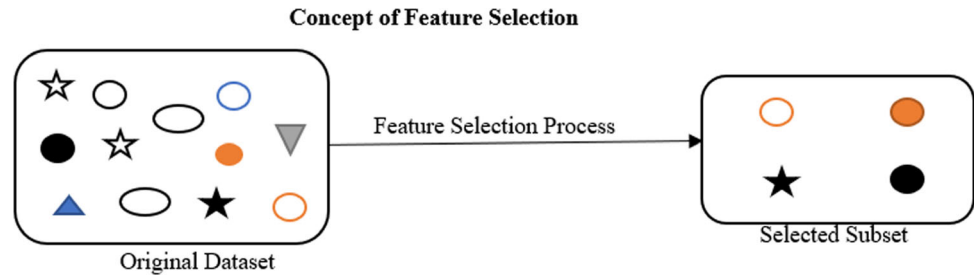


Fig. 2 Illustration of the general feature selection process

On the other hand, the wrapper method predicts the accuracy of the already determined algorithm for learning to generate the selected features' quality. These feature selection models are often luxurious to run on prominent data features. Usually, the wrapper approach includes the classification algorithm, and it interacts with the classifier. Although this method usually presents better results than the filter approach, they are slower and are computationally expensive because they depend on the resource demand of the modeling algorithm [119].

Therefore, the wrapper methods rely on modeling algorithms where every subset is generated and assessed. Generating subset in wrapper method is based on various search strategies. Exponential, sequential, and random techniques are the three search techniques under the wrapper method [119]. Exponential algorithms evaluate several subsets that exponentially increase with the feature space size. This method is almost impossible because of its high computational cost, even though it shows accurate results. Sequential algorithms add or remove features in

sequence (one or more). This may result in local minima once the selected subset is added to or taken from the original dataset, it cannot be further modified. Random algorithms integrate randomness into their search technique, which prevents it from being trapped in local minima. They are often referred to as population-based methods, e.g., metaheuristic algorithm, random generation, simulated annealing, etc. Figure 2 vividly displays the typical feature selection process flowchart from the initial feature set through to subset evaluation. Finally, the embedded approach combines the wrapper and filter methods. In this approach, feature selection is included in the training process and is particular to the learning algorithms [101]. This method may be more efficient in many aspects as they get a solution quicker by avoiding retraining a predictor from the beginning for all the investigated variable subset [85].

We exclude the detailed description of each feature selection method because their detailed explanation is contained in [261, 266]. We show the general wrapper method of the feature selection framework to solve feature selection problems in Fig. 3. The categorization method is shown in Fig. 4, where the spheres in the figure represent the methodology adopted by this paper that defines the way we get to the metaheuristic algorithms.

2.2 Metaheuristic algorithms

Metaheuristic algorithms are generally a higher-level technique that seeks to generate a sufficiently helpful solution to any optimization problem [50] that is complicated and challenging to solve to an optimal level. It has become vital to locate an optimal solution based on insufficient or partial data in the world of inadequate resources such as time and computational capacity. The advent of metaheuristics in finding a solution to this optimization problem has become one of the most remarkable accomplishments of the past decades in research endeavors. Metaheuristic algorithms behave in a stochastic manner because they begin their optimization by picking random solutions. The technique works as a black box, avoiding local optima while exploring the search space effectively and efficiently because of the stochastic nature

Fig. 3 Framework for wrapper feature selection

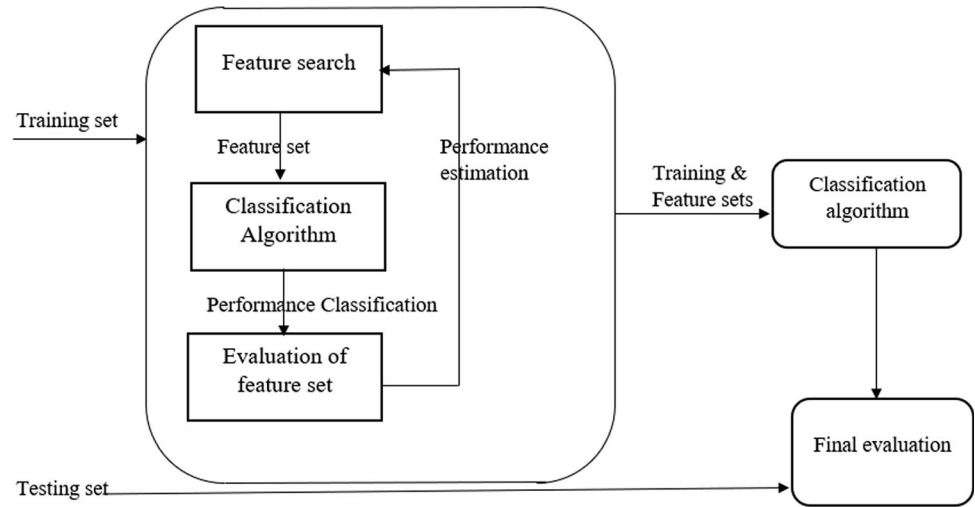
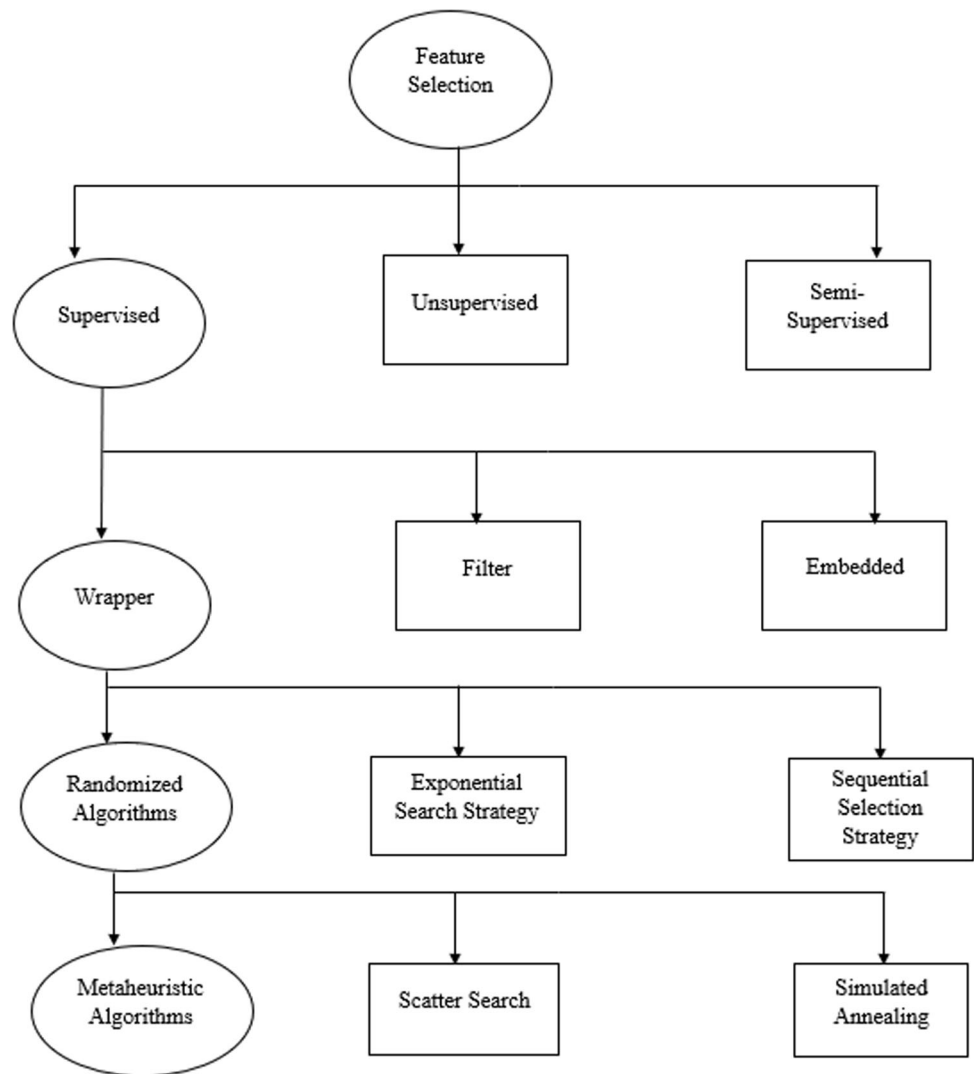


Fig. 4 Categorization of feature selection methods



of the algorithms. One main strength of metaheuristic algorithms is their outstanding ability to avoid premature convergence. Due to its class of global optimization techniques, metaheuristic algorithms have capabilities of both exploration and exploitation. However, they are often found to trap in local optima rather than a global optimum. The primary reason for this is the difficulty of trade-off exploration and exploitation appropriately [257]. These two are also known as diversification and intensification, respectively. In a generation sense, diversification means the ability to search a lot of different sections of the search space. In contrast, intensification implies finding superior solutions within those sections [142]. Although these two are sometimes conflicting goals, a search algorithm should still balance them. A trade-off between exploration and exploitation is substantial for all optimization methods. A compelling trade-off between these two will help to reduce computational cost and device efficient optimization. Metaheuristic algorithms are successfully utilized to solve science and engineering difficulties like civil engineering to design bridges and buildings; electrical engineering to get an optimal solution to power generation; data mining for prediction, classification, clustering, and many more; industrial sector for transportation, job scheduling, facility location problems [8].

The issue of the trade-off between the exploration and exploitation in multiclass feature selection and classification problems received attention from researchers where different methods were adopted to ensure that these conflicting objectives are balanced, such as the work done by Sahebi et al. [207] where their approach made some random changes to the intelligent crossover (IC), intelligent mutation (IM), and their developed inverse operator. The inverse operator was employed to investigate the monitored genes accurately. This introduced an increment in the intelligence of their approach as the exploration and exploitation were done with a greater knowledge of the search space. In Nagpal et al. [170], the study utilized the selection of just the k -best particles as they applied force on each other to maintain an adequate balance between exploration and exploitation. In that study, k was a function of time which decreased linearly to 1, and the study assigned random weights as the coefficient. The two main categories of metaheuristic algorithms are:

a. *Single solution-based metaheuristic algorithms* the goal of most existing feature selection methods is to maximize the classification performance during the search procedure only or combine the classification performance of the feature subset numbers into one objective function [261, 266]. When algorithms work based on one solution per time and involve the local search-based metaheuristics like tabu search, variable

neighborhood search (VNS), and iterated local search [161], this is referred to as a trajectory method. It may often be trapped in local optima due to non-comprehensive exploration of the search space.

b. *Population-based or multiple metaheuristic algorithms* this class of metaheuristics performs a search with numerous initial points similar to swarm-based metaheuristics [161]. The population-based algorithm benefits from scouring the search space for exploration in a useful way. This method is appropriate for searching for global solutions because it can provide global exploration and local exploitation. They are not easily trapped in the local optima because multiple solutions help each other and explore the search space extensively. Due to their benefit of iterating toward the good part of search space is mostly used in real-world problems.

Metaheuristic algorithms have drawn significant attention from researchers to solve feature selection problems in recent decades. In solving this problem, metaheuristic algorithms are categorized into four main aspects depending on their behaviors: evolutionary-based, physics-based, swarm intelligence-based, and human-based [162]. Figure 5 illustrates the categorization of metaheuristic algorithms.

1. *Evolutionary-based algorithms* algorithms in this category are nature-inspired and begin their process by randomly generating their population solutions. The renowned algorithm in this group is the genetic algorithm—GA, which John Holland developed in the 1960s based on Darwin's evolutionary theory. GA has been given much attention with different variants and improvements from researchers and was applied to solve various problems in a real-world situations. Other popular algorithms developed in this group include genetic programming, tabu search, evolution strategy, differential evolution, flower pollination algorithm [270], memetic algorithm [165], and many more.
2. *Swarm Intelligence-based algorithms* are inspired by the social interaction or behavior of birds, animals, insects, fish, and many more. Numerous metaheuristic algorithms have been established within the last two decades in this category, and more are being developed. Several researchers have developed variants of some popular ones, and others have hybridized algorithms from this category. The renowned algorithms in this category are the particle swarm optimization (PSO) algorithm [59]. It was inspired by a group of birds flying through their search space to find the best location. It was developed by Kenneth and Eberhart in 1995 and has gained so much attention due to its rich mathematical basis for solving problems. Other

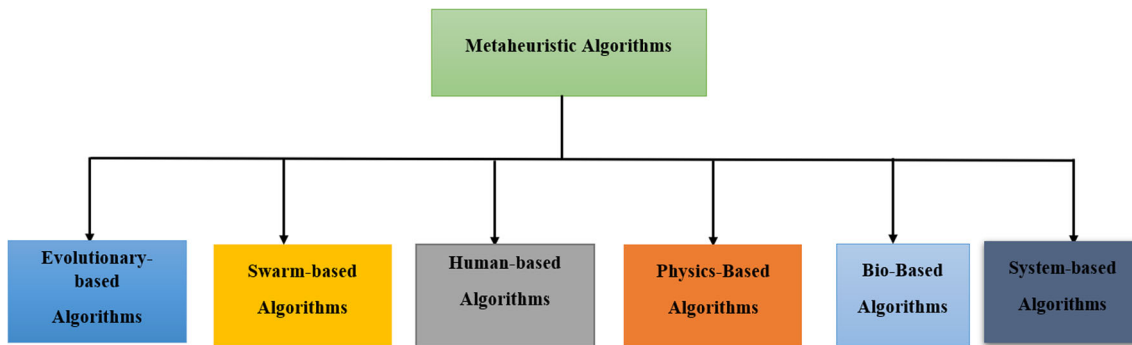


Fig. 5 Classification of metaheuristic algorithms

- algorithms in this group are cuckoo search [267], gray wolf optimizer [154], and krill herd Algorithm [248].
3. *Physics-based algorithms* algorithms that fall in this group draw their inspiration from the laws of physics in the world. Some of the common algorithms in this group include ray optimization [125], gravitational search algorithms [201], galaxy-based search algorithm [100], equilibrium optimizer [72], atom search optimizer [287].
 4. *Human-based algorithms* the algorithms here are inspired by activities performed by or behaviors of humans. Human beings perform various activities that affect their performance, and researchers use these behaviors to develop algorithms. The two most popular algorithms in this category are teaching–learning-based optimization (TLBO) by Rao et al. [200] and league championship algorithms (LCA) by Husseinzadeh Kashan [105]. Others include exchange market algorithm (EMA) by Ghorbani and Babaei [81], seeker optimization algorithm (SOA) by Dai et al. [49], and social-based algorithm (SBA) by Ramezani and Lotfi [197].

2.3 Scientific background

This subsection presents the mathematical background of the feature selection problem and its multiclass approach. The feature selection problem is mathematically represented as follows: If a dataset “s” contains “d” feature numbers, then the problem involves the mechanism for choosing the best subset from “d” features. If we have a dataset $s = \{f_1, f_2, \dots, f_d\}$. This aimed at selecting the best subsets of features from s. Remove subset of $d = \{f_1, f_2, \dots, f_n\}$ where $n < d$ and f_1, f_2, \dots, f_n . This signifies the features of whichever dataset.

In Sánchez-Marño et al. [211], a wrapper-based method of combining ANOVA with functional networks was proposed to estimate the function of f in terms of n input variable, $f = x_1, x_2, x_3, \dots, x_n$ through the

approximation of its component function known as AFN-FS, which serves as our reference point. This same integer function may also be written to be the sum of the 2^n orthogonal summands:

$$f(x_1, x_2, \dots, x_n) f_0 + \sum_{v=1}^{2^n-1} f_v(x_v) \tag{1}$$

Here, v denotes each subset possible with n variable input and f_0 constant corresponds to no argument function. Each of the functional component $f_v(x_v)$ in the above expression is approximated by the AFN technique as:

$$f_v(x_v) = \sum_{j=1}^{2^n-1} c_{vj} p_{vj}(x_v) \tag{2}$$

In the equation above, c_{vj} represent parameters that are supposed to be estimated where p_v represent the orthonormalized basis functions set [35].

A problem of optimization was solved by utilizing the c_{vj} parameters to learn:

$$\text{Minimize } J = \sum_{s=1}^m \epsilon_s^2 = \sum_{s=1}^m (y_s - \hat{y}_s)^2 \tag{3}$$

In Eq. 3, m is the available number of samples and y_s is the output desired for sample s and \hat{y}_s is the output estimated to be attained by:

$$\hat{y}_s = \hat{f}(x_1, x_2, \dots, x_{sn}) f_0 + \sum_{v=1}^{2^n-1} \sum_{j=1}^{k_v} c_{vj} p_{vj}(x_{sv}) \tag{4}$$

Immediately the c_{vj} parameters become learned, it leads to the derivation of two sets of indices called the Global Sensitivity Index (GSI) and Total Sensitivity Index (TSI). The former shows the relevance of the variance in each of the functional components in Eq. 1, while the latter shows each feature’s relevance. These two indices can be expressed as follows:

$$GSI_v = \sum_{j=1}^{k_v} c_{vj}^2, v = 1, 2, \dots, 2^n - 1; TSI_i = \sum_{v=1}^{2^n-1} GSI_v x_i \in v; i = 1, \dots, n. \tag{5}$$

Those two indices allow ascertaining the relevant features individually or combined with the others. To advance its application scope, the AFN-FS was modified due to its limited complexity function in Eq. 4 by employing incremental approximation as follows:

$$\hat{y}_s = f_0 + \sum_{i=1}^n \sum_{j=1}^{k_i} c_{ij}^2 p_j(x_{si}) \tag{6}$$

In Eq. 5, s indicate a particular sample, and the initial feature is represented by n and k_i represent the function set for the estimation of the univariate component of i . The approximation complexity can be increased as soon as some sets of features are cast off by adding other components. This process is iterative to take away as many features as possible. It, however, does not diminish the approximation of the mean accuracy.

Further, since this study focuses on the wrapper approach to multiclass feature selection, we present the mathematical background to the multiclass concept and its problem structure. The multiclass output is denoted by utilizing a distributed representation, that is, giving a set of classes of L , y_s as the output is derived by L elements such that $y_{sl} = 0$ to indicate that it is not included. This changes the optimization challenge of multiclass classification to:

$$J = - \sum_s \sum_{l=1}^L \{y_{sl} \ln \hat{y}_{sl} + (1 - y_{sl}) \ln(1 - \hat{y}_{sl})\} \tag{7}$$

\hat{y}_{sl} is the estimate of the desired output y_{sl} which indicates that sample s might probably belong to class l . The consideration is that the sample belongs to the class having the highest value of \hat{y}_{sl} .

The other approach to multiclass issues entails transforming this problem into many binary ones. The different ways of achieving this task are by using the method of ECOC—error correction output codes [56]. This technique transforms its output by utilizing a matrix $M_{k \times c}$ with the number of classes (c) in the column and the classifiers number (k) as rows. These schemes are often represented [99] as:

- *One-versus-rest* here, one classifier is used for each class ($k = c$) which transforms an issue with c classes into c binary ones.
- *One-versus-one* with 1 vs. 1, each classifier is generated for every class pair. So, we have classifiers $k = \frac{c(c-1)}{2}$

3 Methodology and technique of paper collection

This section explained the procedure employed to select, collect, and review papers. Discussion of keywords or techniques search, databases or data sources, and criteria for inclusion & exclusion were presented. This study followed the systematic review approach [8], and we got a guide from work done by Agushaka & Ezugwu [12].

3.1 Search keywords

The study selected some keywords as our search criteria on the data sources or databases to achieve our review objectives. These include multiclass, feature selection, metaheuristics, and algorithms. Initially, the search activities began in September 2021, and the final search for articles was done in March 2022. The papers from the search output were scanned for relevance to collect additional articles from their in-text citation and reference areas.

3.2 Academic data sources

This study used the appropriate keywords to search and select relevant articles from the available literature. The target was particular to articles published in trustworthy peer-reviewed journals, conference proceedings, and edited books indexed in Scopus, Elsevier, IEEE Xplore, Springer link, Research gate, Google Scholar, and Web of Science databases. Also, the study considered those repositories as having high-quality, highly ranked, and internationally recognized articles published in SCI-index journals and conference proceedings. The number of articles reviewed sums up to 221. The keywords above formed our search bases in the named repositories from 2000 to 2022.

3.3 Inclusion or exclusion criteria

To solely extract relevant articles from the literature, we framed a few criteria for inclusion and exclusion. After scanning the titles, abstracts, methodologies, conclusion, or the entire content in other cases, we include or exclude the articles using the set criteria. Table 1 shows the criteria for selection.

3.4 Eligibility

The eligibility of the articles selected was determined by the set inclusion or exclusion criteria. A total of 40 related articles were selected from Scopus, 67 from Elsevier, 40 from IEEE, 39 from Springer, 13 from Google Scholar, and

Table 1 Criteria for inclusion or exclusion

Inclusion	Exclusion
Articles that utilized the wrapper-based metaheuristic approach to solve feature selection problems	Although we mentioned the various supervised methods, we excluded the work using filter approaches to solve the feature selection problem
Articles published in areas of feature selection between 2000 and 2022 in multiclass situations, either as multiple binaries to solve multiclass problems or as single multiclass problems	Articles employed in binary problem areas within the years specified and the ones used in multiclass situations before the year 2000
Articles from trustworthy peer-reviewed journals, edited books, and conference proceedings from Scopus, Elsevier, IEEE Xplore, Springer link, Research gate, Google Scholar, and Web of Science databases	Articles that were published as part of abstracts, keynote speeches, editorials, and textbooks
Articles that were reported in English Language	Articles that were reported in other languages

27 from Web of Science. On Scopus, 39 articles were published in peer-reviewed journals and one conference proceeding; Elsevier published the highest number of articles; IEEE published 29 peer-reviewed journal articles and 11 articles in conference proceeding; in Springer, 32 articles in peer-reviewed journals and two articles in conference proceeding were included; 25 articles in peer-reviewed journals and two articles in conference proceeding were reviewed from Web of Science, and 11 journal articles and two book chapters were reviewed from Google Scholar which sum up to 221 articles. The number of articles presented here was obtained after applying the inclusion and exclusion criteria. Therefore, more articles were obtained after the search using the aforementioned keywords, which were then excluded based on the exclusion criteria.

4 Metaheuristic algorithms for multiclass feature selection

Most of the classification metaheuristic algorithms are binary. However, several real-world problems are not dichotomous, i.e., 0 or 1, yes or no, true or false, present or absent. Therefore, the need to extend binary feature selection for multiclass becomes essential. Multiclass classification is more complex than a binary variant. Only the decision boundaries of one class are to be known in the binary selection, while in multiclass classification, numerous limitations are involved in binary feature selection. This study considers multiclass variants of metaheuristic algorithms to get the relevant feature. We considered four categories of all multiclass variants of metaheuristic algorithms: evolutionary-based, swarm intelligence base, human-related, physics base, and finally, we examined some hybrid versions that combine two or more

metaheuristic algorithms in solving problems in multiclass feature selection.

4.1 Evolutionary algorithms

We observed that this category of algorithms seems to receive the least attention in solving multiclass feature selection problems. A limited number of these algorithms were developed recently, i.e., from 2000 to 2022. Table 2 shows the list of available metaheuristic algorithms developed within the specified years. The popular genetic algorithm was combined with a support vector machine (SVM) classifier to solve binary or multiclass feature selection problems and parameter optimization in the hospital expense model [235]. Still, its year of development is out of the scope of this study. Although we considered this in the hybrid method. Simon [222] proposed a biogeography-based optimization that was used to solve a real-world sensor selection issue used in the health estimation of aircraft engines. The algorithm's performance was tested on fourteen standard benchmarks and was compared with seven population-based optimization algorithms like GA, DE, ACO, SGA, and PSO. The result shows that BBO and SGA outperformed seven out of fourteen benchmarks. Table 1 presents the list of metaheuristic algorithms for feature selection that falls under evolutionary-based from 2000 to 2022.

4.2 Swarm intelligence

Swarm-based metaheuristic algorithms have recently received the most significant attention from researchers. Swarm intelligence systems comprise a population of simple agents interacting locally and in their immediate environment. They are usually a biological system that draws their inspiration from nature. The agents follow a simple procedure even though no central management

Table 2 List of evolutionary-based metaheuristic algorithms

Algorithm acronym	Name	References
HMO	Honey-bees mating optimization algorithm	Abbass [1]
GBE	Queen-bee evolution	Qin et al. [191]
BBBC	Big bang–big crunch	Erol and Eksin [67]
CMAES	Covariance matrix adaptation evolution strategy	Iruthayarajan and Baskar [108]
IWO	Invasive weed optimization	Rad and Lucas [193]
SWOA	Small-world optimization algorithm	Du et al. [57]
GSO	Group search optimization	He et al. [95]
ICA	Imperialist competitive algorithm	Atashpaz-Gargari and Lucas [25]
GSO	Group search optimizer	He et al. [95]
DS	Differential search algorithm	Civicioglu [42]
BSA	Backtracking search optimization	Civicioglu [44]
CRO	Corel reefs optimization	Salcedo-Sanz et al. [208]
FPA	Flower pollination algorithm	Yang et al. [270]
SFS	Stochastic fractal search	Salimi [210]
SFO	Synergistic fibroblast optimization	Dhivyaprabha et al. [55]

structure controls how the individual agent is meant to behave [213]. A clear benefit here is “autonomy” since the agents are not controlled by external management, and each agent represents a solution to a particular problem. Also, due to their self-coordination, we can say that the swarm is robust, providing no single point of failure. Another benefit deduced from their behavior is “self-organization” [32]. Examples of the renowned algorithms in this group include particle swarm optimization (PSO), ant colony optimization, artificial bee colony (ABC) optimization, and others. Many of these algorithms were recently proven to provide appropriate outcomes in a large type of actual-world applications [59]. Figure 6 shows a general framework for swarm intelligence algorithms that must obey some fundamental phases.

This section describes some swarm intelligence metaheuristic algorithms in selecting multiclass classification

problems with their modifications and the dataset used. Table 3 presents a comprehensive list of SI metaheuristic algorithms developed between 2000 and 2022 to solve feature selection issues. Although this might not be all, it shows the attention given to SI in the last two decades. Moreover, Fig. 7 represents the role of classifiers in feature selection.

4.3 Ant lion optimizer

The ant lion optimizer (ALO) imitates the chasing system of antlions in nature. It was developed by Mirjalili [152, 153] and implemented using five significant steps of chasing prey by the ants: the ant’s random walk, building their traps, catching their prey, entrapping ants, and traps re-building. A new feature selection technique was proposed by Wang et al. [250] based on a modified ALO

Fig. 6 General framework of swarm intelligence algorithms

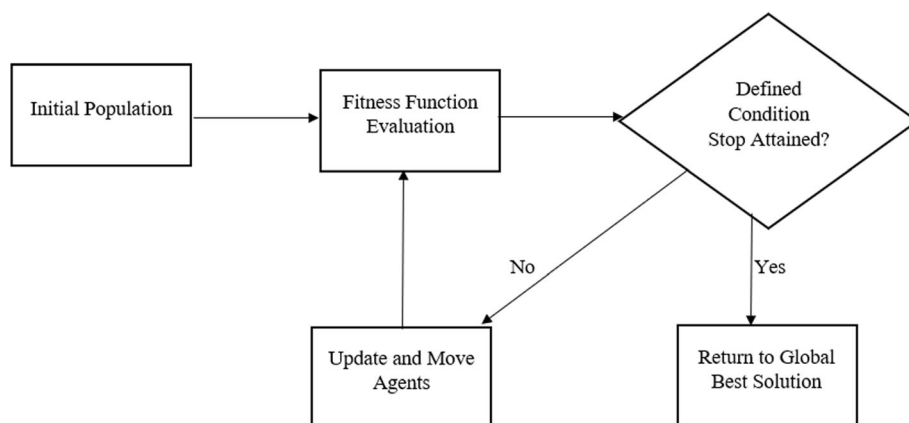
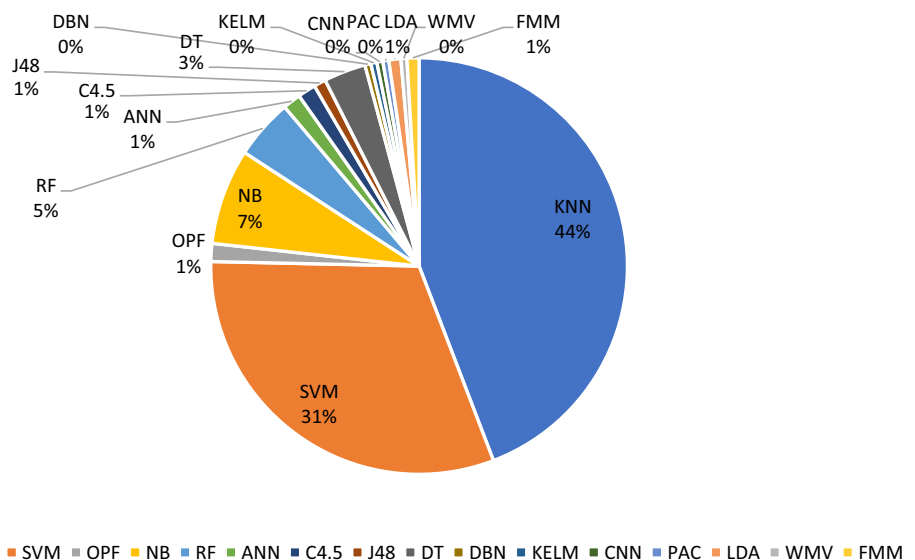


Table 3 List of swarm intelligence-based metaheuristic algorithms

Abbreviation	Algorithm name	References
BFO	Bacteria foraging optimization	Passino [181]
BCA	Bacteria chemotaxis algorithm	Muller et al. [168]
GEP	Gene expression programming	Ferreira [74]
AFSA	Artificial fish swarm algorithm	Neshat et al. [176]
WSO	Wasp swarm optimization	Pinto et al. [182]
CSO	Cat swarm optimization	Chu et al. [40]
ABC	Artificial bee colony	Karaboga and Basturk [121]
MS	Monkey search	Zhang and Liu [280]
FA	Firefly algorithm	Łukasik and Żak [143]
BS	Bacterial swarming	Chu et al. [41]
FS	Fish-school search	Filho et al. [75]
SFS	Stochastic fractal search	Salimi [210]
PFA	Paddy field algorithm	Isaac et al. [109]
BA	Bumblebees algorithm	Comellas and Martínez-Navarro [45]
CS	Cuckoo search	Yang and Deb [267]
GSO	Group search optimizer	He et al. [95]
CGS	Consultant-guided search	Lordache [141]
BA	Bat algorithm	Yang and Gandomi [269]
FA	Fireworks algorithm	Tan and Zhu [232]
FA	Firefly algorithm	Yang [268]
FOA	Fruit fly optimization algorithm	Pan [178]
FPA	Flower pollination algorithm	Yang et al. [270]
ACS	Artificial cooperative search algorithm	Civicioglu [43]
KH	Krill herd algorithm	Gandomi and Alavi [79]
WSA	Wolf search algorithm	Tang et al. [234]
DE	Dolphin echolocation	Kaveh and Farhodi [124]
SSO	Social spider optimization	Cuevas et al. [46]
SSO	Swallow swarm optimization algorithm	Neshat et al. [175]
EVOA	Egyptian vulture optimization algorithm	Sur et al. [230]
GWO	Gray wolf optimization	Mirjalili et al. [154]
SSO	Shark smell optimization	Abedinia et al. [3]
ALO	Ant lion optimizer	Mirjalili [152, 153]
BSA	Bird swarm algorithm	Meng et al. [151]
EHO	Elephant herding optimization	Wang et al. [249]
MFO	Moth flame optimization	Mirjalili [152]
AAA	Artificial algae algorithm	Uymaz et al. [243]
DA	Dragonfly algorithm	Mirjalili [156]
SMO	Spider monkey optimization	Bansal et al. [28]
BSO	Bird swarm algorithm	Meng et al. [151]
CSA	Crow search algorithm	Askarzadeh [24]
WOA	Whale optimization algorithm	Mirjalili and Lewis [157]
MBF	Mouth brooding fish algorithm	Jahani and Chizari [114]
BO	Butterfly optimization	Arora and Singh [22]
GOA	Grasshopper optimization algorithm	Saremi et al. [212]
PBO	Polar bear optimization	Polap and Woźniak [185]
SSA	Salp swarm algorithm	Mirjalili et al. [158]
SHO	Spotted hyena optimizer	Dhiman and Kumar [53]
EPO	Emperor penguin optimizer	Jimoh et al. [117]

Table 3 (continued)

Abbreviation	Algorithm name	References
SSA	Squirrel search algorithm	Jain et al. [115]
BOA	Butterfly optimization algorithm	Arora and Singh [22]
EPC	Emperor penguins colony	Harifi et al. [92]
PFA	Pathfinder algorithm	Yapici and Cetinkaya [271]
HHO	Harris hawks optimization	Heidari et al. [97]
SFO	Sailfish optimizer	Shadravan et al. [218]
SOA	Seagull optimization algorithm	Dhiman and Kumar [54]
BWO	Black window optimization	Premkumar et al. [188]
ChOA	Chimp optimization algorithm	Khishe and Mosavi [130]
MPA	Marine predators algorithm	Faramarzi et al. [73]
MRFO	Manta ray foraging optimization	Zhao et al. [288]
GNDO	Generalized normal distribution optimization	Zhang et al. [282]
RFO	Red fox optimization algorithm	Pořap and Woźniak [186]

Fig. 7 Role of classifiers in feature selection

called MALO and WSVM in reducing the dimension of hyperspectral images. They compared the MALO with other well-known algorithms on some hyperspectral image datasets, and the result showed that the MALO outperformed other methods.

In Zawbaa et al. [277], an optimization technique for the feature selection problems is proposed, which studied the antlion optimizer technique's "chaotic" variance. This chaotic system attempted to advance the trade-off between exploitation and exploration phases. The method was tested with various chaotic maps on some datasets of feature selection. Medjahed et al. [149] presented a method of feature selection and a broad cancer diagnostic procedure using kernel-based learning. The SVM recursive feature elimination (SVM-RFE) was utilized to prefilter the genes.

Next, the SVM-RFE was improved by utilizing binary dragonfly (BDF). Experiments were conducted on six microarray datasets found in the literature. Furthermore, the result demonstrated the approach's efficacy and provided an advanced classification accuracy rate by a reduced number of genes. Moreover [27], a novel search technique for minimal attribute reduction based on rough sets and ALO was proposed. The experiment used University of California Irvine (UCI) repository datasets, and its results indicated that features selected by ALO are well classified with satisfactory accuracy. Emary & Zawbaa [65] modified the ALO and applied it to feature selection with reliance on the Lévy flights called Lévy antlion optimization (LALO) algorithm, which was involved in a wrapper-based model used to pick the combination of an optimal feature that

maximized classification while minimizing the number of features selected.

4.4 Artificial bee colony algorithm

ABC got its inspiration from the intelligent behavior that mimics the bee's food search behavior in the populations. Dervis Karaboga originally developed the ABC in 2005. The base algorithm was employed to perform the local search combined with a random search used for hybrid optimizations. Shunmugapriya and Kanmani [221] proposed a new mixed technique called AC-ABC. The features of ABC and Ant Colony algorithms were combined to improve feature selection. They attempted to eradicate the immobility behavior of the ants, including the associated global search time consumption for original solutions using the chosen bees. The ACO used the exploitation ability of the ABC to ascertain the best ant with the best feature subset, and the bees made use of the generated feature subsets for their food sources. They used thirteen UCI benchmark datasets for the evaluation of their algorithm. Hancer et al. [90] combined a new multi-objective ABC approach based on a feature selection algorithm with the non-dominated sorting technique and genetic operators. They developed two main strategies: ABC with binary representation and ABC with continuous representation. They examined the proposed algorithm performance on 12 benchmark datasets. The results show that the method with the binary model has a better performance in dimensionality reduction and classification accuracy than the continuous model.

Further, Zhang et al. [284] studied a multi-objective feature selection approach known as a two-archive multi-objective artificial bee colony algorithm (TMABC-FS). They proposed two novel operators: diversity-guiding search for bees that are onlookers and convergence-guiding search for used bees to find a class of non-dominated feature subsets with good distribution and merging. Two archives were employed: the external archive and the leader archive, to improve the ability of diverse types of bees to search. The hybrid algorithm was validated using several UCI datasets and was compared with a few traditional algorithms and multi-objective techniques with evidence that TMABC-FS is a robust and efficient method in solving cost-sensitive issues in feature selection. Arslan and Ozturk [23] offered a new method of feature selection based on ABC Programming (ABCP). This approach was referred to as Multi Hive ABCP (MHABCP) to solve a high-dimensional SR problem. They investigated the general performance of the MHABCP and its learning capability using synthetic datasets of actual high-dimensional SR and compared them with other existing methods. The method

outperformed others in choosing relevant features, reducing the data dimensionality.

The authors [252] proposed a quick multi-objective evolutionary feature selection technique, known as FMABC-FS—fast multi-objective ABC. The algorithm was applied to many UCI datasets. The results confirmed that the sample size strategy of the variable is more appropriate to FMABC-FS, with optimal feature subsets having less running time than other algorithms. Almarzouki [18] utilized the ABC algorithm to select lesser genes for the classification of cancer disease. This method employed the CNNs for tumor classification without including labels. In the testing and training phases, three cancer datasets were utilized: kidney, lungs, and brain cancer datasets. This study also presented suggestions on ways to pre-process and modify gene expression further to improve the detection of cancer accuracy in future research.

4.5 Bat algorithm

The echolocation comportment of bats inspired the bat algorithm. This ability to echolocate microbats is interesting as the bats can locate their prey and categorize different insect types in total darkness [269]. Bats send sound waves and receive reflections to find their prey's actual path and location. On reception of the sound waves, the bat that transmits the waves then draws an audio image of its surrounding obstacles and sees them clearly [203]. Since its development in 2011, the bat algorithm has attracted considerable attention, and researchers have modified it and developed its versions to solve feature selection issues. Although a year after its creation, Nakamura et al. [171] produced a binary version called binary bat algorithm (BBA) [171], this review's focus is not on binary classification. Taha et al. [231] presented a hybrid of a nature-inspired bat algorithm and the naive Bayes classifier known as BANB. This approach explored four perspectives: classification accuracy, number of features, feature generalization, and stability. The algorithm's performance was assessed on twelve datasets taken from various domains and then compared with the other three popular algorithms for feature selection. Results indicated that BANB outperformed in choosing fewer features while the classification accuracy was maintained and is more stable in yielding more feature subsets than other methods.

Again, Jeyasingh and Veluchamy [116] proposed a modified bat algorithm (MBA). They used simple random sampling to choose from the subset, some random instances, and remove irrelevant features from an original feature set. They applied the MBA to enhance the classification of radiofrequency to identify breast cancer occurrence. The MBA was modified by using simple random sampling to

choose random instances of the dataset, which reduced the dimension of the feature set, and a random forest trained those selected features for classification. However, using a simple random sample might not be appropriate in all situations and may lead to information loss. Tawhid and Dsouza [237] proposed a hybrid variant of BA with an enhanced PSO, which was used to better the performance of feature selection. They used the PSO algorithm to increase the merging ability of the hybridized algorithm. Saleem et al. [209] developed a modified niche-based bat algorithm (NBBA) with a KNN classifier to solve the feature subset selection issue. The study used over twenty standard UCI repository datasets, and the result showed the capacity of the HBBEPSO in searching feature space. Hammouri et al. [2, 89] adopted the BA to explore features subset optimally and increase the cancer classification accuracy. They combined the wrapper and filter approaches to improve the feature selection performance, where robust mRMR was used as a filter method to choose the highly appropriate feature. The improved BA played the role of the search strategy for the wrapper method of the final elements selected.

4.6 Chimp optimization algorithm

Chimp optimization algorithm (ChOA) is a recently developed metaheuristic method [130] inspired by chimps' sexual motivation and individual intelligence in their collective foraging that is quite different from the other predators. It was originally developed to remove two major problems of local optima trapping and slow convergence in high-dimensional datasets. After it was developed, Wu et al. [256] proposed an enhanced version of the ChOA called enhanced chimp optimization algorithm (EChOA) to solve feature selection problems. This study discovered that despite the four divisions of the hunting strategy baseline algorithm to diversify the population, the algorithm was still limited by local optimum trapping, which led to the EChOA's introduction of highly disruptive polynomial mutation (HDPM) for the population diversity increment. Three strategies were introduced to enhance the population's balance between the exploration and exploitation phases. The study employed support vector machine as a learning algorithm known as EChOA-SVM. The results of the method were compared with other well-known methods and evaluated using seventeen benchmark datasets from the UCI repository, which showed the superiority of the method in terms of higher classification accuracy and lower feature selected on most datasets.

In the following year, Piri et al. [184] proposed a binary multi-objective chimp optimization algorithm (BMO-ChOA) with dual archive and k-nearest neighbors (KNN) classifier to mine relevant medical data aspects. The study

was evaluated using fourteen various dimensions of medical datasets, and the results showed a better performance of the BMOChOA in the features selected and accuracy of performance. The main shortfall of this method is computational complexity and scalability. Furthermore, Pashaei and Pashaei [180] presented two binary versions of the ChOA to tackle the feature selection problem. First, S- and V-shaped transfer functions were utilized to convert the continuous search space to binary, and in the second approach, the crossover operator was used to improve its exploratory behavior. Five high-dimensional biomedical datasets and other datasets from other domains like text, life, and image were adopted to evaluate the performance of the approach. The study was compared with six popular feature selection methods, including PSO, GA, ACO, etc. The method outperformed the other methods in the study in the number of genes selected and the accuracy of classification on most datasets.

4.7 Cuckoo search optimization algorithm

The CSO algorithm was inspired by carefully observing the cuckoo birds' behavior with their reproduction tactic. It is a popular algorithm for solving real-world problems. The way cuckoo birds lay their eggs and reproduction was the basis for forming this algorithm. The CSO algorithm commences with a primary population of this bird as other evolutionary algorithms. The cuckoo's initial population has many eggs in the host bird's nest. Some eggs are more likely to grow into adult birds, while others are recognized and damaged by the host bird. Grown eggs indicate the higher suitability of the area and nest in the search space. The aim of cuckoo as an optimization algorithm is to discover the place for a higher survival rate of most eggs [203]. In 2012, Tiwari [238] proposed a cuckoo-algorithm-based feature selection to solve the face recognition problem. He applied the algorithm to an array of feature vectors extracted by an image's 2-D discrete cosine transform. In the study, the algorithm searched the feature space for an optimal feature and employed the classifier to locate the most matching image in the dataset using Euclidean distance. The algorithm outperformed PSO, which made him conclude that the cuckoo algorithm is more efficient for face recognition. Mousavirad and Ebrahimpour-Komleh [166] introduced a new approach to the cuckoo search algorithm. The best feature subsets were initially chosen using the CSO algorithm, and then KNN was used as the classifier. This proposed algorithm was assessed using some UCI repository's datasets, i.e., Iris, Wine, Pima, Glass Classification, and Breast Cancer which was compared with forward feature selection (FFS) & backward feature selection (BFS), GA-based feature selection (GA-

FS), and PSO-FS. The result indicates an improved performance classification.

Similarly, Elyasigomari et al. [63] proposed a combination of cuckoo optimization and harmony search optimization called minimum redundancy and maximum relevance cuckoo optimization algorithm–harmony search (MRMR-COA-HS) to solve gene selection problems for cancer classification. First, the MRMR was used to select the relevant genes, after which these selected genes were passed into the wrapper system, which combined the novel COA-HS algorithm by utilizing the SVM classifier. They applied the method to some microarray datasets to assess the leave-one-out cross-validation approach. The performance assessment was conducted with other evolutionary algorithms, and it notably outperformed other methods by selecting the lower number of genes while maintaining the highest classification accuracy. Jayaraman and Sultana [113] introduced artificial gravitational cuckoo search algorithm (AGCSA) with particle bee optimized associative memory neural network (PBAMNN) to manage heart disease-related information gotten from the UCI repository of heart disease dataset, which has three hundred and three instances and seventy-five attributed with high dimensionality. The features' dimensionality was selected according to the attribute of the AGCSA with notable reduction, and the selected features were processed by the PBAMNN, which was used to improve the heart disease recognition rate. In Mehedi et al. [150], a method was proposed to improve the multiclass support vector machine (MSVM) classifier's performance by adopting the modified cuckoo search (MCS) called (MCS-MSVM). This approach was applied to minimize the challenge of power quality disturbances. This method showed exceptional 100% classification accuracy when simulated under a noise-free condition and over 98% under various signal-to-noise ratios conditions. This approach was utilized in the electric power network for detesting and classifying power quality disturbance. The MCS reduced the feature dimension and the computational bottleneck by selecting a fewer subset of features.

4.8 Dragon fly algorithm

The dragonfly algorithm (DA) inspired the peculiar flocking compartment of dragonflies in nature. It was developed by Mirjalili [156] and applied to various problems. Zhang et al. [283] propose a new feature selection technique known as IG-MBKH. Information gain (IG) was the basis for a pre-screening feature ranking method, while an improved binary krill herd (MBKH) algorithm was integrated. Its results specified the ability of the IG-MBKH algorithm to achieve convergence improvement, reduced number of feature subsets, and accurate classification

compared to many new algorithms. Cui et al. [48] presented a new algorithm for feature selection called hybrid improved dragonfly algorithm (HIDA), combining the advantages of both improved dragonfly algorithm (IDA) with mRMR to produce a suitable feature subset with a classification of higher accuracy. The performance of HIDA was examined using ten gene datasets and eight datasets from the UCI repository. The result showed the superiority of HIDA. Moreover, in 2021, Chantar et al. [36] also proposed an improved version of the DA as they combined simulated annealing (SA) to resolve the local optima challenge of the DA, and they enhanced the ability of the technique to select the best feature subsets for effective classification. The approach utilized a set of frequently used datasets from the UCI repository to test the performance of the approach. The results showed the superiority of the hybridized approach. Other DA variations and application areas are found [198]. Abdelaziz et al. [2] presented an improved DA for feature selection challenges. This method proposed three variants of the BDA using 19 datasets from the UCI repository, out of which 3 were multiclass datasets and were compared with other literature methods. These approaches, according to the study, outperformed other methods. However, the efficacy of these methods was not tested on more multiclass and large datasets, which will prove their effectiveness in a real-world scenario.

4.9 Firefly algorithm

The firefly algorithm was inspired by the optical connection between fireflies in mating and exchanging information with light flashes and was introduced in 2010 by Yang [268]. Kanimozhi and Latha [120] proposed a new technique in combining the SVM classifier with the earlier proposed binary form of the firefly algorithm to retrieve images. The algorithm was over Corel Caltech and Pascal database images to increase the algorithm's performance with optimal features. Zhang et al. [281] proposed a novel variance of the FA for discriminatory feature selection in regression models and classification to support decision-making procedures by employing approaches to data-based learning. Simulated annealing (SA)-improved local and global solutions, chaotic-accelerated parameters, and diversion mechanisms of feeble solutions to circumvent the possibility of local optimum entrapment and lessen the untimely convergence problem in the baseline FA meta-heuristic algorithm. The study evaluated the efficiency of the proposed model using twenty-nine classifications and eleven datasets of regression benchmark. The technique substantially improved over other well-known FA versions statistically and traditional search methods for various feature selection issues. Chikara et al. [39] presented an

upgraded firefly algorithm DyFA-Dynamic FA for feature selection, which improved convergence rate and reduced computation complexity using dynamic adaptation in blind image steganalysis. The study further reduced the computational complexity by using a hybrid DyFA designed by collaborative filter and wrapper techniques (DyFA), an incremental SVM classifier with radial basis function kernel on ten cross-validations to estimate the efficiency of the DyFA algorithm. Selvakumar and Muneeswaran [215] presented another feature selection technique based on the FA for network intrusion detection issues. The method combines a filter method of feature selection and wrapper methods of feature selection using the C4.5 & Bayesian network to pick the final feature subsets and utilized for KDD CUP 99 datasets. Marie-Sainte and Alalyani [148] proposed a feature selection approach based on the FA. The study employed selected features using an SVM classifier to categorize Arabic texts.

4.10 Flower pollination algorithm

The pollination technique of flowers inspired the flower pollination algorithm (FPA). Some researchers have proposed binary versions of this algorithm, while others modified or combined the binary versions to solve multi-class feature selection problems. In 2015, Zawbaa et al. [278] proposed a multi-objective flower pollination algorithm (MOFPA) optimization hybrid with a rough set for feature selection. It explored the abilities of wrapper and filter-based feature selection techniques. They tested the MOFPA on eight datasets from the UCI data repository. In 2017, Rajamohana et al. [195] proposed an algorithm that extracted features using the adaptive binary flower pollination algorithm (ABFP), a global optimization method applied to spam detection problems. They used naive Bayes (NB) classifier as an objective function. The result showed that the ABFPA technique selected only the informative set and gave a higher accuracy of classification when compared with other competitive techniques. In 2019, Singh and Kaur [226] proposed the FPA algorithm to select optimal features. They used three predefined feature selection algorithms to select the most crucial attribute for detecting anomalies. The performance was compared on over ten features from Kyoto 2006+ for intrusion detection system (IDS).

4.11 Fireworks algorithm

The fireworks (FA) by Tan and Zhu [232] is a swarm-based metaheuristic algorithm inspired by observing fireworks explosions for complex functions of global optimization. Xue et al. [261, 266] proposed a novel mathematical optimization model for supervised classification problems

using FA directly for classification without modification. Based on the training set, a linear equation set was constructed, and an objective function was proposed, which was optimized by FA. 70% of samples were employed as the training set. At the same time, four various datasets were utilized in the experiment. The outcome indicated that the new approach could accurately identify the label set. Xue et al. [264] presented a self-adaptive FA (SaFWA) to solve the optimization classification model competently. The SaFWA utilized four main candidate solution generation strategies (CSGSs) to grow the variety of solutions. They used eight datasets for the experiments to estimate the performance of SaFWA, and the result indicates that the approach is feasible in solving classification problems through SaFWA and the optimization classification model.

4.12 Gray wolf optimizer

The gray wolf optimization algorithm (GWO) was inspired by the chasing procedure of a group of gray wolves in their natural environment [154]. The algorithm emulates the hierarchy of leadership and chasing approach of gray wolves in their natural setting. The GWO has been used recently for solving feature selection problems in data mining. Emary et al. [64] proposed a feature selection method based on multi-objective GWO in searching for the most appropriate and useful features, reducing the feature dimension. The hybrid approach employed the lower computation complexity in the filter method to advance the wrapper method's performances. It was tested using different UCI datasets and achieved much robustness and stability. Li et al. [138] proposed a novel predictive-based framework that hybridized an improved GWO (IGWO) and kernel extreme learning machine (KELM) known as IGWO-KELM and applied it to problems in medical diagnosis. The approach was compared with the base GA and GWO using some well-known disease diagnosis problems using performance metrics: accuracy of classification, the number of selected feature subsets G-mean, specificity, precision, and F-measure. Its result proved to be superior to its counterparts.

Moreover, Too et al. [239] proposed a novel viable binary variant of the gray wolf optimizer (CBGWO) to solve the feature selection challenge in the electromagnetic classification of signals. They extracted some time–frequency features from the STFT coefficient, and the new method was used to evaluate the optimal subset from the initial dataset. Experimental results showed that the CBGWO was superior in feature reduction and classification performance. It also has a very low cost of computation which is appropriate for real-world applications. Sreedharan et al. [229] developed a system for recognizing facial emotions known as facial emotion recognition (FER)

that can analyze essential human facial expressions, like normal, smile, unhappy, angry, amaze, terrified, and irritate. The manner of recognition of the FER system was categorized into four activities, pre-processing, extraction of feature, selection of feature, and classification. After pre-processing, scale-invariant feature transform-based feature extraction was used to extract the features from the facial point. A neural network (NN) based on GWO was utilized to categorize the various emotions from the selected features. Kitonyi and Seger [133] presented a hybridization of a popular metaheuristic optimizer called GWO and gradient descent algorithm to resolve feature selection issues. They first compared the approach with the baseline GWO in twenty-three test functions and developed three binary implementations, and compared the final implementation against two implementations of the binary GWO and binary GWPSO using six medical datasets taken from the UCI repository on the rate of accuracy, the number of features selected subsets, precision, F-measure, and sensitivity metrics. A newly proposed hybridized technique comprised the extended binary cuckoo search, genetic algorithm, and whale optimization algorithm, which aimed to reduce the time required to search a huge database during image retrieval. This approach was compared with other popular classification algorithms like KNN, NB, random forest—RF, CatBoost, considering recall, precision, error rate, F-measure, etc. [118].

4.13 Grasshopper optimization algorithm

Grasshopper optimization algorithm (GOA) was developed by Saremi et al. [212]. It was modeled mathematically, mimicked the grasshopper's swarm behavior in nature, and was applied to solving challenging problems in structural optimization. Aljarah et al. [16] proposed a hybridized approach based on the GOA to enhance the SVM model's parameters and simultaneously find the optima features subset. The method employed eighteen different dimensional benchmark data to test its accuracy. They assessed the performance of the proposed approach with seven other popular algorithms. The results indicated that this approach outperformed other methods in most datasets regarding the classification accuracy and reduced the number of feature subsets selected. Zakeri and Hokmabadi [276] proposed a new feature selection method called GOFs, based on mathematically modeling the interaction between grasshoppers to find food sources. They modified the base GOA to ensure its suitability for feature selection. They were enhanced by statistical measures in the iteration processes to change the same features with the most promising ones. The approach used various publicly available datasets to test its performance. The results produced by this method were compared with twelve other

prominent forms of feature selection with indications of the significance of the GOFs. The study in Khurana and Verma [132] proposed a combination of tuned grasshopper optimization algorithms with classifiers. Their metaheuristic method aims to determine the significantly reduced feature subset from all features and improve the classification performance. They used the random search technique for tuning the classifiers and adopted KNN & SVM. They evaluated the performance of the approach using five multiclass datasets to test accuracy and the area under the curve—AUC. They computed the results with some cross-validation techniques. The researchers compared the result of the proposed method with other algorithms, which showed the technique's performance by outperforming all the compared prominent methods.

4.14 Harris hawk optimization

After the HHO was proposed in 2019, some researchers applied, improved, and hybridized it to solve the feature selection problems. The studies by Hu et al. [102] and Wei et al. [254] presented a hybrid approach of HHO and KELM. Wei et al. worked on an intelligent prototype to predict students' entrepreneurial intention. The HHO was utilized to improve the KELM model. Afterward, Gaussian barebone was used to improve the HHO algorithm to empower the ability of optimization to tune KELM's parameters and identify the compressed feature sets. The method was referred to as GBHHO-KELM. The study adopted the use of thirty benchmark problems of CEC 2014 and re-evaluated some popular techniques. The results showed that the GBHHO-KELM achieved higher stability and better classification accuracy. However, Hu et al. combined improved a binary version of the HHO known as HHSORL and applied the study to predict the severity of Covid-19 virus infection. Hussain et al. [104] later proposed a hybrid version of HHO with the sine-cosine algorithm. The SCA was employed to solve the ineffective exploration of HHO, and the new method was called SCHHO, which was tested with other well-known methods. The outcome increased the convergence speed and produced significant search results with no computational cost added. Qu et al. [192] proposed a hybrid feature selection technique known as VNLHHO-variable neighborhood learning Harris hawks optimizer. They also presented a new activation function to change the incessant solution of VNLHHO into binary values. The NB classifier was used to select genes that can assist in classifying tissues of binary and multiclass cancers.

4.15 Krill herd algorithm

The individual krill behaviors of herding inspired the krill herd (KH) algorithm. This algorithm is based on swarm intelligence and bacterial foraging algorithms introduced by Gandomi and Alavi [79]. The krill movement's objective function finds each krill's minimum distance from food and the herd's highest density. After it was developed, other modifications and hybridization were proposed for feature selection problems. In Wang et al. [247], the study proposed a series of chaotic particle swarm krill herd (CPKH) algorithms to solve optimization tasks within restricted time situations. Various 1-D chaotic maps were used instead of the method's parameter. They assessed the approach on thirty-two benchmark functions and a gear train design difficulty, and the results revealed the accuracy and effectiveness of the CPKH method. Hafez et al. [87] proposed a hybridization method of monkey algorithm (MA) for feature selection and KHA called MAKHA. The system was developed to quickly search the feature space for a near-optimal subset and minimize a given fitness function. This method was utilized to choose the optimal combination of features, thereby reducing the datasets' data dimension to increase the performance classification and select fewer feature subsets. The fitness function employed focused on classification accuracy as the main objective and data size reduction as secondary. It was evaluated on eighteen datasets and proved its advancement over other methods. In Rani and Ramyachitra [199], the fish swarm optimization algorithm with SVM and random forest RF techniques for cancer feature selection and classification reduced only a few features from the datasets. Next, an enhanced krill herd optimization (KHO) technique was used to select the genes, and the RF technique was utilized to categorize the types of cancer. The random forest classification was used for its classification accuracy. They tested the efficiency of this method on ten different gene microarray cancer datasets. The KHO/RF method outperformed the other methods with 100% accuracy of results for most datasets.

4.16 Polar bear optimization

The PBO algorithm was created by Polap and Woźniak [185], with its inspiration drawn from the survival of polar bears in hunting for food in unsuitable weather conditions. The study modeled the polar bear behavior as optima solutions' search engine. The simulated adaptation of polar bears during harsh winter became an advantage for the search exploration and exploitation phases, and the population control was the birth & death mechanism. From the literature, we discover that not much work has been done in

developing other variants of the PBO with application to feature selection. In Haq et al. [91], a binary polar bear optimization was designed to solve the combinatorial problems of scalable unit commitment (UC). The method was also employed to solve the economic dispatch problem using the conventional lambda iteration method. The work in Mirkhan and Çelebi [159] proposed a hybrid technique of rough set and polar bear optimization to find optima feature reducts which is the main goal of feature selection. The rough set theory is a potent technique for measuring the influence of every attribute in a dataset and the effect of removing an attribute on the accuracy of the dataset. The heuristic algorithm plays an important role in avoiding the evaluation of all combinations of features. This method harnesses the advantage of a polar bear by utilizing a dynamic population and its birth and death mechanism to quickly locate optimal solutions by removing irrelevant candidates and retaining the promising ones. The method was able to find better optimal reducts considering the size of the population, time of execution, and number of iterations.

4.17 Red fox optimization

The red fox optimization was developed by Polap and Woźniak [186] and was inspired by the hunting methods in nature of the red fox. The study modeled the exploration phase using the red fox's territorial search for food when they spot a prey afar off as global search and exploitation using movement within the habitat to get closer to the prey before attaching as the local search. This study was initially applied to solve the optimization problem in the engineering field. The algorithm has then been used to tackle feature selection problems. After its development in 2021, Khorami et al. [131] proposed a hybrid red fox optimization with a convolutional neural network to detect the COVID-19 virus from sets of X-ray images. Their approach examined chest X-ray images using a new pipeline machine vision-based system for more precise results. After the X-ray images input was pre-processed in their method, they segmented the region of interest. After that, a combination of gray-level co-occurrence matrix (GLCM) and discrete wavelet transform (DWT) features was extracted from processed images. The results of the method suggest adequate efficiency in diagnosing the COVID-19 virus compared to other methods in the literature.

Moreover, Vaiyapuri et al. [244] developed a hybridized red fox optimizer with a deep-learning-enabled microarray gene expression classification (RFODL-MGEC) method. Genes were the features, and the study aimed to advance classification performance by choosing suitable features. The RFO was utilized to get optima feature subsets, and a bidirectional deep neural network was employed to classify

multiarray gene expression. The neural network parameters used were tuned optimally using the CGO algorithm. The approach was able to perform satisfactorily on the datasets used. However, this method was not used on large-scale datasets, which may reduce the model's performance. Furthermore, Fu et al. [78] presented an improved variance of the red fox, called developed red fox optimization (DRFO), which was applied to detect skin cancer. Like that of Khorami et al. [131], this study employed a pipeline method to accurately diagnose melanoma from some dermoscopy images. They segmented the region of interest using the kernel fuzzy C-means approach after pre-processing of images was done. Lastly, an optimized classification technique, a multi-layer-based perceptron, was used for the final diagnosis.

4.18 Salp swarm optimization

The salp swarm algorithm (SSO) is a recently developed algorithm inspired by nature [158] to solve different optimization issues. The base algorithm's inspiration was inspired by the flooding behavior of salps crossing and hunting in aquatic habitats and was employed to solve engineering difficulties. Ibrahim et al. [106] presented a hybridized optimization technique for feature selection problems. The proposed algorithm combined the SSO algorithm and the PSO called SSAPSO to improve the efficacy of both exploitation and exploration phases. They tested the performance of the SAPS using two investigational sequences. They compared the first with related methods using benchmark functions. In contrast, the other was used to determine the best feature set and remove irrelevance ones from the original datasets using various datasets from the UCI ML repository.

Also, [242] proposed a technique for selecting optimal feature subset in the wrapper method and solving feature selection problems. They included two enhancements into the base SSA: based learning at the starting phase of SSA to improve its population diversity in the search space. Secondly, it included developing and using a new local search algorithm with SSA to enhance its exploitation. The KNN classifier was used in the training data as a fitness function. To validate the performance of the enhanced SSA (ISSA), they applied it to eighteen datasets from the UCI repository. They compared it with four common optimization algorithms and four criteria of assessment. The result showed that ISSA performed better than all the base algorithms in the fitness function accuracy, reduction of features in most datasets, and convergence curve. Neggaz et al. [172] developed a new version of SSA for feature selection known as improved follower of salp swarm algorithm, which used the sine cosine algorithm and disrupts operator (ISSAFD), to update the followers' position

in the SSA by utilizing mathematical functions of sinusoidal as inspired by the sine cosine algorithm (SCA). Twenty datasets were evaluated, four of which were multiclass, higher dimensional. The results showed the efficacy of the ISSAFD in reducing feature dimension and fewer selected features, specificity, accuracy, and sensitivity. The enhancement improved the exploration phase and avoided getting stuck in the local zone. Hegazy et al. [96] improved the structure of basic SSA to enhance the solution accuracy, reliability, and convergence speed and was called ISSA. Inertia weight was added as a new control parameter to adjust the best solution. This new method was tested in the feature selection task and was merged with the KNN for feature selection, where twenty-three datasets from UCI were employed to test the performance of the ISSA algorithm. Both Tubishat et al. [242] and Hegazy et al. [96] did similar work in the same year on the same SSA. However, with a few differences in their methods, both called their improved variant ISSA. Jain and Dharavath [112] presented a feature selection technique that improved the SSOA-salp swarm optimization algorithm called memetic-MSSOA, which they transformed into binary to get the best classification accuracy. They compared the efficacy of the MSSOA with the other five metaheuristic algorithms on UCI datasets. This approach was applied to detect plant diseases with superior performance.

4.19 Whale optimization algorithm

The WOA was inspired by the bubble-net chasing strategy of humpback whales and was proposed by Mirjalili and Lewis [157]. This algorithm comprises three operators in simulating the pursuit of their prey, bubble-net foraging, and encircling prey behavior of the humpback whales. This algorithm has been used to solve many optimization problems in recent times. Mafarja and Mirjalili [146] proposed a novel feature selection based on WOA's wrapper method. They discovered that the algorithm had not been thoroughly applied to feature selection problems. They proposed two binary modifications of the WOA algorithm to explore the best feature subsets as classification, which are (1) roulette wheel and tournament selection mechanisms against a random operator in the process of searching, and (2) mutation and crossover operators were employed in enhancing the WOA algorithm's exploitation phase. They used twenty benchmark datasets in the approach.

Earlier in 2017, the same researchers proposed a hybrid WOA with a simulated annealing technique for feature selection. Their aim for adopting the simulated annealing in their approach was to ensure enhanced exploitation by searching the most capable areas that the WOA algorithm can locate. Nematzadeh et al. [173] presented a filter

method of feature selection outside the scope of this study. Zheng et al. [289] presented a novel hybrid algorithm for feature selection known as the maximum Pearson maximum distance increased whale optimization algorithm (MPMDIWOA). In the first instance, a filter algorithm was proposed named maximum Pearson maximum distance (MPMD) based on Pearson coefficient and correlation distance. Next, parameters were projected in MPMD to fine-tune the weights of redundancy and relevance. Secondly, the revised whale optimization algorithm acted as a wrapper algorithm. They verified this method using ten benchmark UCI machine learning datasets. The outcomes indicated that the algorithm's classification accuracy was meaningfully higher than the other compared algorithms.

In Bui et al. [33], a hybridization of the WOA and adaptive neuro-fuzzy inference system approach called WANFIS was proposed to solve feature selection and land pattern classification problems. The capital of Vietnam, Hanoi, was selected as a case study due to its complex surface. They compared the model's performance with many benchmarked classifiers using standard indicators like the Kappa index and receiver operator characteristics. The result showed the outperformance of WANFIS over other approaches. In Mandal et al. [147], the researchers presented a three-stage framework for feature selection using a wrapper-filter method to detect medical-related diseases. The approach adopted three classifiers: naïve bay, SVM, and KNN, with the whale optimization algorithm as the wrapper-based feature selection method to reduce the feature subset and achieve higher accuracy at stage three of their approach. Earlier in their approach, the filter technique was introduced, including Chi-square, ReliefF, and mutual information. Later in stage two, they applied the XGBoost algorithm to get the best feature set. The efficacy of their approach was evaluated on UCI datasets, and the results displayed better performance over other popular techniques. In Too et al. [241], the authors developed two variants of the WOA called spatial bound whale optimization algorithm (SBWOA) and its simplified version S-SBWOA in solving the multiclass high-dimensional problem of feature selection. The study utilized sixteen high-dimensional feature selection datasets from the Arizona State University repository. The two variants outperformed other methods in producing higher validation accuracies, least mean fitness values, standard deviation, selected features, and computational time. This study showed significantly reduced features, increased accuracy, precision, and F-measure compared to the other eight methods. However, the study could adopt other popular classifiers apart from the kNN in the future to enhance the validation accuracy.

4.20 Human related

This section presents a summary of human-based metaheuristic algorithms for multiclass feature selection. Since human activity defers, researchers have developed and still propose various algorithms that depict the actions of humans to solve complex problems. Although the literature shows limited metaheuristic algorithms developed under this approach, the most popular are the teaching-learning optimization, brainstorm optimization, and league championship algorithms. This section will examine some of these algorithms applied to solve feature selection problems and their application areas. Table 4 presents the list of human-based metaheuristic algorithms developed to solve feature selection challenges over the years of consideration.

4.21 Brainstorm optimization

Brainstorm optimization (BSO) was proposed by and inspired by the human brainstorming process [220] and was applied to data classification. Pourpanah et al. [71] proposed a novel hybridized BSO based and the fuzzy ART-MAP (FAM) model called FAM-BSO feature selection and optimization technique for classification problems. The researchers employed ten benchmark challenges and a case study in the real world to appraise the hybrid's ability. They statistically quantified the results using the bootstrap approach with very high confidence intervals, which showed promising results compared with the original FAM and other procedures. Yun-Tao et al. [275] proposed a new two-phase evolutionary feature selection technique called clustering-guided integer brainstorm optimization algorithm (IBSO-C). The study introduced a new strategy and an integer update scheme for improving the search performance of individuals in BSO. They compared the results with many existing algorithms using real-world datasets. The results indicated that IBSO-C could select fewer feature subsets with high classification accuracy at a less computational cost. Furthermore, Song et al. [228] developed the adaptive mechanism-based BSO (ABSO) algorithm built on chaotic local search. They tested its performance on twenty-nine benchmark functions to ascertain its effectiveness and stability and compared the results with other optimization algorithms. Their results proved that ABSO outperformed those five algorithms considering its stability and convergence accuracy.

4.22 Gaining sharing knowledge-based algorithm

The GSK mimics acquiring and disseminating knowledge in the human lifespan. It was developed by Mohamed et al.

Table 4 List of human-based metaheuristic algorithms

Abbreviation	Algorithm name	References
SWOA	Small-world optimization algorithm	Du et al. [57]
LCA	League championship algorithm	Kashan [122]
SEOA	Social emotional optimization	Xu et al. [258, 260]
BSO	Brainstorm optimization	Shi [220]
TLBO	Teaching–learning-based optimization	Rao et al. [200]
ASO	Anarchic society optimization	Ahmadi-Javid [19]
EMA	Exchange market algorithm	Ghorbani and Babaei [81]
JA	Jaya algorithm	Venkata Rao [245]
GSK	Gaining sharing knowledge-based algorithm	Mohamed et al. [163]
SAR	Search and rescue optimal planning system	Kratzke et al. [134]
CHIO	Coronavirus herd immunity optimization	Al-Betar et al. [13]

[163] for solving continuous space optimization problems. It was developed based on two essential stages: junior and senior gaining and sharing phases. The performance of GSK was verified and analyzed using thirty test problems taken from the CEC2017 benchmark with ten recent well-known metaheuristic algorithms. Results showed the robustness, convergence, and quality of GSK delivering excellent performance in solving optimization problems, particularly on dimensionally severe issues. Agrawal et al. [10] presented an improved form of the GSK algorithm to search for optimal feature subsets. They first represented a binary version of the GSK utilizing a probability estimation operator (Bi-GSK) on the main pillars of GSK and introduced a chaotic map to improve the performance of Bi-GSK. They used twenty-one UCI repository datasets to test the performance of the Bi-GSK, with its results showing that Chebyshev's chaotic map portrayed an improved performance accuracy and convergence rate. It also outperformed other metaheuristic algorithms considering fitness value, efficiency, and the limited number of selected feature subsets, thereby reducing the dimension of the original datasets. Agrawal et al. [11] proposed the GSK algorithm over continuous search space with a total of eight S-shaped and V-shaped transfer functions used to solve binary search space problems. They tested performance on twenty-one UCI repository benchmark datasets and compared the results with some commonly used metaheuristic algorithms. They performed two nonparametric tests to investigate the results, which showed superiority statistically.

4.23 Teaching–learning-based optimization

Rao et al. [200] developed the teaching–learning-based optimization (TLBO) to optimize the problems in mechanical design. The TLBO works on the consequence of teachers' impact on their students. This population-based approach ensures that solutions move to a global

solution. They evaluated the algorithm's effectiveness on five diverse controlled benchmark test functions with different characteristics and real-world applications. Since it was initially proposed, researchers have developed some versions and hybridization to solve feature selection problems. Allam and Nandhini [17] utilized the TLBO for optimizing features in automatically diagnosing breast disease. They employed the naive Bayes classifier to find the individual's fitness and multilayer perceptron (MLP), J48, random forest with logistic regression algorithms to estimate its efficacy. The results confirmed that the scheme generated a higher rate of accuracy on the Wisconsin Diagnosis Breast Cancer (WDBC) dataset in categorizing the benign and malignant tumors. Sevin and Dökeroglu [217] hybridized the TLBO algorithm and extreme learning machines (ELM) called TLBO-ELM to solve data classification problems under which feature selection falls. It was tested on some set of UCI benchmark datasets. Its performance was proven viable for binary and multiclass classification of data problems compared with some commonly used algorithms.

Das et al. [51] proposed FSTLBO, a TLBO based on a feature selection method to find the optimal feature subsets. This method revises the weak features with strong ones, and the results indicate a substantial enhancement when performance was considered compared to some other feature selection models. The method fulfilled its objective of increasing performance, reducing the computational cost, increasing accuracy, removing irrelevant data, and helping faster model learning. Muhammad et al. [167] presented a novel text feature selection technique that used an amalgamation of rough set theory (RST) and TLBO known as RSTLBO. They developed four frameworks in the RSTLBO: the acquisition of standard datasets, dataset pre-processing, utilizing the RSTLBO method; selected feature set used by employing the SVM technique. The result showed that the algorithm produced an improved sentiment analysis. Rajinikanth and Pavithra [196] presented a

density-based modified teaching–learning-based optimization (DMTLO) to select features, KNN to cater for NaN values, and classification was done by SVM & Ensemble. The result proved that the DMTLO outperformed the existing methods by generating the required number of attributes. Some binary variants of this algorithm and application areas of TLBO are not covered in this study, and an example is a work by Chen et al. [38].

4.24 Physics base

Physics-based metaheuristic algorithms were established based on the inspiration of physics laws. Some of these algorithms were applied in solving feature selection problems in different application areas. This section examines a few of these physics-based algorithms and their variants that were proposed in solving feature selection problems in different areas of application. Table 5 presents the list of physics-based metaheuristic algorithms for feature selection issues.

4.25 Equilibrium optimizer

EO is a new metaheuristic algorithm in the physics-based category based on inspiration from control volume mass balance models used to estimate equilibrium and dynamic states and was developed to solve various engineering problems [72]. This algorithm is referred to as one of the very influential, fast, and best performance population-

based algorithms for optimization. A new wrapper-based feature selection algorithm and chaos theory [214] was proposed. In this approach, chaos theory's principles were employed to overcome the slow rate of convergence and its shortcoming of getting entrapped in local optima, which exists in the original EO. Therefore, the approach embeds ten different chaotic maps in optimizing the EO to overcome the difficulties and accomplish a more robust and efficient search mechanism. The researchers also used eight different S & V-shaped transfer functions and tested the performance on fifteen benchmark datasets and four large-scale NLP UCI repository datasets. The result showed that this technique is highly competitive in finding optimal feature subsets.

In Elgamal et al. [61], the study produced a novel metaheuristic optimizer called improved equilibrium optimization algorithm (IEOA) with two significant advancements in the original EO. The first applied elite opposite-based learning (EOBL) to improve its population's diversity, while the other integrated three new local search strategies to prevent its local optima trapping problem. The IEOA enhanced the population's diversity and the classification accuracy, feature subsets selected, and an increased convergence speed. They tested the performance on twenty-one biomedical benchmark UCI repository datasets. The results then showed the outperformance of IEOA over the original EO and other algorithms for most of the datasets used. Moreover [177], the researchers presented a hybrid feature selection approach that is based on the

Table 5 List of physics-based metaheuristic algorithms

Abbreviation	Algorithm name	References
GSA	Gravitational search algorithm	Rashedi et al. [201]
CSS	Charged system search	Kaveh and Talatahari [127]
GbSA	Galaxy-based search algorithm	Hosseini [100]
EMO	Electromagnetic optimization	Cuevas et al. [47]
ACROA	Artificial chemical reaction optimization algorithm	Alatas [14]
Spiral	Spiral optimization	Tamura and Yasuda [233]
BA	Black hole algorithm	Hatamlou [94]
WSA	Water cycle algorithm	Eskandar et al. [68]
RO	Ray optimization	Kaveh and Khayatazad [125]
MBA	Mine blast algorithm	Sadollah et al. [205]
CBO	Collision bodies optimization	Kaveh and Mahdavi [126]
SCA	Sine cosine algorithm	Sindhu et al. [224]
MVO	Multi-verse optimizer	Mirjalili et al. [155]
EFO	Electromagnetic field optimization	Abedinpourshotorban et al. [4]
LAPO	Lightening attachment procedure optimization	Nematollahi et al. [174]
TEO	Thermal exchange optimization	Kaveh and Dadras [123]
EO	Equilibrium optimizer	Faramarzi et al. [72]
ASO	Atom search optimization	Zhao et al. [287]
NRO	Nuclear reaction optimization	Wei et al. [253]
HGSO	Henry gas solubility optimization	Hashim et al. [93]

Relief filter technique and EO called RBEO-LS, which has two phases: first utilized the ReliefR algorithm as a pre-processing step for feature weights assignment and the second used binary EO (BEO) as a wrapper search technique. The performance was tested on sixteen datasets from the UCI database and other high-dimensional biological datasets. The results displayed that RBEO-LS proved superior to other well-known algorithms. Further work in feature selection for biological data classification developed using the EO can be found in Too and Mirjalili [240].

4.26 Gravitational search algorithm

The GSA was introduced by Rashedi et al. [201]. The basis of this algorithm was the gravitational law and mass interactions. The search agents in this approach were an assembly of masses that cooperate based on the laws of motion and Newtonian gravity. The algorithm was compared with some popular search techniques, and the results showed the high performance of the GSA in solving different nonlinear functions. Papa et al. [179] proposed a combination of algorithms that utilized the optimization behavior of GSA and the speed of the optimum-path forest (OPF) classifier called (OPF-GSA) in providing an accurate and quick framework for the selection of features. The experiments on the datasets gotten from UCI and NTL datasets, like classification of images, recognition of vowels, power distribution systems, and fraud detection, were conducted to evaluate the robustness of the OPF-GSA against linear discriminant analysis (LDA), principal component analysis (PCA), and an algorithm based on PSO for selection of features. Nagpal et al. [170] explored the power of the wrapper-based GSA in solving feature selection issues using biomedical datasets. This approach utilized the GSA and KNN to reduce the number of feature subsets while improving prediction accuracy. Ing et al. [107] presented a new system to regulate the best daily arrangement based on variable photovoltaic (PV) output generation and the profile data load. The results indicated that the proposed best daily configuration technique could advance the performance of the distribution network concerning the reduction of power loss, improvement of voltage profile, and switching minimization.

Furthermore, Zhu et al. [290] proposed an improved GSA known as IGSA, which adopted the concept of global memory and the definition of exponential Kbest to improve the baseline GSA. In this approach, the authors improved the exploitation ability of the IGSA by memorizing the optimal solution obtained, thereby preventing the particles from premature convergence and slow movement, maintaining an equilibrium between the exploration and exploitation. Moreover, Taradeh et al. [236], a GSA-based algorithm with evolutionary mutation and crossover

operators was presented to solve the multiclass feature selection problems. The method used both KNN and decision tree classifiers on eighteen popular UCI datasets to assess the method's performance. It was compared with PSO, GA, and the baseline GSA and outperformed the algorithms mentioned. Kumar and John [135] presented a hybridized Gaussian-based particle swarm optimization gravitational search algorithm for wide-ranging feature selection. This method overcame the shortcoming of getting stuck in local optima and large parameter usage plagued by GSA, PSO, and PSOGSA. SVM was used as a classifier and was assessed for different benchmark datasets.

4.27 Sine cosine algorithm

The SCA is a novel metaheuristic algorithm developed by Sindhu et al. [224] for global optimization using a novel position update approach. The approach's position update procedure for each search agent was generated by two coefficients: exploration and exploitation rates. Those coefficients were updated in every run of the algorithm and presented an appropriate balance between the two phases, and the performance was evaluated using some benchmark functions. Experimental results showed faster convergence speed and achievement of global best with higher accuracy. Hafez et al. [88] produced a feature selection system that utilizes the SCA. Typically, the SCA can search the feature region rapidly to get the best or a near-best subset of features by curtailing a particular fitness function. They evaluated the performance of the approach on eighteen datasets which showed advancement when compared with other procedures such as PSO and GA. Khamees & Rashed [129] presented a novel hybrid method for feature selection. The approach combined SCA and CS to exploit and explore the search space to reach the best solution. The performance was tested on four UCI repository medical datasets and was compared with the baseline SCA and other algorithms. The results indicated the method's efficacy in exploring and exploiting the search space, selecting the fewest feature subset possible, reducing datasets' dimensions, and improving a high classification rate with low run-time on all datasets used. Moreover, Rehman et al. [202] developed a novel Multi Sine Cosine algorithm (MSCA), which employed several swarm clusters to explore & exploit the search space to evade the local minima or maxima issues. The researchers evaluated the hybridization performance on different benchmark functions. Statistically, the results showed the superiority of MSCA in terms of convergence when tested against some commonly used metaheuristic algorithms.

5 Hybrid methods

Hybrid metaheuristic algorithms combine two or more best operators of various metaheuristic algorithms to present a new enhanced or superior version of the existing ones. Many hybrid algorithms have been proposed recently, and hybridization is attracting much attention from the research community. In feature selection, several hybrid metaheuristic algorithms have been proposed to solve this problem in application to biomedical data, image classification, data mining, and many more to obtain optimal feature subsets from irrelevant, redundant original datasets. These hybrid algorithms help eliminate the possibility of trapping in local optima, efficient & effective search space exploration, avoiding premature convergence, and making better exploration. This section explains some of the recently proposed hybrid approaches to selecting the best feature subset in different application areas.

In addition, Yousefpour et al. [274] presented a hybrid method with two metaheuristic algorithms to discover the best feature subset. This study treated this task in two stages: first was to obtain local solutions using filter–wrapper methods to reduce the urge dimensional feature space; second, they used two metaheuristic algorithms to select optimal subset when integrated with harmony search (HS) and GA. The experimental results on three mostly used datasets in sentiment analysis prove the approach's efficacy over other base methods in classification accuracy. Sindhu et al. [225] proposed a hybrid wrapper-based feature selection technique of biogeography-based optimization (BBO) and sine cosine algorithm (SCA) called IBBO for solving feature selection issues. This method introduced the position update mechanism of the SCA algorithm into BBO to improve variety within the habitats. The results showed the outperformance of most of the datasets when the performance was checked against fourteen benchmark datasets from the UCI repository, with seven benchmark test functions used.

Furthermore, Pirgazi et al. [183] presented a hybridized filter–wrapper metaheuristic gene selection approach based on shuffled frog-leaping algorithm (SFLA) and the IWSSr method for high dimension datasets. They implemented two main phases in their system: filter, which used ReliefF for weighing feature, and wrapper, which used the SFLA and IWSSr algorithms to perform effective feature search. The result of the experiment showed a more compact feature set reduction with high classification accuracy. In Mohammadzadeh [164], the study used the natural process of whale optimization & flower pollination algorithms and an opposition-based learning approach to achieve the algorithm's accuracy and convergence speed. The proposed algorithm's performance was evaluated using ten

UCI datasets in spam email detection. Compared with some metaheuristic algorithms, the results were efficacious in the classification accuracy and average feature selection reduction size. Moreover, Dey et al. [52], the study proposed a hybrid feature selection method of metaheuristic using a golden ratio optimization (GRO) and equilibrium optimization (EO) algorithms called the golden ratio-based equilibrium optimization (GREO) algorithm applied in recognizing speech emotion. They fed the selected features by the model into the XGBoost classifier. They considered linear prediction cepstral coefficients (LPCC) and linear predictive coding (LPC)-based features as input and optimized using the GREO algorithm. Two standard datasets were used to assess its high recognition accuracy and outperformed other well-known metaheuristic algorithms for feature selection.

Similarly, Qian et al. [190] proposed an upgraded combined feature selection algorithm comprised of non-linear inertia weight binary particle swarm optimization with shrinking encircling and exploration mechanism (NBPSOSEE) with sequential backward selection (SBS) known as NBPSOSEE-SBS, to select the best feature subset and applied in electric charge recovery risk. The experiment results proved the effectiveness of NBPSOSEE-SBS in reducing the significant number of irrelevant features and improving the prediction results in terms of the lower execution time compared with a well-known algorithm with seven other popular wrapper-based features subset selection techniques used in the prediction of risk of ECR for power customers. Xue et al. [263] proposed a modern hybrid selection algorithm comprising the GA and PSO to enhance the search capabilities of this model, and KNN was utilized as the classifier. The method's performance was used in some simulations using the learning array from UCI as a benchmark dataset. Finally, in this section is the work of [6], the work presented a new hybrid binary variance of improved chaotic crow search and particle swarm optimization algorithm (ECCSPSOA with kNN as the classifier in solving feature selection challenges. The version of CSA was hybridized with PSO for better search strategies and convergence into the optimal global solution within the search space. The method used 15 UCI datasets from popular optimization algorithms with six different performance metrics. The findings demonstrated the efficacy of ECCSPSOA in obtaining a median accuracy rate of 89.67% over the fifteen datasets. The approach outperformed some commonly used methods considering standard deviation and fitness value by getting the least value in thirteen and eight of the datasets considered. However, a major drawback of the ECCSPSOA is its selection of more features on seven of fifteen datasets used. A newly proposed hybridized technique comprised the extended binary cuckoo search, genetic algorithm, and

whale optimization algorithm aimed at reducing the time required to search a huge database during image retrieval. This approach was compared with other popular classification algorithms like KNN, NB, RF CatBoost, and many more, considering recall, precision, error rate, and F-measure [118]. Isaac et al. [109] proposed a hybrid competitive coevolution model for feature selection that utilized two nature-inspired paddy field algorithm and spider monkey optimization. This approach employed SVM as the classifier and was evaluated on two datasets, producing better results than each algorithm individually. They applied the efficacy of this hybridization to diagnose pulmonary emphysema disease. Basu et al. [29] presented a combination of harmony search and adaptive hill climbing approach to feature selection. The researchers used deep learning based on CNNs—convolutional neural networks to extract features. The performance was better than other popular algorithms detecting the Covid-19 virus on CT scan datasets.

6 Issues and challenges

Although metaheuristic algorithms have solved a series of feature selection problems, there are still some noticeable issues and challenges that we will discuss in this section.

6.1 Scalability and stability

In our world today, datasets are growing exponentially, ranging from thousands to millions. The metaheuristic algorithms developed in solving feature selection problems must be scalable to handle the increasing volume of selecting the best feature subsets from the high-dimensional datasets available. The scalability of feature selection algorithms has been regarded as a great problem requiring a sufficient number of samples to get accurate statistics. Bolón-Canedo et al. [31] noted that little attention had been given to the scalability of the feature selection methods as against the training classifiers. The classifier to be used by the algorithm must also be scalable. Otherwise, it would not be able to handle the classification task. Therefore, in designing an algorithm, we conclude that scalability must be incorporated into the design of algorithms from the start. Extra attention needs to be assumed to the scalability of feature selection and classification to keep pace with the increasing growth of data.

The stability of designed algorithms is another vital issue to be addressed. As defined by Aggarwal et al. [7], stability is the sensitivity of the selection process to data agitation in the training set. It was discovered that after the perturbation is introduced to training samples, features with shallow stability can be selected using common feature

selection techniques. In most cases, feature selection algorithms do not find the same subset for various sample datasets when attempting to obtain the best subsets classification. Alelyani et al. [15] discovered that the fundamental traits of data can significantly disturb an algorithm's stability. A stable feature selection algorithm is as vital as classification accuracy. The authors in [31, 128, 246] proposed some solutions to stabilize feature selection algorithms. However, developing feature selection algorithms for high precision and stability classification is still a challenge.

6.2 Multiclass classifiers

The choice of an appropriate classifier for any feature selection algorithm is key to its success in obtaining the best solution. There are different proposed classifiers used with metaheuristic algorithms to solve feature selection problems, such as k-near neighborhood (KNN), support vector machine (SVM), naive Bayesian (NB), artificial neural network (ANN), random forest (RF), kernel extreme learning machine (KELM), fuzzy rule-based (FR), C4.5, ID3, optimum-path forest (OPF). KNN happens to be the most popular classifier used in the literature, as indicated in Fig. 7, and can be applied for large dimensional datasets. SVM is the next most used classifier for feature selection but mainly was applied to medical datasets like cancer detection, artery diseases, and intrusion detection systems. The role of the other classifiers is indicated in Fig. 7 to describe the classifier for multiclass classification problems, Lin [140] presented an efficient classifier to deal with the issue of continuous data explosion and computational complexity, which has deteriorated the performance and accuracy of the classification models by adopting multivariate statistical analyses. The study exploits the two advantages of multivariate statistical analyses: their ability to explore the relationships between variables and locate the most illustrating features of the examined data and their ability to solve problems that are stuck by high dimensionality. The study applied the number one advantage to select relevant feature subsets and the number two to generate the multivariate classifier. The experimental results indicated that their model could significantly improve the classification training time and still maintain accuracy in multi-class classification problems.

Also, their classifier's discrimination degree outperformed other well-known classifiers. Additional studies were done that modified or hybridized different classifiers to practical multiclass feature selection problems. Among these are [103] hybridized feature selection and SVM recursive feature elimination (SVM-RFE) to examine classification accuracy in multi-class problems on Dermatology and Zoo databases. Atallah et al. [26] whose work

designed an intelligent kidney transplant prediction method to solve the prediction problem by modifying the KNN. A system that categorized the mammogram images into malignant, benign, and normal was created by Punithavathi and Devakumari [189]. The images were preprocessed and extracted features from the region to train the modified SVM and KNN classifier. Moreover, Ezenkwu et al. [69] combined the SVM and random forest classifiers, and Sesmero et al. [216] formalized and evaluated an ensemble of classifiers designed to resolve multiclass problems.

In Yijing et al. [272], the study realized that recent work found in the literature rarely focuses on imbalanced learning situations in multiclass learning. However, more attention was given to binary imbalance cases and balance situations where the regular classifiers display their strength. They believed that since imbalanced data differ in their imbalance ratio, class numbers, and dimension, the classifiers' performances in learning from diverse datasets are not the same. This motivated the proposal of a system of manifold classifiers called adaptive multiclass classifier system (AMCS), which can handle multiclass imbalanced learning that can differentiate various types of imbalanced data. They combined three major components in AMCS: feature selection, ensemble learning, and resampling, which were individually selected discriminatively for various imbalanced kinds of data. They applied the proposed AMCS in recognition of oil-bearing reservoirs. The results showed the accuracy of the AMCS in recognizing layers characters of some good logging data. Mustaqem et al. [169] conducted a study to classify cardiac arrhythmia disease into one of sixteen categories using a wrapper-based feature selection method and SVM for multiclass classification. Dataset from the ICU repository was employed to test the performance of their approach, and one-against-one and SVM techniques showed superiority over other classifiers as they got the data accuracy of 81.11% on 80/20 and 92.07% on 90/10 data splits. In Lausser et al. [137], the study incorporates feature selection processes in multiclass classifiers considering low cardinal high-dimensional datasets. Their feature selection method does not precisely fit into wrapper, filter, and embedded categories. They saw their two feature selection network examples as lightly related to the multiclass classifier system's structure applied to diagnostic phenotype prediction. They evaluated the performance ability of this approach using RD, SVM, and KNN on some multiclass microarray datasets, which proved superior to other methods on most of the datasets used.

6.3 Datasets

In proving the performance of a metaheuristic algorithm for feature selection problems, choosing the

suitable datasets is key in selecting the best optima subsets for the classification and training of classifiers. The datasets provide a set of scientifically proven data from diverse application areas for the performance evaluation of any algorithm. We found that most researchers employed the various datasets, for example, Iris, Wine, Breast Cancer, and Heart disease, from the UCI machine learning repository for performance evaluation from the literature reviewed. Only a few studies employed datasets from other warehouses, such as [63, 149], whose work employed microarray datasets [226], utilized the Kyoto 2006+ datasets [229], tested their approach using JAFFE database and the Cohn–Kanade, [179] experimented using datasets from both UCI and NTL datasets, Dey et al. [52] conducted the performance evaluation of their hybrid approach using speech emotion recognition (SER) datasets in Surrey Audio-Visual Expressed Emotion (SAVEE) and Emotional DB (EmoDB). Sometimes, there is a benchmark challenge because the performance evaluation depends on a particular dataset, classes, and classifiers. Comparing feature selection metaheuristic algorithms with a single dataset and the same kind of classifiers is better. However, some classifiers are renowned for multiclass feature selection problems, while others a suitable for binary issues. Table 6 presents a summary of commonly used datasets for feature selection problems, references of research work that used them, and the dataset's attributes or resource location for their attributes.

6.4 Objective function construction

A wrapper-based algorithm optimizes a particular objective function in selecting the best feature subset(s). Due to the high number of binary feature selection-related work in the literature, the objective function formulated includes either classification accuracy maximization or minimization of the number of subsets selected. Moreover, to combine the conflicting objectives, most works in the literature constructed multi-objective functions to solve feature selection issues [9, 62, 65, 146, 223, 277] and converted the multi-objective problem into a single objective problem through the application of weights on both objectives then performed the algorithm learning. This method has effectively and efficiently optimized the fitness function and located the optimal feature subset within particular datasets. This method has efficiently and effectively optimized the fitness function and located the optimal feature subset within specific datasets.

Mathematically, the objective function can be represented as $Z = ax + by$ where a & b are constraints, x & y are variables, and Z is the objective function that can either be minimized or maximized. Any problem that pursues minimizing or maximizing a linear function depending on

Table 6 Summary of commonly used datasets for feature selection problems

Datasets	Dataset characteristics	Attribute	Associated tasks	Instances	Attributes	Missing values	Authors
Iris	Multivariate	Real	Classification	150	4	No	Chaudhary et al. [37], Bhardwaj et al. [30], Khamees and Rashed [129], Mousavirad and Ebrahimpour-Komleh [166], Sadat and Saniee [204], Sánchez-Marroño et al. [211], Shunmugapriya and Kanmani [221], Xue et al. [262]
Breast Cancer Wisconsin	Multivariate	Real	Classification	569	32	No	Allam and Nandhini [17], Jeyasingh and Veluchamy [116], Medjahed et al. [149], Sadat and Saniee [204]
Wine	Classification	Integer, real	Classification	178	13	No	Farhad et al. [71], Hegazy et al. [96], Mafarja et al. [144], Mohmmadzadeh [164], Wu et al. [255], Xu [259], Zhang et al. [286]
Zoo	Multivariate	Categorical, integer	Classification	101	17	No	Chantar et al. [36], Huang et al. [103], Mafarja and Mirjalili [145], Mafarja and Mirjalili [146], Mohmmadzadeh [164], Pourpanah et al. [71], Yijing et al. [272], Zawbaa et al. [277]
Pima Indian Diabetics	Multivariate, time series	Categorical, integer	N/A	N/A	20	N/A	Aljarah et al. [16], Jeyasingh and Veluchamy [116], Mousavirad and Ebrahimpour-Komleh [166], Shunmugapriya and Kanmani [221]

some constraints determined by any set of linear inequalities is said to be an optimization problem. It can be a single objective or multi-objective optimization problem. A generic multi-objective optimization problem [86] can formally be expressed as:

$$\begin{aligned}
 &\text{minimize} \\
 &(x) : (f_1(x), f_2(x), \dots, f_M(x)) \\
 &\text{Subject to : } x_L \leq x \leq x_U \tag{8} \\
 &g_a(x) \leq 0, a = 1, \dots, p \\
 &h_b(x) \leq 0, a = 1, \dots, p
 \end{aligned}$$

Here, $x \in \mathbb{R}^D$, D being the number of problem’s variables, x_L and x_U represent the lower bound and the upper bound of the variables, respectively f_1, f_2, \dots, f_M represent the objective function of M to be optimized, which depends on p & q , inequality & equality constraints, respectively. To extend the population-based optimization problems to multi-objective problems, the authors [82] used three broad categories of methodologies, which are discussed below:

1. *Pareto-based method* These methods utilize Pareto-dominance relations to assess the population’s quality. Methods here are still in use today; however, it seems that their ability to solve problems that are more than three objectives (many-objective) is somehow inadequate [110]. Fonseca & Fleming [76] first used the Pareto dominance relation in a multi-objective genetic algorithm. The individuals in the population were

ranked using Eq. 9 after evaluating the objective function.

$$r_i = 1 + p_i, \tag{9}$$

where p_i represent the number of dominating individuals of the decision vector x_i .

2. *Decomposition-based methods* These methods use vector weighting and a scalarizing function to break down a multi-objective problem into smaller subproblems that are single-objective. When these subproblem sets are solved, it is assumed that a decent estimate of the Pareto front is attained. Available evidence suggests that this way of dealing with multi-objective problems is greatly scalable for multiple objectives problems. However, difficulties still exist in resolving this form of method [82].
3. *Indicator-based methods* This kind of procedure for solving multi-objective problems is considered promising, as in [251], which is dependent on developed metrics in measuring the quality of the set of solutions gotten from a population-based optimization technique. The hypervolume is the most notable of these many indicators presented in the multi-objective optimization setting [292]. Zitzler et al. [291] noted that the need to compare analysis relating to the strength and

weaknesses of various algorithms birthed the introduction of these many indicators.

In Mirjalili et al. [158], the researchers applied the multi-objective optimization problem to solve the design of engineering problems. The equation in their study without losing generality is given as:

$$\begin{aligned} \text{Minimize: } & F(\vec{x}) = \{f_1(\vec{x}), f_1(\vec{x}), \dots, f_o(\vec{x})\} \\ \text{Subject to : } & g_i(\vec{x}) \geq 0, i = 1, 2, \dots, m \\ & h_i(\vec{x}) \leq 0, a = 1, 2, \dots, p \\ & lb_i \leq ub_i, i = 1, 2, \dots, n \end{aligned} \quad (10)$$

where o represent the objectives' number, m and p , inequality and equality constraints, respectively, and lb_i & ub_i are upper & lower bounds of the i th variables.

6.5 Evaluation criteria for performance checking

Assessing a model's performance reveals how well it performs on hidden data. In the practical sense, making predictions data for the future is the essence of the model. Hence, it is a significant problem that predictive models intend to solve. As a result, there is a serious need to comprehend the context before deciding on a suitable metric. Each model tries to address a difficulty with an objective using a separate dataset [5]. Numerous evaluation metrics are used in the literature in assessing the performance of wrapper-based feature selection metaheuristic algorithms. The recall and precision were in data classification in computer science; area under curve (AUC) was employed in radar signals; specificity and sensitivity were used in medical classification. There are other means to measure the performance of algorithms, and a few of the well-known performance metrics are discussed here which were not reviewed in [8]:

1. *Standard deviation (SD)* This measures the similarity between various solutions run. A high SD shows substantial changes in the solution as the function executes iteratively. It can be mathematically formulated as:

$$\sigma^j = \sqrt{\frac{\sum_{i=0}^N (S_i^j - \mu)^2}{N}} \quad (11)$$

2. *Average of solution* This is the ratio of the selected features and the overall number of features in the original dataset. It can be mathematically represented as:

$$\mu^j = \frac{\sum_{i=0}^N S_i^j}{N} \quad (12)$$

7 Application area

Multiclass feature selection has been applied in different areas of human endeavors. This section details the diverse application areas of multiclass feature selection. Moreover, a few studies combined feature selection methods with deep learning and ML approaches and are applied in some real-world scenarios, which are as well mentioned in this section.

The major areas of multiclass feature selection found in the literature include facial expression classification [229], cancer diagnosis or detection [66, 116, 149, 187], network intrusion detection [80, 194, 215, 262], text classification [132, 148, 265], classification and detection of other diseases [37, 113, 196, 206], image retrieval [118, 120]. Other application areas include oil-bearing recognition [169], cardiac arrhythmia disease detection [169], phenotype diagnosis [137], classification of power quality disturbance [150], human activity recognition [98], network traffic classification [30, 70, 219], emotion detection from speech signals [273], and sentiment classification [84].

In Bhardwaj et al. [30] and Shi et al. [219], robust techniques were proposed to solve traffic classification challenges in the imbalance dataset, although both studies employed different approaches. The former is called their method global optimization approach (GOA), which selects the best features and recognizes stable ones. This method combined various popular feature selection methods to get optimal feature subsets using different traffic datasets to solve network traffic problems. They proposed a novel goodness measure inside the random forest, outperforming the well-known feature selection methods in traffic classification. While the latter adopted feature selection and deep learning approaches, which proved to have a better outcome in overcoming the negative effect of multiclass imbalance and concept drift issues associated with the machine learning methods. This approach utilized three main phases: to begin with, useless features were taken out using symmetric uncertainty, after which deep learning was introduced to the already selected features to reduce dimensionality and generate features, and lastly, the removal of redundant features, which was done by weighted symmetric uncertainty (WSU). In [34, 83], an approach was proposed to use radiographic or X-ray images to detect the Covid-19 virus using the combination of feature selection techniques with deep learning models, i.e., CNN and deep neural network, respectively. The latter

study was carried out in Pakistan with an outstanding result, which has been adopted for use in the Pakistan radiology department. Both studies showed more than a 98% accuracy rate, suggesting that combining deep learning and feature selection techniques could produce a high accuracy rate. Figure 8 shows the areas of application of multiclass feature selection in various fields of human endeavor. Recent research works were conducted to improve the model's accuracy.

8 Discussion and future directions

This survey presents a study on various multiclass feature selection techniques that exist in the literature that have been applied to high-dimensional dataset classification. This study revealed the strength and weaknesses of feature selection metaheuristic algorithms discussed in this section, given some gaps for future research in this domain.

Furthermore, we found that several metaheuristic algorithms' variants have not been developed to solve multiclass feature selection problems. These algorithms include AAA, ACS, BCO, BMO, CGS, CHIO, CSO, EMA EPC, EVOA, ES, FBIO, GbSA, HSO, HGSO, LCA, MBA, PFA, SFL, SSA, SSDO, TCO, TGSR, VCS, VPL, WSA, WSO, and many more. These algorithms will advance classification when their binary and multiclass versions are developed. Several existing methods could also be advanced to solve real-world feature selection problems that are often multiclass. The literature shows that

researchers have faced challenges in obtaining the classification problem's best subset.

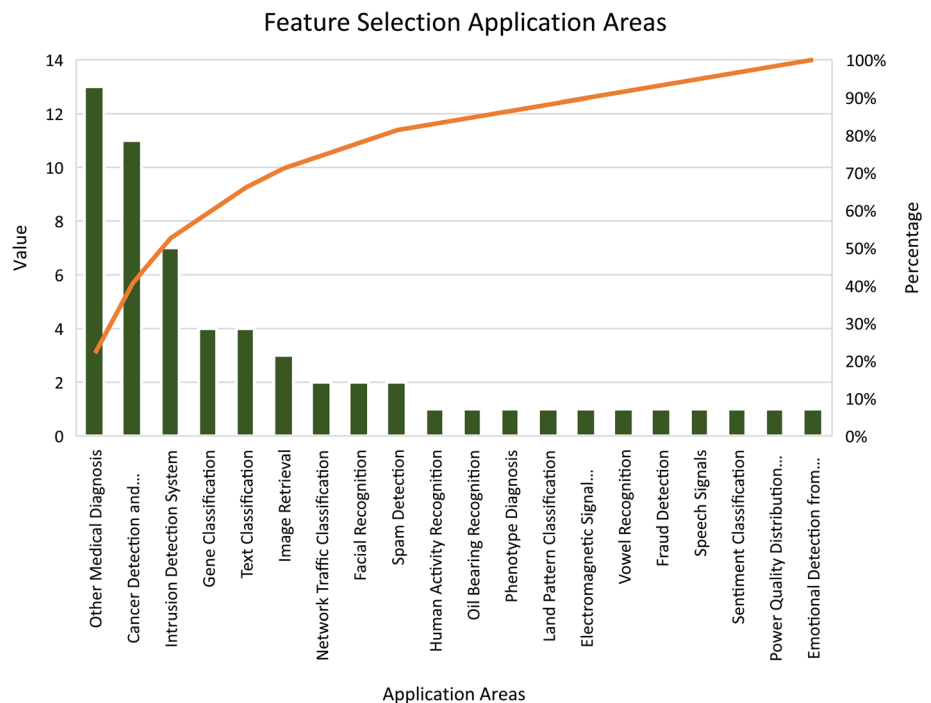
Although, all metaheuristic algorithms perform differently in different datasets and problems. However, some of them have the following strengths:

1. Ability to converge to a true global optimum.
2. Good exploration and exploitation.
3. Ease of implementation.
4. Ability to perform a local and global search.
5. It is suitable for dynamic applications since some can quickly adapt to change.

The limitations of some of them and possible solutions include:

1. Existing methods are unscalable and are usually unstable when dealing with multiple class and higher-dimensional datasets.
2. Fewer or limited multiclass classifiers exist in the literature compared with binary ones. The strategy for tackling multiclass problems could include using a limited multiclass classifier or decomposing them into multiple binary forms and solving them iteratively using a binary classifier, which may be computationally costly.
3. Existing algorithms suffer from slow convergence rates due to random generation of movement.
4. They are prone to get trapped in local optima or premature convergence. Certainly, metaheuristic algorithms' use in feature selection is evolving. Upcoming research work can concentrate on combining

Fig. 8 Multiclass feature selection application areas



metaheuristic algorithms to prevent the shortcomings of single ones. Although developing this framework may be demanding, it will present more effective and satisfactory results.

5. They contain complex operators for selection and crossover.
6. High computational time. Most of the alerts generated by algorithms in intrusion detection systems tend to be fault alarm rates which increases the detection rate due to unrelated & incomplete features and duplication in the intrusion detection systems' datasets. To overcome these challenges and ensure the development of more accurate and efficient IDS models, some researchers have utilized preprocessing methods such as feature selection, normalization, and hybrid modeling techniques that were usually applied. Therefore, we recommend more future work hybridizing IDS models with algorithms to create the feature selection for IDS with higher predictive capacity. Also, IDS should be modeled as multiclass problems where detection is classified into severe, mild, high, medium, or low. Moreover, future work should reduce the false alert rate of the metaheuristic feature selection methods.
7. They must tune many metaheuristic algorithms' hyperparameter values, leading to premature convergence. Future research should be directed toward verifying the control parameters for metaheuristic algorithms. Few works found in the literature explore that particular area(s). The area that can help is hyperparameters for metaheuristic algorithms in different control parameters test values during the evaluation stage of assessing the algorithm's practicality. The accuracy of the network for a precise duty depends on the parameters' structure.
8. Many of the algorithms were designed for binary or real search space only.

Due to the limitations identified, we, therefore, propose that instead of presenting new algorithms belonging to the swarm and physics-based methods, emphasis should be placed on improving and hybridizing the existing metaheuristic algorithms in these areas to minimize the identified disadvantages and are directly applied to solving multiclass feature selection problems. Although, limited algorithms have been proposed in evolutionary and human-based categories. We, therefore, suggest more novel work in evolution and human-based metaheuristic algorithms to solve feature selection difficulties that are particularly multiclass since there is still a significant gap in the development of metaheuristic algorithms specifically for multiclass classification methods. Many techniques in the literature used for multiclass problems combined two or more existing binary methods or hybridization of other

forms; however, little or nothing was done mainly for multiclass problems.

The literature moreover revealed that researchers had faced diverse challenges in obtaining the best-selected feature sets in multiclass classification problems as single classifiers seem not to be as effective as combining more than one, which can take advantage of the strengths of those classifiers into one improved approach. The literature's most utilized classifiers for feature selection problems are KNN and SVM, which implies that future studies can be done considering other classifiers. Some classifiers have attracted less usage in classification issues. Furthermore, we found from the literature that more work was done in applying feature selection problems to areas like intrusion detection systems, cancer detection, image classification, multimedia, text classification [139] and many more, but little or no work was done in drug classification, theft detection, and weather prediction. These can be good areas of application for researchers in this domain to explore.

Finally, few works were found in the literature where deep learning models combined with feature selection techniques proved highly accurate in classification accuracy of not below 98% [34, 83, 219]. Therefore, we conclude that future studies can be undertaken to harness the deep learning approach in combination with feature selection techniques. Also, our study discovered that this hybrid approach had been conducted and applied to disease diagnosis or detection only, which particularly influences the medical science field. In future research, this can also be extended to other real-world situations such as the field of technology (spam mail detection, security threat classification) and the industry—customer purchase behavioral prediction, and more, not just to solve medical problems that currently dominate the research endeavors. The future development of heuristic or metaheuristic approaches may tend toward the use of other classifiers popular, creation of more hybrid local search techniques with metaheuristic methods for more accurate prediction as studied in [20, 36, 60, 136], multi-objective binary techniques creation, application to solving medical diagnosis challenges [180], hybridized wrapper-filter approaches are expected to be developed, and more real-world application areas would be embarked upon [184].

9 Conclusion

The current value of data for knowledge discovery in the digital world has positioned feature selection in an active and evolving mode. The choice of meaningful data helps humans remove noisy or useless data subsets from an original dataset. These assist the different machine learning

models in learning from a meaningful set of data for prediction and solving real-life problems. Moreover, the rise in the quantity of digital data from various sources like web pages, image repositories, databases, social media platforms, and many more indicates the improvement in data classification.

The recent rapid data creation, exchange, and sharing of data make data analysis, extraction, and knowledge retrieval a very daunting task. First, there is a need to reduce the dimensionality of the available irrelevant or redundant data to excerpt knowledge and obtain insight from such data. The feature selection process is a crucial data preprocessing stage that helps minimize the predictive model's data dimension. Though this is a very complicated and demanding task in terms of computation, if not done correctly, it may defeat the purpose of removing relevant feature subsets and the suitability of any predictive model in real life.

Feature selection is essential in enabling any model's faster performance, eradicating loud, less useful data, improving the model's accuracy and precision, removing unwanted features, and increasing the data testing generalization. Although the traditional feature selection methods have been adopted for classification tasks, these approaches have failed to significantly reduce the high dimension of the feature space, producing inaccurate and inefficient prediction models. An evolving method like metaheuristic optimization has provided an emerging standard for feature selection as they yield exciting results that are precise for best classification as again the traditional approaches. Metaheuristic techniques have the inherent ability to effectively improve the accuracy of computational demands, storage, and categorization. Therefore, metaheuristics have been applied more and more in different fields. Nevertheless, a small detail about best practices for case-by-case usage of these evolving feature selection approaches is known. The findings in the literature continue to reveal the most effective ways(s) that, if not accurately performed, may alter the predictive model's real-world application, precision, and performance.

Based on the reviewed literature, feature selection in machine learning has attracted much attention. A systematic review was conducted in this paper, emphasizing wrapper-based metaheuristic algorithms to solve multiclass feature selection problems with metaheuristic algorithms. Other studies were done in this area center on binary issues, text classification, multi-objective feature selection by introducing reliability to cater for missing data, smarm-based approach, evolutionary-based approach, and many more, but we found no work in the literature on multiclass feature selection problems considering the four categories earlier mentioned. Several techniques aimed at improving

the performance of metaheuristics have been considered. The little work done in multiclass feature selection as found in the literature combined two or more binary approaches to help in multiclass situations. However, some metaheuristic algorithms can present a higher performance than others to solve feature selection difficulties in high-dimensional datasets.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Abbass HA (2001) MBO: marriage in honey bees optimization-a Haplometrosis polygynous swarming approach. In: Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat No.01TH8546) (vol 1, pp 207–214). <https://doi.org/10.1109/CEC.2001.934391>
2. Hammouri AI, Mafarja M, Al-Betar MA, Doush IA, Awadallah MA (2020) An improved dragonfly algorithm for feature selection. *Knowl-Based Syst* 203(106131):1–16
3. Abedinia O, Amjady N, Ghasemi A (2016) A new metaheuristic algorithm based on shark smell optimization. *Complexity* 21(5):97–116. <https://doi.org/10.1002/cplx.21634>
4. Abedinpourshotorban H, Shamsuddin SM, Beheshti Z, Jawawi DNA (2016) Electromagnetic field optimization: a physics-inspired metaheuristic optimization algorithm. *Swarm Evol Comput* 26:8–22. <https://doi.org/10.1016/j.swevo.2015.07.002>
5. Abiodun EO, Alabdulatif A, Abiodun OI, Alawida M, Alabdulatif A, Alkhawaldeh RS (2021) A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities. In: *Neural computing and applications*. Deutschland: Springer Science and Business Media. <https://doi.org/10.1007/s00521-021-06406-8>
6. Adamu A, Abdullahi M, Junaidu SB, Hassan IH (2021) An hybrid particle swarm optimization with crow search algorithm for feature selection. *Mach Learn Appl* 6:100108. <https://doi.org/10.1016/j.mlwa.2021.100108>
7. Aggarwal CC, Kong X, Gu Q, Han J, Yu PS (2014) Active learning: a survey. *Data Classif Algorithms Appl*. <https://doi.org/10.1201/b17320>
8. Agrawal P, Abutarboush HF, Ganesh T, Mohamed AW (2021) Metaheuristic algorithms on feature selection: a survey of one decade of research (2009–2019). *IEEE Access* 9:26766–26791. <https://doi.org/10.1109/ACCESS.2021.3056407>
9. Agrawal P, Ganesh T, Mohamed AW (2021) A novel binary gaining–sharing knowledge-based optimization algorithm for feature selection. *Neural Comput Appl* 33(11):5989–6008. <https://doi.org/10.1007/s00521-020-05375-8>
10. Agrawal P, Ganesh T, Mohamed AW (2021) Chaotic gaining sharing knowledge-based optimization algorithm: an improved metaheuristic algorithm for feature selection. *Soft Comput* 25(14):9505–9528. <https://doi.org/10.1007/s00500-021-05874-3>
11. Agrawal P, Ganesh T, Oliva D, Mohamed AW (2021) S-shaped and v-shaped gaining–sharing knowledge-based algorithm for feature selection. *Appl Intell* 52(1):81–112

12. Agushaka JO, Ezugwu AE (2022) Initialisation approaches for population-based metaheuristic algorithms: a comprehensive review. *Appl Sci* (Switz). <https://doi.org/10.3390/app12020896>
13. Al-Betar MA, Alyasseri ZAA, Awadallah MA, Abu Doush I (2021) Coronavirus herd immunity optimizer (CHIO). *Neural Comput Appl* 33(10):5011–5042. <https://doi.org/10.1007/s00521-020-05296-6>
14. Alatas B (2011) ACROA: artificial chemical reaction optimization algorithm for global optimization. *Expert Syst Appl* 38(10):13170–13180. <https://doi.org/10.1016/j.eswa.2011.04.126>
15. Alelyani S, Liu H, and Wang L (2011) The effect of the characteristics of the dataset on the selection stability. In: *Proceedings-international conference on tools with artificial intelligence, ICTAI*, pp 970–977. <https://doi.org/10.1109/ICTAI.2011.167>
16. Aljarah I, Al-Zoubi AM, Faris H, Hassonah MA, Mirjalili S, Saadeh H (2018) Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm. *Cogn Comput* 10(3):478–495. <https://doi.org/10.1007/s12559-017-9542-9>
17. Allam M, Nandhini M (2018) Feature optimization using teaching learning based optimization for breast disease diagnosis. *Int J Recent Technol Eng* 7(4):78–85
18. Almarzouki HZ (2022) Deep-learning-based cancer profiles classification using gene expression data profile. *J Healthc Eng.* <https://doi.org/10.1155/2022/4715998>
19. Ahmadi-Javid A (2011) Anarchic society optimization: a human-inspired method. In: *2011 IEEE congress of evolutionary computation, CEC 2011* (pp 2586–2592). <https://doi.org/10.1109/CEC.2011.5949940>
20. Ahmed S, Sheikh KH, Mirjalili S, Sarkar R (2022) Binary simulated normal distribution optimizer for feature selection: theory and application in COVID-19 datasets. *Expert Syst Appl* 200:116834. <https://doi.org/10.1016/j.eswa.2022.116834>
21. Armanfard N, Reilly JP, Komeili M (2016) Local feature selection for data classification. *IEEE Trans Pattern Anal Mach Intell* 38(6):1217–1227. <https://doi.org/10.1109/TPAMI.2015.2478471>
22. Arora S, Singh S (2019) Butterfly optimization algorithm: a novel approach for global optimization. *Soft Comput* 23(3):715–734. <https://doi.org/10.1007/s00500-018-3102-4>
23. Arslan S, Ozturk C (2019) Multi hive artificial bee colony programming for high dimensional symbolic regression with feature selection. *Appl Soft Comput J* 78:515–527. <https://doi.org/10.1016/j.asoc.2019.03.014>
24. Askarzadeh A (2016) A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. *Comput Struct* 169:1–12. <https://doi.org/10.1016/j.compstruc.2016.03.001>
25. Atashpaz-Gargari E, Lucas C (2007) Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In: *Proceedings 2007 IEEE congress on evolutionary computation, CEC 2007* (pp 4661–4667). <https://doi.org/10.1109/CEC.2007.4425083>
26. Atallah DM, Badawy M, El-Sayed A (2019) Intelligent feature selection with modified K-nearest neighbor for kidney transplantation prediction. *SN Appl Sci* 1(10):1–17. <https://doi.org/10.1007/s42452-019-1329-z>
27. Azar AT, Banu N, Koubaa A (2020) Rough set based ant-lion optimizer for feature selection. In: *Proceedings-2020 6th conference on data science and machine learning applications, CDMA 2020* (pp 81–86). <https://doi.org/10.1109/CDMA47397.2020.00020>
28. Bansal JC, Sharma H, Jadon SS, Clerc M (2014) Spider Monkey Optimization algorithm for numerical optimization. *Memet Comput* 6(1):31–47. <https://doi.org/10.1007/s12293-013-0128-0>
29. Basu A, Sheikh KH, Cuevas E, Sarkar R (2022) COVID-19 detection from CT scans using a two-stage framework. *Expert Syst Appl* 193:116377. <https://doi.org/10.1016/j.eswa.2021.116377>
30. Bhardwaj A, Tiwari A, Bhardwaj H, Bhardwaj A (2016) A genetically optimized neural network model for multi-class classification. *Expert Syst Appl* 60:211–221. <https://doi.org/10.1016/j.eswa.2016.04.036>
31. Bolón-Canedo V, Rego-Fernández D, Peteiro-Barral D, Alonso-Betanzos A, Guijarro-Berdiñas B, Sánchez-Marño N (2018) On the scalability of feature selection methods on high-dimensional data. *Knowl Inf Syst* 56(2):395–442
32. Brezočnik L, Fister I, Podgorelec V (2018) Swarm intelligence algorithms for feature selection: a review. *Appl Sci* (Switz) 8(9):1521. <https://doi.org/10.3390/app8091521>
33. Bui QT, van Pham M, Nguyen QH, Nguyen LX, Pham HM (2019) Whale optimization algorithm and adaptive neuro-fuzzy inference system: a hybrid method for feature selection and land pattern classification. *Int J Remote Sens* 40(13):5078–5093. <https://doi.org/10.1080/01431161.2019.1578000>
34. Canayaz M (2021) MH-COVIDNet: diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images. *Biomed Signal Process Control* 64(2020):102257. <https://doi.org/10.1016/j.bspc.2020.102257>
35. Castillo E, Sánchez-Marño N, Alonso-Betanzos A, Castillo C (2007) Functional network topology learning and sensitivity analysis based on ANOVA decomposition. *Neural Comput* 19(1):231–257. <https://doi.org/10.1162/neco.2007.19.1.231>
36. Chantar H, Tubishat M, Essgaer M, Mirjalili S (2021) Hybrid binary dragonfly algorithm with simulated annealing for feature selection. *SN Comput Sci* 2(4):1–11. <https://doi.org/10.1007/s42979-021-00687-5>
37. Chaudhary A, Kolhe S, Kamal R (2016) A hybrid ensemble for classification in multiclass datasets: an application to oilseed disease dataset. *Comput Electron Agric* 124:65–72. <https://doi.org/10.1016/j.compag.2016.03.026>
38. Chen X, Xu B, Yu K, Du W (2018) Teaching-learning-based optimization with learning enthusiasm mechanism and its application in chemical engineering. *J Appl Math.* <https://doi.org/10.1155/2018/1806947>
39. Chikara RR, Sharma P, Singh L (2018) An improved dynamic discrete firefly algorithm for blind image steganalysis. *Int J Mach Learn Cybern* 9(5):821–835. <https://doi.org/10.1007/s13042-016-0610-3>
40. Chu S, Tsai P, Pan J (2006) *Cat Swarm Optimization*. Springer-Verlag Berlin Heidelberg, 4099 (March 2014), pp 854–858. <https://doi.org/10.1007/11801603>
41. Chu Y, Mi H, Liao H, Ji Z, Wu Q H (2008) A fast bacterial swarming algorithm for high-dimensional function optimization. In: *2008 IEEE congress on evolutionary computation (IEEE world congress on computational intelligence)* (pp 3135–3140). <https://doi.org/10.1109/CEC.2008.4631222>
42. Civicioglu P (2012) Transforming geocentric cartesian coordinates to geodetic coordinates by using differential search algorithm. *Comput Geosci* 46:229–247. <https://doi.org/10.1016/j.cageo.2011.12.011>
43. Civicioglu P (2013) Artificial cooperative search algorithm for numerical optimization problems. *Inf Sci* 229:58–76. <https://doi.org/10.1016/j.ins.2012.11.013>
44. Civicioglu P (2013) Backtracking search optimization algorithm for numerical optimization problems. *Appl Math Comput* 219(15):8121–8144. <https://doi.org/10.1016/j.amc.2013.02.017>
45. Comellas F, Martínez-Navarro J (2009) Bumblebees: a multi-agent combinatorial optimization algorithm inspired by social insect behaviour. In: *2009 world summit on genetic and evolutionary computation, 2009 GEC summit-proceedings of the*

- 1st ACM/SIGEVO summit on genetic and evolutionary computation, GEC'09 (pp 811–814). <https://doi.org/10.1145/1543834.1543949>
46. Cuevas E, Cienfuegos M, Zaldívar D, Pérez-Cisneros M (2013) A swarm optimization algorithm inspired in the behavior of the social-spider. *Expert Syst Appl* 40(16):6374–6384. <https://doi.org/10.1016/j.eswa.2013.05.041>
 47. Cuevas E, Oliva D, Zaldívar D, Pérez-Cisneros M, Sossa H (2012) Circle detection using electro-magnetism optimization. *Inf Sci* 182(1):40–55. <https://doi.org/10.1016/j.ins.2010.12.024>
 48. Cui X, Li Y, Fan J, Wang T, Zheng Y (2020) A hybrid improved dragonfly algorithm for feature selection. *IEEE Access* 8:155619–155629. <https://doi.org/10.1109/ACCESS.2020.3012838>
 49. Dai C, Chen W, Song Y, Zhu Y (2010) Seeker optimization algorithm: a novel stochastic search algorithm for global numerical optimization. *J Syst Eng Electron* 21(2):300–311. <https://doi.org/10.3969/j.issn.1004-4132.2010.02.021>
 50. Dankolo MN, Radzi NHM, Sallehuddin R, Mustaffa NH (2017) A study of metaheuristic algorithms for high dimensional feature selection on microarray data. *AIP Conf Proc*. <https://doi.org/10.1063/1.5012198>
 51. Das H, Chakraborty S, Acharya B, Sahoo AK (2020) Optimal selection of features using teaching-learning-based optimization algorithm for classification. *Appl Intell Decis Mak Mach Learn*. <https://doi.org/10.1201/9781003049548-11>
 52. Dey A, Chattopadhyay S, Singh PK, Ahmadian A, Ferrara M, Sarkar R (2020) A hybrid meta-heuristic feature selection method using golden ratio and equilibrium optimization algorithms for speech emotion recognition. *IEEE Access* 8:200953–200970. <https://doi.org/10.1109/ACCESS.2020.3035531>
 53. Dhiman G, Kumar V (2017) Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications. *Adv Eng Softw* 114:48–70. <https://doi.org/10.1016/j.advengsoft.2017.05.014>
 54. Dhiman G, Kumar V (2019) Seagull optimization algorithm: theory and its applications for large-scale industrial engineering problems. *Knowl-Based Syst* 165:169–196. <https://doi.org/10.1016/j.knsys.2018.11.024>
 55. Dhivyaprabha TT, Subashini P, Krishnaveni M (2018) Synergistic fibroblast optimization: a novel nature-inspired computing algorithm. *Front Inf Technol Electron Eng* 19(7):815–833. <https://doi.org/10.1631/FITEE.1601553>
 56. Dietterich TG, Bakiri G (1995) Solving multiclass learning problems via error-correcting output codes. *J Artif Intell Res* 2:263–286. <https://doi.org/10.1613/jair.105>
 57. Du H, Wu X, Zhuang J (2006) Small-world optimization algorithm for function optimization BT. In: Jiao L, Wang L, Gao, Liu J, and Wu F (Eds) *Advances in Natural Computation* (pp 264–273). Berlin :Springer
 58. Dy JG, Brodley CE (2004) Feature selection for unsupervised learning. *J Mach Learn Res* 5:845–889
 59. Eberhart R, Kennedy J (1995) Particle swarm optimization. *Proc IEEE Int Conf Neural Netw* 4:1942–1948
 60. Elgamel ZM, Yasin NBM, Tubishat M, Alswaiti M, Mirjalili S (2020) An improved harris hawks optimization algorithm with simulated annealing for feature selection in the medical field. *IEEE Access* 8:186638–186652. <https://doi.org/10.1109/ACCESS.2020.3029728>
 61. Elgamel ZM, Yasin NM, Sabri AQM, Sihwail R, Tubishat M, Jarrah H (2021) Improved equilibrium optimization algorithm using elite opposition-based learning and new local search strategy for feature selection in medical datasets. *Computation* 9(6):68. <https://doi.org/10.3390/computation9060068>
 62. El-Kenawy ESM, Eid MM, Saber M, Ibrahim A (2020) MbGWO-SFS: modified binary grey wolf optimizer based on stochastic fractal search for feature selection. *IEEE Access* 8:107635–107649. <https://doi.org/10.1109/ACCESS.2020.3001151>
 63. Elyasigomari V, Lee DA, Screen HRC, Shaheed MH (2017) Development of a two-stage gene selection method that incorporates a novel hybrid approach using the cuckoo optimization algorithm and harmony search for cancer classification. *J Biomed Inform* 67:11–20. <https://doi.org/10.1016/j.jbi.2017.01.016>
 64. Emary E, Yamany W, Hassanien AE, Snel V (2015) Multi-objective gray-wolf optimization for attribute reduction. *Procedia Comput Sci* 65:623–632. <https://doi.org/10.1016/j.procs.2015.09.006>
 65. Emary E, Zawbaa HM (2019) Feature selection via Lévy Antlion optimization. *Pattern Anal Appl* 22(3):857–876. <https://doi.org/10.1007/s10044-018-0695-2>
 66. Engchuan W, Chan JH (2015) Pathway activity transformation for multi-class classification of lung cancer datasets. *Neurocomputing* 165:81–89. <https://doi.org/10.1016/j.neucom.2014.08.096>
 67. Erol OK, Eksin I (2006) A new optimization method: big bang-big crunch. *Adv Eng Softw* 37(2):106–111. <https://doi.org/10.1016/j.advengsoft.2005.04.005>
 68. Eskandar H, Sadollah A, Bahreininejad A, Hamdi M (2012) Water cycle algorithm-a novel metaheuristic optimization method for solving constrained engineering optimization problems. *Comput Struct* 110–111:151–166. <https://doi.org/10.1016/j.compstruc.2012.07.010>
 69. Ezenkwu CP, Akpan UI, Stephen BU-A (2021) A class-specific metaheuristic technique for explainable relevant feature selection. *Mach Learn Appl* 6:100142. <https://doi.org/10.1016/j.mlwa.2021.100142>
 70. Fahad A, Tari Z, Khalil I, Almalawi A, Zomaya AY (2014) An optimal and stable feature selection approach for traffic classification based on multi-criterion fusion. *Futur Gener Comput Syst* 36:156–169. <https://doi.org/10.1016/j.future.2013.09.015>
 71. Pourpanah F, Shi Y, Lim CP, Hao Qi, Tan CJ (2019) Feature selection based on brain storm optimization for data classification. *Appl Soft Comput J* 80:761–775
 72. Faramarzi A, Heidarinejad M, Mirjalili S, Gandomi AH (2020) Marine predators algorithm: a nature-inspired metaheuristic. *Expert Syst Appl* 152:113377. <https://doi.org/10.1016/j.eswa.2020.113377>
 73. Faramarzi A, Heidarinejad M, Stephens B, Mirjalili S (2020) Equilibrium optimizer: a novel optimization algorithm. *Knowl-Based Syst* 191:105190. <https://doi.org/10.1016/j.knsys>
 74. Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst* 13(2):87–129. <https://doi.org/10.1590/s0104-07072011000400008>
 75. Filho CJAB, Neto FB, de L, Lins AJCC, Nascimento AIS, Lima MP (2008) A novel search algorithm based on fish school behavior. In: 2008 IEEE international conference on systems, man and cybernetics (pp 2646–2651). <https://doi.org/10.1109/ICSMC.2008.4811695>
 76. Fonseca C, Fleming P (1993) Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. In: *Conference on genetic algorithms* (vol 423, pp 416–423)
 77. Fodor IK, Fodor IK (2002) A survey of dimension reduction techniques (No. UCRL-ID-148494). Lawrence Livermore National Lab, CA (US)
 78. Fu Z, An J, Yang Q, Yuan H, Sun Y, Ebrahimi H (2022) Skin cancer detection using Kernel Fuzzy C-means and developed

- Red Fox Optimization algorithm. *Biomed Signal Process Control* 71:103160. <https://doi.org/10.1016/j.bspc.2021.103160>
79. Gandomi AH, Alawi AH (2012) Krill herd: a new bio-inspired optimization algorithm. *Commun Nonlinear Sci Numer Simul* 17(12):4831–4845. <https://doi.org/10.1016/j.cnsns.2012.05.010>
 80. Ghanem WAHM, Jantan A (2020) A new approach for intrusion detection system based on training multilayer perceptron by using enhanced Bat algorithm. *Neural Comput Appl* 32(15):11665–11698. <https://doi.org/10.1007/s00521-019-04655-2>
 81. Ghorbani N, Babaei E (2014) Exchange market algorithm. *Appl Soft Comput J* 19:177–187. <https://doi.org/10.1016/j.asoc.2014.02.006>
 82. Giagkiozis I, Purshouse RC, Fleming PJ (2015) An overview of population-based algorithms for multi-objective optimisation. *Int J Syst Sci* 46(9):1572–1599. <https://doi.org/10.1080/00207721.2013.823526>
 83. Gilanie G, Bajwa UI, Waraich MM, Asghar M, Kousar R, Kashif A, Aslam RS, Qasim MM, Rafique H (2021) Coronavirus (COVID-19) detection from chest radiology images using convolutional neural networks. *Biomed Signal Process Control* 66:102490. <https://doi.org/10.1016/j.bspc.2021.102490>
 84. Gokalp O, Tasci E, Ugur A (2020) A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. *Expert Syst Appl* 146:113176. <https://doi.org/10.1016/j.eswa.2020.113176>
 85. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
 86. Habib A, Singh HK, Chugh T, Ray T, Miettinen K (2019) A multiple surrogate assisted decomposition-based evolutionary algorithm for expensive multi/many-objective optimization. *IEEE Trans Evol Comput* 23(6):1000–1014. <https://doi.org/10.1109/TEVC.2019.2899030>
 87. Hafez AI, Hassanien AE, Zawbaa HM, Emary E (2015) Hybrid Monkey algorithm with krill herd algorithm optimization for feature selection. In: 2015 11th international computer engineering conference (ICENCO) (pp 273–277). <https://doi.org/10.1109/ICENCO.2015.7416361>
 88. Hafez AI, Zawbaa HM, Emary E, Hassanien AE (2016) Sine cosine optimization algorithm for feature selection. In: 2016 international symposium on innovations in intelligent systems and applications (INISTA), pp 1–5. IEEE
 89. Hammouri AI, Mafarja M, Al-Betar MA, Awadallah MA, Abu-Doush I (2020) An improved dragonfly algorithm for feature selection. *Knowl-Based Syst* 203:106131. <https://doi.org/10.1016/j.knsys.2020.106131>
 90. Hancer E, Xue B, Zhang M, Karaboga D, Akay B (2018) Pareto front feature selection based on artificial bee colony optimization. *Inf Sci* 422:462–479. <https://doi.org/10.1016/j.ins.2017.09.028>
 91. Haq SSU, Aftab Ahmad FI, Nawaz T, Majeed A, Shahzar MS, Sattar MK (2020) A novel binary variant model of swarm inspired polar bear optimization algorithm employed for scalable unit commitment. *Int Trans Electric Energy Syst*. <https://doi.org/10.1002/2050-7038.12711>
 92. Harifi S, Khalilian M, Mohammadzadeh J, Ebrahimnejad S (2019) Emperor Penguins Colony: a new metaheuristic algorithm for optimization. *Evol Intel* 12(2):211–226. <https://doi.org/10.1007/s12065-019-00212-x>
 93. Hashim FA, Houssein EH, Mabrouk MS, Al-Atabany W, Mirjalili S (2019) Henry gas solubility optimization: a novel physics-based algorithm. *Futur Gener Comput Syst* 101:646–667. <https://doi.org/10.1016/j.future.2019.07.015>
 94. Hatamlou A (2013) Black hole: a new heuristic optimization approach for data clustering. *Inf Sci* 222:175–184. <https://doi.org/10.1016/j.ins.2012.08.023>
 95. He S, Wu QH, Saunders JR (2009) Group search optimizer: an optimization algorithm inspired by animal searching behavior. *IEEE Trans Evol Comput* 13(5):973–990. <https://doi.org/10.1109/TEVC.2009.2011992>
 96. Hegazy AE, Makhlof MA, El-Tawel GS (2020) Improved salp swarm algorithm for feature selection. *J King Saud Univ-Comput Inf Sci* 32(3):335–344. <https://doi.org/10.1016/j.jksuci.2018.06.003>
 97. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: algorithm and applications. *Futur Gener Comput Syst* 97:849–872. <https://doi.org/10.1016/j.future.2019.02.028>
 98. Helmi AM, Al-Qaness MAA, Dahou A, Damaševičius R, Krilavičius T, Elaziz MA (2021) A novel hybrid gradient-based optimizer and grey wolf optimizer feature selection method for human activity recognition using smartphone sensors. *Entropy* 23(8):1065. <https://doi.org/10.3390/e23081065>
 99. Holmes G, Pfahringer B, Kirkby R, Frank E, Hall M (2002) Multiclass alternating decision trees BT. In: Elomaa LT, Mannila H, Toivonen H (eds) *Machine learning: ECML 2002*. Springer, Berlin, pp 161–172
 100. Hosseini HS (2011) Principal components analysis by the galaxy-based search algorithm: a novel metaheuristic for continuous optimisation. *Int J Comput Sci Eng* 6(1/2):132. <https://doi.org/10.1504/ijcse.2011.041221>
 101. Hoque N, Bhattacharyya DK, Kalita JK (2014) MIFS-ND: a mutual information-based feature selection method. *Expert Syst Appl* 41(14):6371–6385. <https://doi.org/10.1016/j.eswa.2014.04.019>
 102. Hu J, Han Z, Heidari AA, Shou Y, Ye H, Wang L, Huang X, Chen H, Chen Y, Wu P (2021) Detection of COVID-19 severity using blood gas analysis parameters and Harris hawks optimized extreme learning machine. *Comput Biol Med* 142:105166. <https://doi.org/10.1016/j.combiomed.2021.105166>
 103. Huang ML, Hung YH, Lee WM, Li RK, Jiang BR (2014) SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *Sci World J*. <https://doi.org/10.1155/2014/795624>
 104. Hussain K, Neggaz N, Zhu W, Houssein EH (2021) An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection. *Expert Syst Appl* 176:114778. <https://doi.org/10.1016/j.eswa.2021.114778>
 105. Husseinzadeh Kashan A (2014) League championship algorithm (LCA): an algorithm for global optimization inspired by sport championships. *Appl Soft Comput J* 16:171–200. <https://doi.org/10.1016/j.asoc.2013.12.005>
 106. Ibrahim RA, Ewees AA, Oliva D, Abd Elaziz M, Lu S (2019) Improved salp swarm algorithm based on particle swarm optimization for feature selection. *J Ambient Intell Humaniz Comput* 10(8):3155–3169. <https://doi.org/10.1007/s12652-018-1031-9>
 107. Ing KG, Mokhlis H, Illias HA, Aman MM, Jamian JJ (2015) Gravitational search algorithm and selection approach for optimal distribution network configuration based on daily photovoltaic and loading variation. *J Appl Math*. <https://doi.org/10.1155/2015/894758>
 108. Iruthayarajan W, Baskar S (2010) Covariance matrix adaptation evolution strategy based design of centralized PID controller. *Expert Syst Appl* 37(8):5775–5781. <https://doi.org/10.1016/j.eswa.2010.02.031>
 109. Isaac A, Nehemiah HK, Dunston SD, Elgin Christo VR, Kannan A (2022) Feature selection using competitive coevolution of bio-inspired algorithms for the diagnosis of pulmonary emphysema. *Biomed Signal Process Control* 72:103340. <https://doi.org/10.1016/j.bspc.2021.103340>

110. Ishibuchi H, Tsukamoto N, and Nojima Y (2008) Evolutionary many-objective optimization: a short review. In: IEEE congress on evolutionary computation (pp 2419–2426)
111. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* 22(1):4–37. <https://doi.org/10.1109/34.824819>
112. Jain S, Dharavath R (2021) Memetic salp swarm optimization algorithm based feature selection approach for crop disease detection system. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-021-03406-3>
113. Jayaraman V, Sultana HP (2019) Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-019-01193-6>
114. Jahani E, Chizari M (2018) Tackling global optimization problems with a novel algorithm—Mouth Brooding Fish algorithm. *Appl Soft Comput* 62:987–1002. <https://doi.org/10.1016/j.asoc.2017.09.035>
115. Jain M, Singh V, Rani A (2019) A novel nature-inspired algorithm for optimization: Squirrel search algorithm. *Swarm Evolut Comput* 44:148–175. <https://doi.org/10.1016/j.swevo.2018.02.013>
116. Jeyasingh S, Veluchamy M (2017) Modified bat algorithm for feature selection with the Wisconsin Diagnosis Breast Cancer (WDBC) dataset. *Asian Pac J Cancer Prev* 18(5):1257–1264. <https://doi.org/10.22034/APJCP.2017.18.5.1257>
117. Jimoh RG, Abisoye OA, Uthman MMB (2018) Classification and feature selection of symptomatic and climatic based malaria parasite counts using support vector machine
118. Johari PK, Gupta RK (2021) Improved feature selection techniques for image retrieval based on metaheuristic optimization. *Int J Comput Sci Netw Secur* 21(1):40–48
119. Jović A, Brkić K, Bogunović N (2015) A review of feature selection methods with applications. In: 2015 38th international convention on information and communication technology, electronics and microelectronics, MIPRO 2015-proceedings (pp 1200–1205). <https://doi.org/10.1109/MIPRO.2015.7160458>
120. Kanimozhi T, Latha K (2015) An integrated approach to region based image retrieval using firefly algorithm and support vector machine. *Neurocomputing* 151(P3):1099–1111. <https://doi.org/10.1016/j.neucom.2014.07.078>
121. Karaboga D, Basturk B (2007) Artificial Bee Colony (ABC) optimization algorithm for solving constrained optimization problems. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 4529 LNAI (pp 789–798). https://doi.org/10.1007/978-3-540-72950-1_77
122. Kashan AH (2009) League championship algorithm: a new algorithm for numerical function optimization. *SoCPaR 2009-Soft Comput Pattern Recognit*. <https://doi.org/10.1109/SoCPaR.2009.21>
123. Kaveh A, Dadras A (2017) A novel meta-heuristic optimization algorithm: thermal exchange optimization. *Adv Eng Softw* 110:69–84. <https://doi.org/10.1016/j.advengsoft.2017.03.014>
124. Kaveh A, Farhoudi N (2013) A new optimization method: dolphin echolocation. *Adv Eng Softw* 59:53–70. <https://doi.org/10.1016/j.advengsoft.2013.03.004>
125. Kaveh A, Khayatazad M (2012) A new meta-heuristic method: Ray optimization. *Comput Struct* 112–113:283–294. <https://doi.org/10.1016/j.compstruc.2012.09.003>
126. Kaveh A, Mahdavi VR (2014) Colliding bodies optimization: a novel meta-heuristic method. *Comput Struct* 139:18–27. <https://doi.org/10.1016/j.compstruc.2014.04.005>
127. Kaveh A, Talatahari S (2010) A novel heuristic optimization method: Charged system search. *Acta Mech* 213(3–4):267–289. <https://doi.org/10.1007/s00707-009-0270-4>
128. Khaire UM, Dhanalakshmi R (2019) Stability of feature selection algorithm: a review. *J King Saud Univ-Comput Inf Sci*. <https://doi.org/10.1016/j.jksuci.2019.06.012>
129. Khamees M, Rashed AAB (2020) Hybrid SCA-CS optimization algorithm for feature selection in classification problems. *AIP Conf Proc* 2290:040001. <https://doi.org/10.1063/5.0028662>
130. Khishe M, Mosavi MR (2020) Chimp optimization algorithm. *Expert Syst Appl* 149:113338. <https://doi.org/10.1016/j.eswa.2020.113338>
131. Khorami E, Mahdi Babaei F, Azadeh A (2021) Optimal diagnosis of COVID-19 based on convolutional neural network and red fox optimization algorithm. *Comput Intell Neurosci*. <https://doi.org/10.1155/2021/4454507>
132. Khurana A, Prakash Verma O (2020) A fine tuned model of grasshopper optimization algorithm with classifiers for optimal text classification. In: 2020 IEEE 17th India council international conference, INDICON 2020 (pp 1–7), IEEE. <https://doi.org/10.1109/INDICON49873.2020.9342432>
133. Kitonyi PM, Segera DR (2021) Hybrid gradient descent grey wolf optimizer for optimal feature selection. *Biomed Res Int*. <https://doi.org/10.1155/2021/2555622>
134. Kratzke TM, Stone LD, Frost JR (2010) Search and rescue optimal planning system. In: 13th conference on information fusion, Fusion 2010 (pp 1–8). <https://doi.org/10.1109/icif.2010.5712114>
135. Kumar S, John B (2021) A novel gaussian based particle swarm optimization gravitational search algorithm for feature selection and classification. *Neural Comput Appl* 33(19):12301–12315. <https://doi.org/10.1007/s00521-021-05830-0>
136. Kurtuluş E, Yıldız AR, Sait SM, Bureerat S (2020) A novel hybrid Harris hawks-simulated annealing algorithm and RBF-based metamodel for design optimization of highway guardrails. *Mater Test* 62(3):251–260. <https://doi.org/10.3139/120.111478>
137. Lausser L, Szekely R, Schirra LR, Kestler HA (2018) The influence of multi-class feature selection on the prediction of diagnostic phenotypes. *Neural Process Lett* 48(2):863–880. <https://doi.org/10.1007/s11063-017-9706-3>
138. Li Q, Chen H, Huang H, Zhao X, Cai ZN, Tong C, Liu W, Tian X (2017) An enhanced grey wolf optimization based feature selection wrapped Kernel extreme learning machine for medical diagnosis. *Comput Math Methods Med*. <https://doi.org/10.1155/2017/9512741>
139. Li W, Miao D, Wang W (2011) Two-level hierarchical combination method for text classification. *Expert Syst Appl* 38(3):2030–2039. <https://doi.org/10.1016/j.eswa.2010.07.139>
140. Lin HY (2012) Efficient classifiers for multi-class classification problems. *Decis Support Syst* 53(3):473–481. <https://doi.org/10.1016/j.dss.2012.02.014>
141. Lordache S (2010) Consultant-guided search—a new meta-heuristic for combinatorial optimization problems. In: Proceedings of the 12th annual genetic and evolutionary computation conference, GECCO '10, pp 225–232. <https://doi.org/10.1145/1830483.1830526>
142. Lozano M, García-Martínez C (2010) Hybrid metaheuristics with evolutionary algorithms specializing in intensification and diversification: overview and progress report. *Comput Oper Res* 37(3):481–497. <https://doi.org/10.1016/j.cor.2009.02.010>
143. Łukasik S, Żak S (2009) Firefly algorithm for continuous constrained optimization tasks BT. In: NT Nguyen, R Kowalczyk, S-M Chen (Eds), Computational collective intelligence Semantic web social networks and multiagent systems, pp 97–106. Berlin: Springer

144. Mafarja M, Eleyan D, Abdullah S, Mirjalili S (2017). S-shaped vs. V-shaped transfer functions for ant lion optimization algorithm in feature selection problem. In: ACM international conference proceeding series, Part F130522. <https://doi.org/10.1145/3102304.3102325>
145. Mafarja MM, Mirjalili S (2017) Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. *Neurocomputing* 260:302–312. <https://doi.org/10.1016/j.neucom.2017.04.053>
146. Mafarja M, Mirjalili S (2018) Whale optimization approaches for wrapper feature selection. *Appl Soft Comput* 62:441–453. <https://doi.org/10.1016/j.asoc.2017.11.006>
147. Mandal M, Singh PK, Ijaz MF, Shafi J, Sarkar R (2021) A Tristage wrapper-filter feature selection framework for disease classification. *Sensors*. <https://doi.org/10.3390/s21165571>
148. Marie-Sainte S, Alalyani N (2020) Firefly algorithm based feature selection for arabic text classification. *J King Saud Univ-Comput Inf Sci* 32(3):320–328. <https://doi.org/10.1016/j.jksuci.2018.06.004>
149. Medjahed SA, Saadi TA, Benyettou A, Ouali M (2017) Kernel-based learning and feature selection analysis for cancer diagnosis. *Appl Soft Comput J* 51:39–48. <https://doi.org/10.1016/j.asoc.2016.12.010>
150. Mehedi IM, Ahmadipour M, Salam Z, Ridha HM, Bassi H, Rawa MJH, Ajour M, Abusorrah A, Abdullah MP (2021) Optimal feature selection using modified cuckoo search for classification of power quality disturbances. *Appl Soft Comput* 113:107897. <https://doi.org/10.1016/j.asoc.2021.107897>
151. Meng XB, Gao XZ, Lu L, Liu Y, Zhang H (2016) A new bio-inspired optimisation algorithm: bird swarm algorithm. *J Exp Theor Artif Intell* 28(4):673–687. <https://doi.org/10.1080/0952813X.2015.1042530>
152. Mirjalili S (2015) Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm. *Knowl-Based Syst* 89:228–249. <https://doi.org/10.1016/j.knosys.2015.07.006>
153. Mirjalili S (2015) The ant lion optimizer. *Adv Eng Softw* 83:80–98. <https://doi.org/10.1016/j.advengsoft.2015.01.010>
154. Mirjalili SM, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
155. Mirjalili S, Mirjalili SM, Hatamlou A (2016) Multi-verse optimizer: a nature-inspired algorithm for global optimization. *Neural Comput Appl* 27(2):495–513. <https://doi.org/10.1007/s00521-015-1870-7>
156. Mirjalili S (2016) Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Comput Appl* 27(4):1053–1073. <https://doi.org/10.1007/s00521-015-1920-1>
157. Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
158. Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017) Salp swarm algorithm: a bio-inspired optimizer for engineering design problems. *Adv Eng Softw* 114:163–191. <https://doi.org/10.1016/j.advengsoft.2017.07.002>
159. Mirkhan A, Çelebi N (2021) Finding the optimal features reduct, a hybrid model of rough set and polar bear optimization. In: Kahraman C, Onar SC, Oztaysi B, Sari IU, Cebi S, Tolga AC (eds) *Intelligent and fuzzy techniques: smart and innovative solutions*. Springer International Publishing, Cham, pp 1596–1603
160. Mitra P, Murthy CA, Pal SK (2002) Unsupervised feature selection using feature similarity. *IEEE Trans Pattern Anal Mach Intell* 24(3):301–312. <https://doi.org/10.1109/34.990133>
161. Mladenović N, Hansen P (1997) Variable neighborhood search. *Comput Oper Res* 24(11):1097–1100
162. Mohamed AW, Abutarboush HF, Hadi AA, Mohamed AK (2021) Gaining-sharing knowledge based algorithm with adaptive parameters for engineering optimization. *IEEE Access* 9:65934–65946. <https://doi.org/10.1109/ACCESS.2021.3076091>
163. Mohamed AW, Hadi AA, Mohamed AK (2020) Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm. *Int J Mach Learn Cybern* 11(7):1501–1529. <https://doi.org/10.1007/s13042-019-01053-x>
164. Mohammadzadeh H (2020) Case study email spam detection of two metaheuristic algorithm for optimal feature selection. May, pp 1–25. <https://doi.org/10.20944/preprints202001.0309.v2>
165. Moscato P, Cotta C, Mendes A (2004) Memetic algorithms (pp 53–85). https://doi.org/10.1007/978-3-540-39930-8_3
166. Mousavirad S, Ebrahimpour-Komleh H (2014) Wrapper feature selection using discrete cuckoo optimization algorithm. *Int J Mechatron Electric Comput Technol* 4(11):709–721
167. Muhammad A, Abdullah S, Sani NS (2021) Optimization of sentiment analysis using teaching-learning based algorithm. *Comput Mater Contin* 69(2):1783–1799. <https://doi.org/10.32604/cmc.2021.018593>
168. Muller SD, Marchetto J, Airaghi S, Kournoutsakos P (2002) Optimization based on bacterial chemotaxis. *IEEE Trans Evol Comput* 6(1):16–29. <https://doi.org/10.1109/4235.985689>
169. Mustaqeem A, Anwar SM, Majid M (2018) Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants. *Comput Math Methods Med*. <https://doi.org/10.1155/2018/7310496>
170. Nagpal S, Arora S, Dey S, Shreya S (2017) Feature selection using gravitational search algorithm for biomedical data. *Procedia Comput Sci* 115:258–265. <https://doi.org/10.1016/j.procs.2017.09.133>
171. Nakamura RYM, Pereira LAM, Costa KA, Rodrigues D, Papa JP, Yang XS (2012) BBA: a binary bat algorithm for feature selection. *Braz Symp Comput Gr Image Process*. <https://doi.org/10.1109/SIBGRAPI.2012.47>
172. Neggaz N, Ewees AA, Elaziz MA, Mafarja M (2020) Boosting salp swarm algorithm by sine cosine algorithm and disrupt operator for feature selection. *Expert Syst Appl*. <https://doi.org/10.1016/j.eswa.2019.113103>
173. Nematzadeh H, Enayatifar R, Mahmud M, Akbari E (2019) Frequency based feature selection method using whale algorithm. *Genomics* 111(6):1946–1955. <https://doi.org/10.1016/j.ygeno.2019.01.006>
174. Nematollahi AF, Rahiminejad A, Vahidi B (2017) A novel physical based meta-heuristic optimization method known as lightning attachment procedure optimization. *Appl Soft Comput J* 59:596–621. <https://doi.org/10.1016/j.asoc.2017.06.033>
175. Neshat M, Sepidnam G, Sargolzaei M (2013) Swallow swarm optimization algorithm: a new method to optimization. *Neural Comput Appl* 23(2):429–454. <https://doi.org/10.1007/s00521-012-0939-9>
176. Neshat M, Sepidnam G, Sargolzaei M, Toosi AN (2014) Artificial fish swarm algorithm: a survey of the state-of-the-art, hybridization, combinatorial and indicative applications. *Artif Intell Rev* 42(4):965–997. <https://doi.org/10.1007/s10462-012-9342-2>
177. Ouadfel S, Abd Elaziz M (2022) Efficient high-dimension feature selection based on enhanced equilibrium optimizer. *Expert Syst Appl* 187:115882. <https://doi.org/10.1016/j.eswa.2021.115882>
178. Pan WT (2012) A new fruit fly optimization algorithm: taking the financial distress model as an example. *Knowl-Based Syst* 26:69–74. <https://doi.org/10.1016/j.knosys.2011.07.001>
179. Papa JP, Pagnin A, Schellini SA, Spadotto A, Guido RC, Ponti M, Chiacchia G, Falcão AX (2011) Feature selection Through

- gravitational search algorithm Department of Computing University of Sao Paulo University of Campinas Institute of Computing. *Sort* 2(1):2052–2055
180. Pashaei E, Pashaei E (2022) An efficient binary chimp optimization algorithm for feature selection in biomedical data classification. *Neural Comput Appl* 34(8):6427–6451. <https://doi.org/10.1007/s00521-021-06775-0>
 181. Passino KM (2002) Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Syst Mag* 22(3):52–67. <https://doi.org/10.1109/MCS.2002.1004010>
 182. Pinto P, Runkler TA, Sousa JM (2005) Wasp swarm optimization of logistic systems BT. In: Ribeiro B, Albrecht RF, Dobnikar A, Pearson DW, Steele NC (eds) *Adaptive and natural computing algorithms*. Springer, Vienna, pp 264–267
 183. Pirgazi J, Alimoradi M, Abharian TE, Olyaei MH (2019) An Efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets. *Sci Rep* 9(1):18580. <https://doi.org/10.1038/s41598-019-54987-1>
 184. Piri J, Mohapatra P, Pradhan MR, Acharya B, Patra TK (2022) A binary multi-objective chimp optimizer with dual archive for feature selection in the healthcare domain. *IEEE Access* 10:1756–1774. <https://doi.org/10.1109/ACCESS.2021.3138403>
 185. Polap D, Woźniak M (2017) Polar bear optimization algorithm: Meta-heuristic with fast population movement and dynamic birth and death mechanism. *Symmetry*. <https://doi.org/10.3390/sym9100203>
 186. Połap D, Woźniak M (2021) Red fox optimization algorithm. *Expert Syst Appl* 166:114107. <https://doi.org/10.1016/j.eswa.2020.114107>
 187. Prabukumar M, Agilandeeswari L, Ganesan K (2019) An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier. *J Ambient Intell Humaniz Comput* 10(1):267–293. <https://doi.org/10.1007/s12652-017-0655-5>
 188. Premkumar K, Vishnupriya M, Babu TS, Manikandan BV, Thamizhselvan T, Ali AN, Islam MR, Kouzani AZ, Parvez Mahmud MA (2020) Black widow optimization-based optimal PI-controlled wind turbine emulator. *Sustainability (Switzerland)* 12(24):1–19. <https://doi.org/10.3390/su122410357>
 189. Punithavathi V, Devakumari D (2020) A hybrid algorithm with modified SVM and KNN for classification of mammogram images using medical image processing with data mining techniques. *Eur J Mol Clin Med* 7(10):2956–2965
 190. Qian S, Shi Y, Wu H, Shang S (2020) An improved hybrid feature selection algorithm for electric charge recovery risk. *Math Probl Eng*. <https://doi.org/10.1155/2020/8479341>
 191. Qin LD, Jiang QY, Zou ZY, Cao YJ (2004) A queen-bee evolution based on genetic algorithm for economic power dispatch. In: 39th international universities power engineering conference, 2004. UPEC 2004, vol 1, pp 453–456
 192. Qu C, Zhang L, Li J, Deng F, Tang Y, Zeng X, Peng X (2021) Improving feature selection performance for classification of gene expression data using Harris Hawks optimizer with variable neighborhood learning. *Brief Bioinform* 22(5):bbab097. <https://doi.org/10.1093/bib/bbab097>
 193. Rad HS, Lucas C (2007) A recommender system based on invasive weed optimization algorithm. *IEEE Congr Evolut Comput* 2007:4297–4304. <https://doi.org/10.1109/CEC.2007.4425032>
 194. Rais HM, Mehmood T (2018) Dynamic ant colony system with three level update feature selection for intrusion detection. *Int J Netw Secur* 20(1):184–192. [https://doi.org/10.6633/IJNS.201801.20\(1\).20](https://doi.org/10.6633/IJNS.201801.20(1).20)
 195. Rajamohana SP, Umamaheswari K, Abirami B (2017) Adaptive binary flower pollination algorithm for feature selection in review spam detection. In: 2017 international conference on innovations in green energy and healthcare technologies (IGEHT), 1–4. IEEE
 196. Rajinikanth N, Pavithra L (2021) Heart diseases prediction for optimization based feature selection and classification using machine learning methods. *IJACSA Int J Adv Comput Sci Appl* 12(2):636–643
 197. Ramezani F, Lotfi S (2013) Social-based algorithm (SBA). *Appl Soft Comput J* 13(5):2837–2856. <https://doi.org/10.1016/j.asoc.2012.05.018>
 198. Ranjini SR, Murugan S (2017) Memory based hybrid dragonfly algorithm for numerical optimization problems. *Expert Syst Appl* 83:63–78. <https://doi.org/10.1016/j.eswa.2017.04.033>
 199. Rani RR, Ramyachitra D (2017) Krill herd optimization algorithm for cancer feature selection and random forest technique for classification. In: 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS) (pp 109–113), IEEE
 200. Rao RV, Savsani VJ, Vakharia DP (2011) Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems. *CAD Comput Aided Des* 43(3):303–315. <https://doi.org/10.1016/j.cad.2010.12.0155>
 201. Rashedi E, Nezamabadi-pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. *Inf Sci* 179(13):2232–2248. <https://doi.org/10.1016/j.ins.2009.03.004>
 202. Rehman MZ, Khan A, Ghazali R, Aamir M, Nawi NM (2021) A new Multi Sine-Cosine algorithm for unconstrained optimization problems. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0255269>
 203. Rostami M, Berahmand K, Nasiri E, Forouzandeh S (2021) Review of swarm intelligence-based feature selection methods. *Eng Appl Artif Intell* 100:104210
 204. Sadat N, Saniee M (2022) Cell separation algorithm with enhanced search behaviour in miRNA feature selection for cancer diagnosis. *Inf Syst* 104:101906. <https://doi.org/10.1016/j.is.2021.101906>
 205. Sadollah A, Bahreininejad A, Eskandar H, Hamdi M (2013) Mine blast algorithm: a new population based algorithm for solving constrained engineering optimization problems. *Appl Soft Comput J* 13(5):2592–2612. <https://doi.org/10.1016/j.asoc.2012.11.026>
 206. Saeed NA, Al-Tai ZTM (2019) Feature selection using hybrid dragonfly algorithm in a heart disease predication system. *Int J Eng Adv Technol* 8(6):2862–2867. <https://doi.org/10.35940/ijeat.F8786.088619>
 207. Sahebi G, Movahedi P, Ebrahimi M, Pahikkala T, Plosila J, Tenhunen H (2020) GeFeS: a generalized wrapper feature selection approach for optimizing classification performance. *Comput Biol Med* 125:103974. <https://doi.org/10.1016/j.compbiomed.2020.103974>
 208. Salcedo-Sanz S, Del Ser J, Landa-Torres I, Gil-López S, Portilla-Figueras JA (2014) The coral reefs optimization algorithm: a novel metaheuristic for efficiently solving optimization problems. *Sci World J*. <https://doi.org/10.1155/2014/739768>
 209. Saleem N, Zafar K, Sabzwari AF (2019) Enhanced feature subset selection using niche-based bat algorithm. *Computation*. <https://doi.org/10.3390/COMPUTATION7030049>
 210. Salimi H (2015) Stochastic Fractal Search: a powerful metaheuristic algorithm. *Knowl-Based Syst* 75:1–18. <https://doi.org/10.1016/j.knsys.2014.07.025>
 211. Sánchez-Marroño N, Alonso-Betanzos A, and Calvo-Estévez RM (2009) A wrapper method for feature selection in multiple classes datasets. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 5517 LNCS(PART 1), pp 456–463. https://doi.org/10.1007/978-3-642-02478-8_57

212. Saremi S, Mirjalili S, Lewis A (2017) Grasshopper optimisation algorithm: theory and application. *Adv Eng Softw* 105:30–47. <https://doi.org/10.1016/j.advengsoft.2017.01.004>
213. Saw T, Myint H (2019) Swarm intelligence based feature selection for high dimensional classification: a literature survey. *Int J Comput (IJC)* 33(1):69–83
214. Sayed GI, Khoriba G, Haggag MH (2021) A novel chaotic equilibrium optimizer algorithm with S-shaped and V-shaped transfer functions for feature selection. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-021-03151-7>
215. Selvakumar B, Muneeswaran K (2019) Firefly algorithm based feature selection for network intrusion detection. *Comput Secur* 81:148–155. <https://doi.org/10.1016/j.cose.2018.11.005>
216. Sesmero MP, Alonso-Weber JM, Gutierrez G, Ledezma A, Sanchis A (2015) An ensemble approach of dual base learners for multi-class classification problems. *Inf Fusion* 24:122–136. <https://doi.org/10.1016/j.inffus.2014.09.002>
217. Sevin E, Dökeroglu T (2019) A novel hybrid teaching-learning-based optimization algorithm for the classification of data by using extreme learning machines. *Turk J Electr Eng Comput Sci* 27(2):1523–1533
218. Shadravan S, Naji HR, Bardsiri VK (2019) The Sailfish Optimizer: a novel nature-inspired metaheuristic algorithm for solving constrained engineering optimization problems. *Eng Appl Artif Intell* 80:20–34. <https://doi.org/10.1016/j.engappai.2019.01.001>
219. Shi H, Li H, Zhang D, Cheng C, Cao X (2018) An efficient feature generation approach based on deep learning and feature selection techniques for traffic classification. *Comput Netw* 132:81–98. <https://doi.org/10.1016/j.comnet.2018.01.007>
220. Shi Y (2011) Brainstorm optimization algorithm. In: *International conference in swarm intelligence* (pp 303–309). Berlin: Springer
221. Shunmugapriya P, Kanmani S (2017) A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid). *Swarm Evol Comput* 36:27–36. <https://doi.org/10.1016/j.swevo.2017.04.002>
222. Simon D (2008) Biogeography-based optimization. *IEEE Trans Evol Comput* 12(6):702–713. <https://doi.org/10.1109/TEVC.2008.919004>
223. Simumba N, Okami S, Kodaka A, Kohtake N (2021) Comparison of profit-based multi-objective approaches for feature selection in credit scoring. *Algorithms* 14(9):60. <https://doi.org/10.3390/a14090260>
224. Sindhu R, Ngadiran R, Yacob YM, Zahri NAH, Hariharan M (2017) Sine-cosine algorithm for feature selection with elitism strategy and new updating mechanism. *Neural Comput Appl* 28(10):2947–2958. <https://doi.org/10.1007/s00521-017-2837-7>
225. Sindhu R, Ngadiran R, Yacob YM, Hanin Zahri NA, Hariharan M, Polat K (2019) A hybrid SCA inspired BBO for feature selection problems. *Math Probl Eng*. <https://doi.org/10.1155/2019/9517568>
226. Singh AP, Kaur A (2019) Flower Pollination Algorithm for feature analysis of Kyoto 2006+ data set. *J Inf Optim Sci* 40(2):467–478. <https://doi.org/10.1080/02522667.2019.1580886>
227. Song L, Smola A, Gretton A, Borgwardt KM, Bedo J (2007). Supervised feature selection via dependence estimation. In: *Proceedings of the 24th international conference on machine learning* (pp 823–830)
228. Song Z, Yan X, Zhao L, Fan L, Tang C, Ji J (2021) Adaptive self-scaling brain-storm optimization via a chaotic search mechanism. *Algorithms*. <https://doi.org/10.3390/a14080239>
229. Sreedharan NPN, Ganesan B, Raveendran R, Sarala P, Dennis B, Rajakumar Boothalingam R (2018) Grey wolf optimisation-based feature selection and classification for facial emotion recognition. *IET Biom* 7(5):490–499. <https://doi.org/10.1049/iet-bmt.2017.0160>
230. Sur C, Sharma S, Shukla A (2013) Egyptian vulture optimization algorithm—a new nature inspired meta-heuristics for Knapsack problem BT. In: Meesad P, Unger H, Boonkrong S (eds) *The 9th international conference on computing and information technology (IC2IT2013)*. Springer, Berlin, pp 227–237
231. Taha AM, Mustapha A, der Chen S (2013) Naive bayes-guided bat algorithm for feature selection. *Sci World J*. <https://doi.org/10.1155/2013/325973>
232. Tan Y, Zhu Y (2010) Fireworks algorithm for optimization. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 6145 LNCS(PART 1), pp 355–364. https://doi.org/10.1007/978-3-642-13495-1_44
233. Tamura K, Yasuda K (2011) Spiral optimization. *IEEE Int Conf Syst Man Cybern* 1:1759–1764
234. Tang R, Fong S, Yang X-S, Deb S (2012) Wolf search algorithm with ephemeral memory. In: *Seventh international conference on digital information management (ICDIM 2012)*, pp 165–172. <https://doi.org/10.1109/ICDIM.2012.6360147>
235. Tao Z, Huiling L, Wenwen W, Xia Y (2019) GA-SVM based feature selection and parameter optimization in hospitalization expense modeling. *Appl Soft Comput J* 75:323–332. <https://doi.org/10.1016/j.asoc.2018.11.001>
236. Taradeh M, Mafarja M, Heidari AA, Faris H, Aljarah I, Mirjalili S, Fujita H (2019) An evolutionary gravitational search-based feature selection. *Inf Sci* 497:219–239. <https://doi.org/10.1016/j.ins.2019.05.038>
237. Tawhid MA, Dsouza KB (2018) Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems. *Appl Comput Inf* 16(1–2):117–136. <https://doi.org/10.1016/j.aci.2018.04.001>
238. Tiwari V (2012) Face recognition based on cuckoo search algorithm. *Ind J Comput Sci Eng* 3(3):401–405
239. Too J, Abdullah AR, Saad NM, Ali NM, Tee W (2018) A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification. *Computers*. <https://doi.org/10.3390/computers7040058>
240. Too J, Mirjalili S (2021) General learning equilibrium optimizer: a new feature selection method for biological data classification. *Appl Artif Intell* 35(3):247–263. <https://doi.org/10.1080/08839514.2020.1861407>
241. Too J, Mafarja M, Mirjalili S (2022) Spatial bound whale optimization algorithm: an efficient high-dimensional feature selection approach. *Neural Comput Appl* 33(23):16229–16250. <https://doi.org/10.1007/s00521-021-06224-y>
242. Tubishat M, Idris N, Shuib L, Abushariah MAM, Mirjalili S (2020) Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection. *Expert Syst Appl*. <https://doi.org/10.1016/j.eswa.2019.113122>
243. Uymaz SA, Tezel G, Yel E (2015) Artificial algae algorithm (AAA) for nonlinear global optimization. *Appl Soft Comput J* 31:153–171. <https://doi.org/10.1016/j.asoc.2015.03.003>
244. Vaiyapuri T, Alaskar H, Aljohani E, Shridevi S (2022) Red Fox Optimizer with data-science-enabled microarray. *Appl Sci* 12:4172
245. Venkata Rao R (2016) Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *Int J Ind Eng Comput* 7(1):19–34. <https://doi.org/10.5267/j.ijiec.2015.8.004>
246. Wald R, Khoshgoftaar TM, Napolitano A (2013) Stability of filter- and wrapper-based feature subset selection. *Proc Int Conf Tools Artif Intell ICTAI*. <https://doi.org/10.1109/ICTAI.2013.63>

247. Wang GG, Gandomi AH, Alavi AH (2013) A chaotic particle-swarm krill herd algorithm for global numerical optimization. *Kybernetes* 42(6):962–978. <https://doi.org/10.1108/K-11-2012-0108>
248. Wang G, Guo L, Gandomi AH, Cao L, Alavi AH, Duan H, Li J (2013) Lévy-flight krill herd algorithm. *Math Probl Eng*. <https://doi.org/10.1155/2013/682073>
249. Wang GG, Deb S, Coelho LDS (2016) Elephant herding optimization. In: *Proceedings-2015 3rd international symposium on computational and business intelligence, ISCBi 2015*, pp 1–5. <https://doi.org/10.1109/ISCBi.2015.8>
250. Wang M, Wu C, Wang L, Xiang D, Huang X (2019) A feature selection approach for hyperspectral image based on modified ant lion optimizer. *Knowl-Based Syst* 168:39–48. <https://doi.org/10.1016/j.knsys.2018.12.031>
251. Wang R, Purshouse R, Fleming P (2012) Preference inspired co-evolutionary algorithms for many-objective optimisation. *IEEE Trans Evolut Comput* 17:474
252. Wang XH, Zhang Y, Sun XY, Wang YL, Du CH (2020) Multi-objective feature selection based on artificial bee colony: an acceleration approach with variable sample size. *Appl Soft Comput J*. <https://doi.org/10.1016/j.asoc.2019.106041>
253. Wei Z, Huang C, Wang X, Han T, Li Y (2019) Nuclear reaction optimization: a novel and powerful physics-based algorithm for global optimization. *IEEE Access* 7:1–9. <https://doi.org/10.1109/ACCESS.2019.2918406>
254. Wei Y, Lv H, Chen M, Wang M, Heidari AA, Chen H, Li C (2020) Predicting entrepreneurial intention of students: an extreme learning machine with Gaussian Barebone Harris Hawks Optimizer. *IEEE Access* 8:76841–76855. <https://doi.org/10.1109/ACCESS.2020.2982796>
255. Wu S, Hu Y, Wang W, Feng X, Shu W (2013) Application of global optimization methods for feature selection and machine learning. *Math Probl Eng*. <https://doi.org/10.1155/2013/241517>
256. Wu D, Zhang W, Jia H, Leng X (2021) Simultaneous feature selection and support vector machine optimization using an enhanced chimp optimization algorithm. *Algorithms*. <https://doi.org/10.3390/a14100282>
257. Xu J, Zhang J (2014) Exploration-exploitation tradeoffs in metaheuristics: survey and analysis. In: *Proceedings of the 33rd Chinese control conference, CCC 2014*, pp 8633–8638. <https://doi.org/10.1109/ChiCC.2014.6896450>
258. Xu Y, Cui Z, Zeng J (2010) Social emotional optimization algorithm for nonlinear constrained optimization problems BT. In: Panigrahi BK, Das S, Suganthan PN, Dash SS (eds) *Swarm, evolutionary, and memetic computing*. Springer, Berlin, pp 583–590
259. Xu Y (2016) K-nearest neighbor-based weighted multi-class twin support vector machine. *Neurocomputing* 205:430–438. <https://doi.org/10.1016/j.neucom.2016.04.024>
260. Xu Z, King I, Lyu MRT, Jin R (2010) Discriminative semi-supervised feature selection via manifold regularization. *IEEE Trans Neural Netw* 21(7):1033–1047. <https://doi.org/10.1109/TNN.2010.2047114>
261. Xue B, Zhang M, Browne WN, Yao X (2016) A survey on evolutionary computation approaches to feature selection. *IEEE Trans Evol Comput* 20(4):606–626. <https://doi.org/10.1109/TEVC.2015.2504420>
262. Xue Y, Jia W, Zhao X, Pang W (2018) An evolutionary computation based feature selection method for intrusion detection. *Secur Commun Netw*. <https://doi.org/10.1155/2018/2492956>
263. Xue Y, Aouari A, Mansour RF, Su S (2021) A hybrid algorithm based on PSO and GA for feature selection. *J Cyber Secur* 3(2):117–124. <https://doi.org/10.32604/jcs.2021.017018>
264. Xue Y, Zhao B, Ma T, Pang W (2018) A self-adaptive fireworks algorithm for classification problems. *IEEE Access* 6:44406–44416. <https://doi.org/10.1109/ACCESS.2018.2858441>
265. Chen X-W, Zeng X, van Alphen D (2006) Multi-class feature selection for texture classification. *Pattern Recogn Lett* 27(14):1685–1691
266. Xue Y, Binping Z, Tinghuai M (2016) Classification based on fireworks algorithm. In: *International conference on bio-inspired computing-theories and applications*, pp 39–40. Springer, Singapore
267. Yang X-S, Deb S (2009) Cuckoo search via Lévy flights. *IEEE World Congr Nature Biol Inspir Comput NaBIC 2009*:210–214
268. Yang XS (2009) Firefly algorithms for multimodal optimization. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 5792 LNCS, pp 169–178. https://doi.org/10.1007/978-3-642-04944-6_14
269. Yang XS, Gandomi AH (2012) Bat algorithm: a novel approach for global engineering optimization. *Eng Comput (Swansea, Wales)* 29(5):464–483. <https://doi.org/10.1108/02644401211235834>
270. Yang XS, Karamanoglu M, He X (2014) Flower pollination algorithm: a novel approach for multiobjective optimization. *Eng Optim* 46(9):1222–1237. <https://doi.org/10.1080/0305215X.2013.832237>
271. Yapici H, Cetinkaya N (2019) A new meta-heuristic optimizer: pathfinder algorithm. *Appl Soft Comput J* 78:545–568. <https://doi.org/10.1016/j.asoc.2019.03.012>
272. Yijing L, Haixiang G, Xiao L, Yanan L, Jinling L (2016) Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data. *Knowl-Based Syst* 94:88–104. <https://doi.org/10.1016/j.knsys.2015.11.013>
273. Yogesh CK, Hariharan M, Ngadiran R, Adom AH, Yaacob S, Berkai C, Polat K (2017) A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal. *Expert Syst Appl* 69:149–158. <https://doi.org/10.1016/j.eswa.2016.10.035>
274. Yousefpour A, Ibrahim R, Hamed HNA, Yokoi T (2016) Integrated feature selection methods using metaheuristic algorithms for sentiment analysis. *Intelligent information and database systems*. Springer, Berlin, pp 129–140
275. Yun-Tao J, Wan-Qiu Z, Chun-Lin H (2021) A clustering-guided integer brain storm optimizer for feature selection in high-dimensional data. *Discret Dyn Nat Soc*. <https://doi.org/10.1155/2021/8462493>
276. Zakeri A, Hokmabadi A (2019) Efficient feature selection method using real-valued grasshopper optimization algorithm. *Expert Syst Appl* 119:61–72. <https://doi.org/10.1016/j.eswa.2018.10.021>
277. Zawbaa HM, Emary E, Grosan C (2016) Feature selection via chaotic antlion optimization. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0150652>
278. Zawbaa HM, Hassanien AE, Emary E, Yamany W, Parv B (2015) Hybrid flower pollination algorithm with rough sets for feature selection. In: *2015 11th international computer engineering conference (ICENCO)*, pp 278–283. IEEE
279. Žerovnik J (2015) Heuristics for NP-hard optimization problems—simpler is better!? *Logist Sustain Transp* 6(1):1–10. <https://doi.org/10.1515/jlst-2015-0006>
280. Zhang X, Liu S (2008) Interval algorithm for global numerical optimization. *Eng Optim* 40(9):849–868. <https://doi.org/10.1080/03052150802056188>
281. Zhang L, Mistry K, Lim CP, Neoh SC (2018) Feature selection using firefly optimization for classification and regression models. *Decis Support Syst* 106:64–85. <https://doi.org/10.1016/j.dss.2017.12.001>

282. Zhang Y, Jin Z, Mirjalili S (2020) Generalized normal distribution optimization and its applications in parameter extraction of photovoltaic models. *Energy Convers Manag* 224:113301. <https://doi.org/10.1016/j.enconman.2020.113301>
283. Zhang G, Hou J, Wang J, Yan C, Luo J (2020) Feature selection for microarray data classification using hybrid information gain and a modified binary krill herd algorithm. *Interdiscip Sci Comput Life Sci* 12(3):288–301. <https://doi.org/10.1007/s12539-020-00372-w>
284. Zhang Y, Cheng S, Shi Y, Wei GD, Zhao X (2019) Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm. *Expert Syst Appl* 137:46–58. <https://doi.org/10.1016/j.eswa.2019.06.044>
285. Zhao Z, Liu H (2007) Semi-supervised feature selection via spectral analysis. In: *Proceedings of the 2007 SIAM international conference on data mining* (pp 641–646). Society for industrial and applied mathematics.
286. Zhang Y, Song XF, Gong DW (2017) A return-cost-based binary firefly algorithm for feature selection. *Inf Sci* 418–419:561–574. <https://doi.org/10.1016/j.ins.2017.08.047>
287. Zhao W, Wang L, Zhang Z (2019) Atom search optimization and its application to solve a hydrogeologic parameter estimation problem. *Knowl-Based Syst* 163:283–304. <https://doi.org/10.1016/j.knosys.2018.08.030>
288. Zhao W, Zhang Z, Wang L (2020) Manta ray foraging optimization: an effective bio-inspired optimizer for engineering applications. *Eng Appl Artif Intell* 87:103300. <https://doi.org/10.1016/j.engappai.2019.103300>
289. Zheng Y, Li Y, Wang G, Chen Y, Xu Q, Fan J, Cui X (2019) A novel hybrid algorithm for feature selection based on whale optimization algorithm. *IEEE Access* 7:14908–14923. <https://doi.org/10.1109/ACCESS.2018.2879848>
290. Zhu L, He S, Wang L, Zeng W, Yang J (2019) Feature selection using an improved gravitational search algorithm. *IEEE Access* 7:114440–114448. <https://doi.org/10.1109/ACCESS.2019.2935833>
291. Zitzler E, Thiele L, Laumanns M, Fonseca C, da Fonseca V (2003) Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Trans Evol Comput* 7(2):117–132
292. Zitzler E, Thiele L (1999) Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Trans Evol Comput* 3(4):257–271

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

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