IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 9, 2021, accepted January 28, 2021, date of publication February 2, 2021, date of current version February 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056407

Metaheuristic Algorithms on Feature Selection: A Survey of One Decade of Research (2009-2019)

PRACHI AGRAWAL[1], HATTAN F. ABUTARBOUSH[2], TALARI GANESH[1], AND ALI WAGDY MOHAMED[3,4]

[1]Department of Mathematics and Scientific Computing, National Institute of Technology at Hamirpur, Hamirpur 177005, India
[2]College of Engineering, Electrical Engineering Department, Taibah University, Medina 42353, Saudi Arabia
[3]Operations Research Department, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza 12613, Egypt
[4]Wireless Intelligent Networks Center (WINC), School of Engineering and Applied Sciences, Nile University, Giza 16453, Egypt

Corresponding author: Prachi Agrawal (prachiagrawal202@gmail.com)

The work of Prachi Agrawal was supported by the Ministry of Human Resource Development (MHRD Govt. of India).

ABSTRACT Feature selection is a critical and prominent task in machine learning. To reduce the dimension of the feature set while maintaining the accuracy of the performance is the main aim of the feature selection problem. Various methods have been developed to classify the datasets. However, metaheuristic algorithms have achieved great attention in solving numerous optimization problem. Therefore, this paper presents an extensive literature review on solving feature selection problem using metaheuristic algorithms which are developed in the ten years (2009-2019). Further, metaheuristic algorithms have been classified into four categories based on their behaviour. Moreover, a categorical list of more than a hundred metaheuristic algorithms is presented. To solve the feature selection problem, only binary variants of metaheuristic algorithms have been reviewed and corresponding to their categories, a detailed description of them explained. The metaheuristic algorithms in solving feature selection problem are given with their binary classification, name of the classifier used, datasets and the evaluation metrics. After reviewing the papers, challenges and issues are also identified in obtaining the best feature subset using different metaheuristic algorithms. Finally, some research gaps are also highlighted for the researchers who want to pursue their research in developing or modifying metaheuristic algorithms for classification. For an application, a case study is presented in which datasets are adopted from the UCI repository and numerous metaheuristic algorithms are employed to obtain the optimal feature subset.

INDEX TERMS Binary variants, classification, feature selection, literature review, metaheuristic algorithms.

I. INTRODUCTION

The real-world problems mostly include a large number of data, and handling the data becomes a very complex and prominent task. A dataset contains a large number of attributes/features. Not always, all the features are necessary to get useful information from the datasets. Some of the features may be irrelevant, redundant, which degrades the performance of the model. Therefore, to reduce the size of the original datasets while maintaining the accuracy of performance is the main aim of feature reduction problem. In feature reduction, feature construction and feature selection take part. The feature extraction or construction constructs a new set of features from original datasets while feature selection selects the relevant features from the original dataset. This paper focuses only on feature selection problems.

A feature selection problem is one of the most challenging tasks in machine learning. If a set contains n number of features, total 2^n subsets are possible from which the best subset has to be picked. It will be most difficult when n tends to a large number because it can not be possible to evaluate the performance of the model at each subset. Hence, to handle the situation, the various methodology has been proposed. Exhaustive search, greedy search, random search etc. are such techniques which have been applied to feature selection problems to find the best subset. Most of the methods suffer from premature convergence, enormous complexity, high computational cost. Therefore, metaheuristic algorithms get much attention to deal with this type of conditions. They are the most efficient and effective techniques and are capable

The associate editor coordinating the review of this manuscript and approving it for publication was Firooz B. Saghezchi.

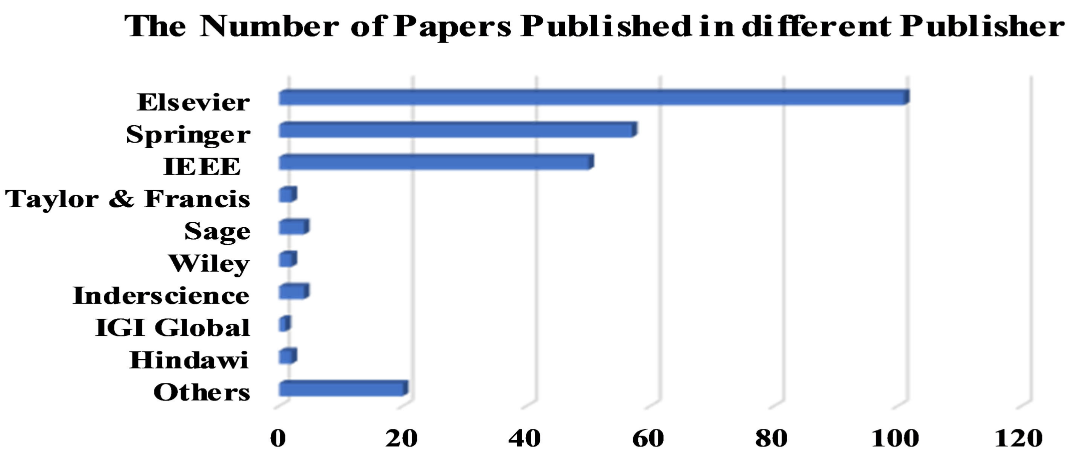

FIGURE 1. The number of papers published regarding the developing of metaheuristic algorithms and feature selection problem.

of finding the best subset of features while maintaining the accuracy of the model.

In the last three decades, several metaheuristic algorithms have been designed to solve various kinds of optimization problems. This study provides an extensive literature survey on metaheuristic algorithms which are developed in the last ten years (2009-2019) and applied to various applications of feature selection problems. Due to its various applications in different fields such as text mining, image processing, bioinformatics, industrial applications, computer vision etc. feature selection becomes an exciting research problem. There are lots of papers published in different publishers regarding the developing metaheuristic algorithm and feature selection problems that is shown in Figure 1. From the figure, it can be observed that Elsevier Publisher publishes more papers in top tier journals (Experts Systems with Applications (IF = 5.452), Applied Soft Computing (IF = 5.472), Knowledge-Based Systems (IF = 5.921), Information Sciences (IF = 5.910), Neurocomputing (IF = 3.317)). In Springer publication, the papers are published in top tier journals Neural Computing and Applications (IF = 4.774), Applied Intelligence (IF = 3.325), Soft Computing (IF = 3.050) etc. The other publishers consist of IOS Press, Massachusetts Institute of Technology (MIT) press, Citeseer, ACM digital library, Hindawi publishing house, Multidisciplinary Digital Publishing Institute (MDPI).

Earlier, a literature survey has been found on feature selection in which non-evolutionary algorithms have been considered. Xue *et al.* [1] presented a survey on evolutionary approaches which mainly focus on genetic algorithm, particle swarm optimization, ant colony optimization, genetic programming in detail. Lee *et al.* [2] studied feature selection in multimedia applications in which they provided an extensive literature survey of seventy related papers from 2001 to 2017. Remeseuro and Bolon-Canedo [3] reviewed feature selection methods on medical problems. Sharma and Kaur [4] presented a systematic review on nature-inspired algorithms to feature selection problem, especially in medical datasets. They provided a categorization of binary and chaotic algorithms for different nature-inspired algorithms. Literature review has been presented in different fields such as sentiment analysis [5], bioinformatics [6], ensemble learning [7] using metaheuristic algorithms.

The main contribution of presenting this study is given as:

(a) This paper presents the definitions and techniques of feature selection problem, and basic concepts of metaheuristic algorithms are thoroughly explained.
(b) The metaheuristic algorithms are classified, and a list of metaheuristic algorithms is given.
(c) It presents an extensive literature of binary metaheuristic algorithms for feature selection problem.
(d) The literature is represented with the vital factor of wrapper feature selection techniques such as the description of the classifier, name of the used datasets, evaluation metrics etc.
(e) It explains the issues and challenges to develop an algorithm in solving feature selection problems. It also presents the evaluation metrics formula to investigate the performance.
(f) Finally, the research gaps and the future work are also presented to enhance the research work.
(g) Eight benchmark datasets have been considered from the UCI repository to show the application of feature selection using wrapper based techniques.
(h) Five metaheuristic algorithms are taken from the literature to implement on feature selection problem.
(i) The results are compared with evaluation metrices i.e. average fitness value, average classification accuracy, average number of feature selected and average computational time.

The organization of the paper as follows: Section II presents the preliminaries for the feature selection and metaheuristic algorithms. The extensive literature on feature selection using

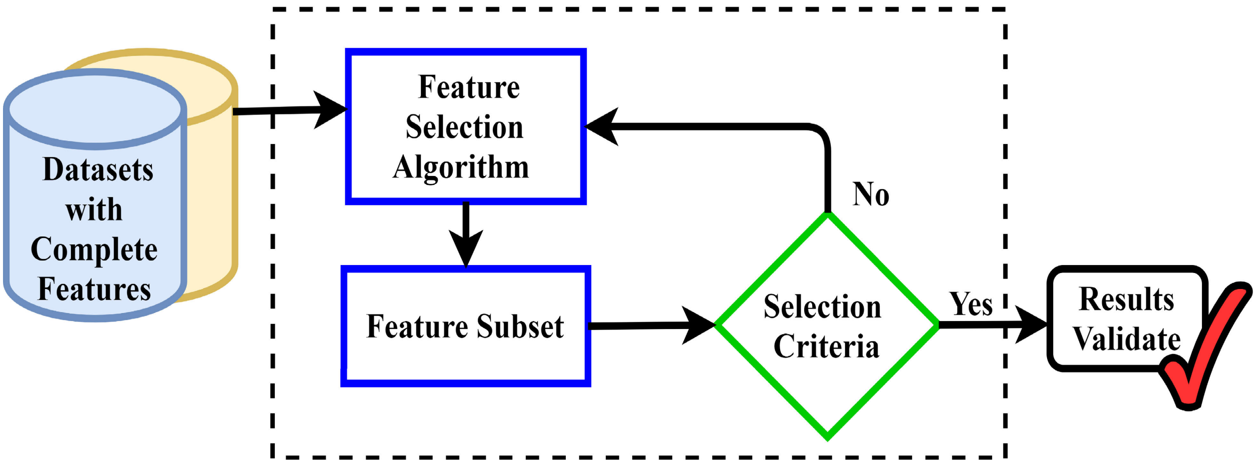

FIGURE 2. Overall feature selection process.

metaheuristic algorithms is given in Section III. The issues and challenges are presented in Section IV and Section V illustrates a case study based on feature selection problem. Section VI suggests the future work based on wrapper feature techniques. The concluding remarks are shown in Section VI.

II. BACKGROUND

This section presents the detailed description of the feature selection problem with the mathematical model and the definitions, concepts and the classifications of metaheuristic algorithms.

A. FEATURE SELECTION

Feature selection deals with inappropriate, irrelevant, or unnecessary features. It is a process that extracts the best features from the datasets [8]. Feature selection is one of the most critical and challenging problems in machine learning. The various applications of the feature selection problem can be demonstrated in different fields. There are some applications such as biomedical problems (to find the best gene from candidate gene) [9]; text mining (to find the best terms word or phrases) [10]; image analysis (to select the best visual contents pixels, colour) [11] etc. Mathematically, a feature selection problem can be formulated in the following way:

Assume a dataset $'S'$ contains $'d'$ number of features. Then the working mechanism of feature selection problem is to select relevant features among $'d'$ features.

Given dataset $S = \{f_1, f_2, f_3, \ldots, f_d\}$

The objective is to select the best subsets of features from S.

Extract Subset $D = \{f_1, f_2, f_3, \ldots, f_n\}$ where, $n < d$ and $f_1, f_2, f_3, \ldots, f_n$ represents the features/attributes of any dataset. Figure 2 depicts the working mechanism of the feature selection process. From the figure, it can be observed that there are five main components of the feature selection process, i.e. original dataset, selection of feature subset, evaluation of feature subset, selection criterion and validation. Several feature selection methods are developed to obtain the best subset of features. Generally, the techniques are classified into three categories filter, wrapper and embedded methods [12]–[15]. Filter methods are independent of learning or classification algorithm. It always focuses on the general characteristics of the data [16]. Wrapper methods always include the classification algorithm and interact with the classifier. These are computationally expensive methods than the filter and also provide more accurate results as compared to filter methods. Embedded methods are a combination of filters and wrapper methods. In embedded methods, the feature selection is a part of the training process and training process held with the classifier. Moreover, the embedded methods use learning algorithm in its process, they will be considered in wrapper approaches category [17].

Wrapper approaches present better results in comparison with filter methods, but they are slower than filters methods. Wrapper methods depend on the modelling algorithm in which every subset is generated and then evaluated. Subset generation in wrapper methods is based on the different search strategy. Jovic *et al.* [18] differentiates search techniques into three categories; exponential, sequential and randomized selection strategy. In the exponential method, the number of evaluated features increases exponentially with the size of features. Although this method shows accurate results, it is not practically possible to apply because of the high computational cost. The examples for exponential search strategy are exhaustive search, branch and bound method [19], [20]. Sequential algorithms include or remove features sequentially. Once a feature is included or removed in the selected subset, it can not be further changed that leads to local optima. Some sequential algorithms are linear forward selection, floating forward or backward selection, best first etc. Randomized algorithms include randomness to explore the search space, which saves the algorithms from trapping into local optima. Randomized algorithms are commonly known as population-based approaches for example

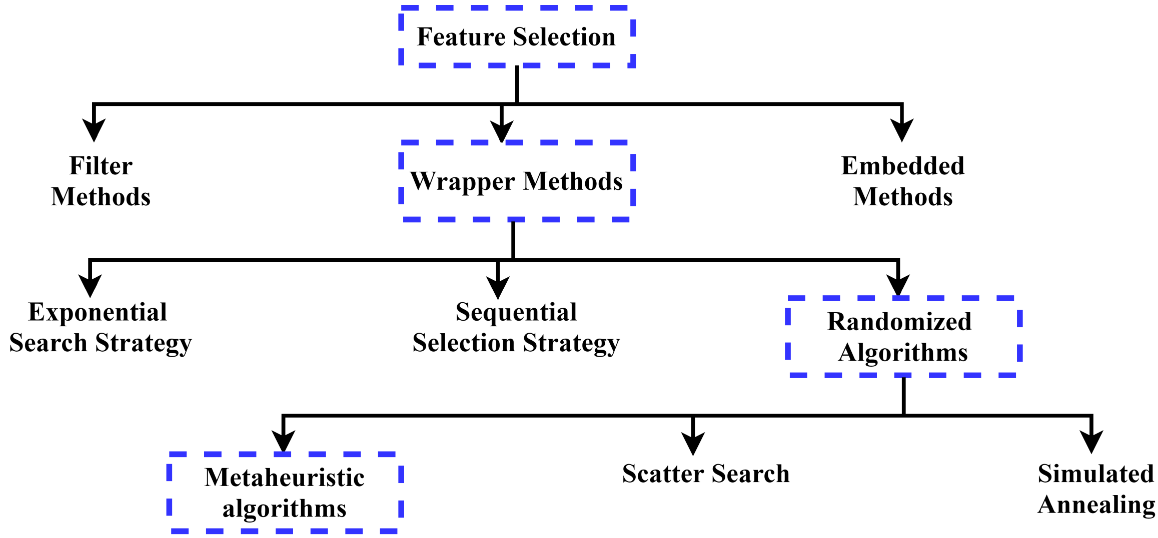

FIGURE 3. Classification of feature selection methods.

simulated annealing, random generation, metaheuristic algorithms etc. [21].

We do not present a detailed description of every method of the feature selection process. The detailed explanation of each method can be found in [1]. The flow chart of categorization of methods for solving feature selection is shown in Figure 3. In the figure, the dashed line box represents the methodology of this paper which describes how we reach to metaheuristic algorithms.

B. METAHEURISTIC ALGORITHMS

Metaheuristic algorithms are optimization methods that obtain the optimal (near-optimal) solution of optimization problems. These algorithms are derivative-free techniques and, have simplicity, flexibility and capability to avoid local optima [22]. The behaviour of metaheuristic algorithms are stochastic; they start their optimization process by generating random solutions. It does not require to calculate the derivative of search space like in gradient search techniques. The metaheuristic algorithms are flexible and straightforward due to the simple concept and easy implementation. The algorithms can be modified easily according to the particular problem. The main property of metaheuristic algorithms is that they have a remarkable ability to prevent the algorithms from premature convergence. Due to the stochastic behaviour of algorithms, the techniques work as a black box and avoid local optima and explore the search space efficiently and effectively. The algorithms make a tradeoff between its two main essential aspects exploration and exploitation [23], [24]. In the exploration phase, the algorithms investigate the promising search space thoroughly, and exploitation comes for the local search of promising area(s) that are found in the exploration phase. They are successfully applied to various engineering and sciences problems, e.g. in electrical engineering (to find the optimal solution for power generation), industrial fields (scheduling jobs, transportation, vehicle routing problem, facility location problem), in civil engineering (to design the bridges, buildings), communication (radar design, networking), data mining (classification, prediction, clustering, system modelling) etc.

Metaheuristic algorithms classify into the following two main categories;

(i) **Single solution based metaheuristic algorithms:** These techniques start their optimization process with one solution, and their solution is updated during the iterations. It may lead to trapping into local optima and also does not explore the search space thoroughly.

(ii) **Population (multiple) solution based metaheuristic algorithms:** Initially, these algorithms generate a population of solutions and start their optimization process. The population of solutions update with the number of generations/iterations. The algorithms are beneficial for avoiding local optima as multiple solutions assist each other and have a great exploration of search space. They also have the quality of jump towards the promising part of search space. Therefore, population-based algorithms use in solving most of the real-world problems.

Researchers pay great attention to metaheuristic algorithms because of their characteristics. Several algorithms have been designed and solved different types of problems. Based on their behaviour, the metaheuristic algorithms can be divided into four categories; evolution-based, swarm intelligence-based, physics-based and human-related algorithms [25]. The categorization of the algorithms is depicted in Figure 4.

(1) **Evolution based algorithms:** It is inspired from the natural evolution and start their process with randomly generated population of solutions. In these type of algorithms, the best solutions are put togther to create new individuals. The new individuals are formed using mutation, crossover and select the best solution. The most

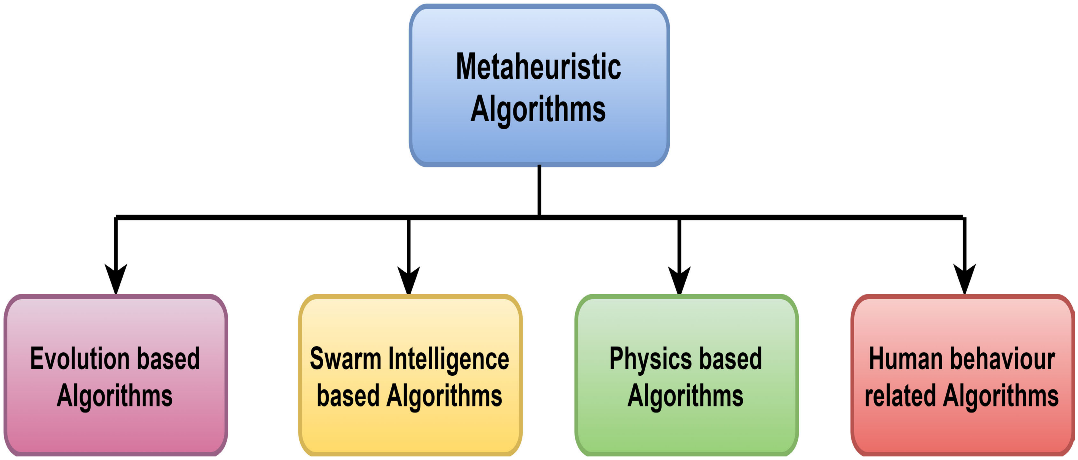

FIGURE 4. Classification of metaheuristic algorithms.

popular algorithm in this category is Genetic algorithm (GA) that is based on Darwin evolution technique [26]. There are other algorithms such as evolution strategy [27], genetic programming [28], tabu search [29], differential evolution [30] etc.

(2) **Swarm intelligence-based algorithms:** These algorithms are inspired by the social behaviours of insects, animals, fishes or birds etc. The popular technique is Particle Swarm Optimization (PSO) developed by Kennedy and Eberhart [31]. It is inspired by the behaviour of a group of birds that fly throughout the search space and find their best location (position). Ant Colony optimization [32], Honey bee swarm optimization algorithm [33], monkey optimization [34] etc are the examples of swarm intelligence algorithms.

(3) **Physics based algorithms:** These are inspired by the rules of physics in the universe. Simulated annealing [35], Harmony search [36] etc come under physics-based algorithms.

(4) **Human behaviour related algorithms:** These techniques are purely inspired by human behaviour. Every human being has its way of doing activities that affect its performance. It motivates researchers to develop the algorithms. The popular algorithms are Teaching learning-based optimization algorithm (TLBO) [37], League Championship algorithm [38] etc.

It is worth mentioning here that there are many metaheuristic algorithms developed from 1966 to till now. In this paper, we present the literature of those algorithms which are developed or proposed since 2009 to 2019 (ten-year span). According to the category, the list of metaheuristic algorithms are presented in Table 1, 2, 3, 4. The first column of the tables present the abbreviation; the second column gives the name of the algorithms, which is followed by the developed year. These algorithms have been applied to solve many real-world applications but, this paper is restricted to present the application in feature selection problems. Therefore, in the following sections, the algorithms which have been developed in ten years span and applied to feature selection problems are discussed.

TABLE 1. List of Evolution based algorithms from 2009-2019.

Abbreviation	Algorithm	Year
DS [39]	Differential search algorithm	2011
BSA [40]	Backtracking search optimization	2013
SFS [41]	Stochastic fractal search	2014
SFO [42]	Synergistic fibroblast optimization	2018

III. METAHEURISTIC ON FEATURE SELECTION

It describes the metaheuristic algorithm, which has been used in solving the feature selection problem. Binary vectors representations are considered to obtain the relevant feature. In the designed algorithm, a solution vector is represented by (10101100.....) this implies that 1 means that a particular feature is selected and 0 means that feature is not selected in the subset. Hence, this section investigates all binary variants of metaheuristic algorithms in detail. The first section describes the evolution-based algorithms; the second describes the swarm intelligence based algorithms, third demonstrates the physics-based algorithms, and the fourth one is for the human-related algorithm. And the last section is for the hybrid algorithms, which are a combination of two or more metaheuristic algorithms that have been used for classification problems.

A. EVOLUTION BASED ALGORITHMS

From Table 1, it can be seen that there are very few algorithms are developed in evolution based category from 2009-2019. Gan and Duan [123] proposed a chaotic differential search algorithm for image processing and it has been combined with lateral inhibition to edge extraction and image enhancement. Negahbani *et al.* [124] used differential search algorithm for the diagnosis of coronary artery

TABLE 2. List of Swarm Intelligence based algorithms from 2009-2019.

Abbreviation	Algorithm	Year
B [43]	Bumblebees algorithm	2009
PFA [44]	Paddy field algorithm	2009
CS [45]	Cuckoo search	2009
GSO [46]	Group search optimizer	2009
CGS [47]	Consultant guided search	2010
BA [48]	Bat algorithm	2010
TCO [49]	Termite colony optimization	2010
HS [50]	Hunting Search	2010
ES [51]	Eagle strategy	2010
HSO [52]	Hierarchical swarm optimization	2010
FA [53]	Firefly algorithm	2010
FOA [54]	Fruit fly optimization algorithm	2011
ECO [55]	Eco inspired evolutionary algorithm	2011
WSA [56]	Weightless swarm algorithm	2011
FPA [57]	Flower pollination algorithm	2012
BMO [58]	Bird mating optimizer	2012
ACS [59]	Artificial cooperative search algorithm	2012
KH [60]	Krill herd algorithm	2012
FROGSIM [61]	Japanese tree frogs calling algorithm	2012
OptBees [62]	The OptBees algorithm	2012
WSA [63]	Wolf search algorithm	2012
TGSR [64]	The great Salmon run algorithm	2012
DE [65]	Dolphin echolocation	2013
SSO [66]	Swallow swarm optimization algorithm	2013
EVOA [67]	Egyptian vulture optimization algorithm	2013
CSO [68]	Chicken swarm optimization	2014
AMO [69]	Animal migration optimization	2014
GWO [22]	Grey wolf optimization	2014
SSO [70]	Shark smell optimization	2014
ALO [71]	Ant lion optimizer	2015
BSA [72]	Bird swarm algorithm	2015
VCS [73]	Virus Colony search	2015
AAA [74]	Artificial algae algorithm	2015
DA [75]	Dragonfly algorithm	2015
DSOA [76]	Dolphin swarm optimization algorithm	2016
CSA [77]	Crow search algorithm	2016
WOA [78]	Whale optimization algorithm	2016
MBF [79]	Mouth brooding fish algorithm	2017
ABO [80]	Artificial Butterfly Optimization	2017
SHO [81]	Selfish herd optimizer	2017
GOA [82]	Grasshopper optimization algorithm	2017
SSA [83]	Salp swarm algorithm	2017
SHO [84]	Spotted hyena optimizer	2017
EPO [85]	Emperor penguin optimizer	2018
SSA [86]	Squirrel search algorithm	2018
BOA [87]	Butterfly optimization algorithm	2019
EPC [88]	Emperor Penguins Colony	2019

disease with fuzzy c-means that was used as a classifier. The performance of the proposed approach has been evaluated using accuracy, sensitivity and specificity measures. Zhang *et al.* [125] proposed binary backtracking algorithm for wind speed forecasting in which extreme learning machine was employed for feature selection. Binary backtracking algorithm was developed using a sigmoidal function that transforms the continuous variables to binary variables. To identify the Leukemia cancer symptoms, Dhal *et al.* [126] implemented the stochastic fractal search algorithm to provide optimal identification. The developed algorithm was compared with other classical methods and achieved high accuracy. Besides, a binary stochastic fractal search was developed to classify the galaxy colour images with extreme machine learning [127].

TABLE 3. List of Physics based algorithms from 2009-2019.

Abbreviation	Algorithm	Year
GSA [89]	Gravitational search algorithm	2009
CSS [90]	Charged system search	2010
GbSA [91]	Galaxy based search algorithm	2011
EMO [92]	Electro magnetism optimization	2011
ACROA [93]	Artificial chemical reaction optimization algorithm	2011
Spiral [94]	Sprial optimization	2011
BH [95]	Black hole algorithm	2012
WCA [96]	Water Cycle algorithm	2012
CSO [97]	Curved Space optimization	2012
RO [98]	Ray optimization	2012
MBA [99]	Mine blast algorithm	2013
GBMO [100]	Gases Brownian Motion Optimization	2013
ACMO [101]	Atmosphere clouds model	2013
KGMO [102]	Kinetic gas molecule optimization	2014
CBO [103]	Colliding bodies optimization	2014
WSA [104]	Weighted superposition attraction	2015
LSA [105]	Lightning search algorithm	2015
SCA [106]	Sine cosine algorithm	2016
WEO [107]	Water Evaporation Optimization	2016
MVO [108]	Multi-verse optimizer	2016
LAPO [109]	Ligtning attachment procedure optimization	2017
ES [110]	Electro-search algorithm	2017
TEO [111]	Thermal exchange optimization	2017
F3EA [112]	Find fix finish exploit analyze	2019

TABLE 4. List of Human behaviour related algorithms from 2009-2019.

Abbreviation	Algorithm	Year
LCA [38]	League championship algorithm	2009
HIA [113]	Human inspired algorithm	2009
SEOA [114]	Social emotional optimization	2010
BSO [115]	Brain storm optimization	2011
TLBO [37]	Teaching learning based optimization	2011
ASO [116]	Anarchic society optimization	2012
SLC [117]	Soccer League competition	2013
SBA [118]	Social based algorithm	2013
EMA [119]	Exchange market algorithm	2014
GCO [120]	Group Counseling optimization	2014
JA [121]	Jaya Algorithm	2016
VPL [122]	Volleyball premier league algorithm	2018
GSK [25]	Gaining sharing knowledge based algorithm	2019

B. SWARM INTELLIGENCE BASED ALGORITHMS

This section presents a detailed description of some well-known algorithms which are based on the different swarm behaviour and modified to solve feature selection problems. We have given our best to present the review of algorithms with their modifications, many datasets used and some other information which will be useful to provide a quick idea about the published research paper.

1) CUCKOO SEARCH

Cuckoo search algorithm was developed by observing the behaviour of cuckoo birds and their reproduction strategy. It is very well known and popular algorithm and has achieved great success in solving various real-world problems. However, several binary versions of cuckoo search algorithm have been developed in solving binary optimization problems. In 2012, Tiwari [128] used CS algorithm for face recognition problem. Firstly, features were extracted using discrete cosine transformation which worked as host egg in CS algorithm. It proved its efficiency for finding out the most matching image in face recognition. To obtain the optimal feature subset, Rodrigues *et al.* [129] proposed a binary CS algorithm (BCS) using a function which converts the continuous variable to its binary form. BCS has successfully applied to two datasets of theft detection in a power system with optimum path forest classifier and obtained that BCS was the very suitable and fastest approach in solving feature selection for industrial datasets. Gunavathi and Premalatha [130] used CS algorithm for classification in microarray gene data of cancer. The features were ranked according to T and F-statistics, and KNN classifier was used as a fitness function. The results showed that CS algorithm obtained 100% accuracy in most of the datasets. Sudha and Selvarajan [131] proposed an enhanced CS algorithm to find the optimal features for breast cancer classification. In ECS, classification accuracy was used as an objective function and evaluated using KNN classifier. Salesi and Cosma [132] extended BCS algorithm in which pseudobinary mutation neighbourhood search was designed to solve feature selection problems of biomedical datasets. Pandey *et al.* [133] introduced binary binomial cuckoo search algorithm to select the best features and applied to fourteen benchmark datasets of UCI [134]. Although lots of applications of machine learning have been solved by developing different versions of CS algorithm [135]–[137].

2) BAT ALGORITHM

It is inspired by the behaviour of bats and very popular in solving various real-world problems. In solving feature selection problems, firstly Nakamura *et al.* [138] developed a binary version of BA (BBA) with sigmoid function to restrict the position of bat's to binary variables. The accuracy was calculated using optimum path forest classifier, and BBA was applied to five datasets. Laamari and Kamel [139] improved the performance of BBA by applying other V-shaped transfer function with SVM classifier to intrusion detection systems. Rodrigues *et al.* [140] used different transfer functions for obtaining binary-based optimization techniques with optimum path forest classifier. The binary BA presented best results with hyperbolic tangent transfer function in most of the datasets. Enache and Sgârciu [141] improved the version of BBA with different classifiers such as SVM and C4.5 and applied to NSL-KDD datasets for intrusion detection systems. Yang *et al.* [142] modified BA to improve the diversity of the population of bats so that it made a good balance between exploration and exploitation and solve the feature selection problem. To classify the MR brain tumour image, Kaur *et al.* [143] modified BA by combining Fisher and parameter-free bat algorithm for good exploration. The dataset has been taken from UCI repository and applied with SVM classifier. The evaluation measure such as the number of features, computational time and accuracy of the classifier were used. Naik *et al.* [144] used BBA with one-pass generalized classifier neural network to estimate the number of selected features. The fitness function was formed using sensitivity and specificity with the classification accuracy. The proposed approaches were evaluated using six available microarray datasets, and the obtained results were compared with the different type of classifiers. Moreover, several modifications have been done to improve the performance of BBA and applied to various real-world classification problems [145]–[148].

3) FIREFLY ALGORITHM

It imitates the mechanism of firefly mating and exchange of information using light flashes. Emary *et al.* [149] proposed first binary version of firefly algorithm (FFA) to solve feature selection problems using a threshold value. The developed algorithm made a good exploration quality which found a quick solution to the problem. It has been applied to benchmark datasets of UCI with KNN classifier. Kanimozhi and Latha [150] presented a technique for image retrieval by using SVM classifier and FFA. The main aim was to increase the performance of the algorithm with optimal features, and the algorithm has been tested over Corel Caltech and Pascal database images. To predict the disease, cardiotocogram data has been used with SVM and FFA by Subha and Murugan [151]. Zhang *et al.* [152] proposed a return cost based FFA for a public dataset in which binary movement operator has been used to update the position of fireflies. The proposed approach proved that it was very competitive with other algorithms. A self-adaptive FFA has been developed for feature selection using mutual information criterion [153]. Two modifications have been done for getting rid of trapping into local optima. Twelve datasets has been considered for evaluating the performance of the algorithm. Xu *et al.* [154] combined binary FFA with opposition based learning algorithm in solving the feature selection problem and applied to ten datasets. For network intrusion detection, FFA algorithm has been used with C4.5 and Bayesian networks classifier and utilized for KDD CUP 99 datasets [155]. For more versions of FFA in various applications of feature selection, interest researchers can be found in [156]–[159].

4) FLOWER POLLINATION ALGORITHM

FPA is inspired from the pollination procedure of flowers. Several binary variants of FPA were developed to solve feature selection problem. Firstly, Rodrigues *et al.* [160] proposed a binary constrained version of FPA (BFPA) in which binary solutions were obtained after generating local pollinations. The BFPA run with optimum path forest classifier, which calculated the performance accuracy. The obtained results were compared with other state-of-the-art metaheuristic algorithms PSO, HS and FA and proved that BFPA is also suitable to get the optimal feature subsets. To improve the performance of BFPA, Zawbaa and Emary [161] used KNN classifier with a modified binary variant of FPA. In binary version, a threshold value was used to get a binary string from continuous variables. Moreover, the proposed algorithm was evaluated on a multi-objective function for classification and regression datasets. It showed that the algorithm outperformed PSO, GA and BA.

Rajamohana *et al.* [162] used different values of λ parameter and profounded an adaptive scheme of BFPA i.e. ABFPA and solved feature selection problem. For text classification, Majidpour *et al.* [163] used BFPA with ada-boost algorithm, which worked as a classifier. Sayed *et al.* [164] considered clonal selection algorithm for exploitation and FPA for exploration to form the binary clonal FPA. The proposed CFPA was applied to three datasets i.e. Australian, breast cancer and german number and obtained the results. Yan *et al.* [165] improved the proposed version by considering a group strategy to avoid local optima, adoptive transfer function was used for binary encoding, and Gaussian mutation strategy was used for exploitation. Based on these modifications, improved version IBCFPA was applied to six biomedical datasets with different classifiers (SVM, KNN and NB) and obtained optimal feature subsets for every dataset.

5) KRILL HERD ALGORITHM

It is based on the movement of Antarctic krills to search their food and to increase the density. Rodrigues *et al.* [166] solved feature selection problem by introducing the binary variant of KH algorithm (BKH) that generated binary vectors by evaluating the transfer operator. Optimum path forest used as a classifier and the proposed algorithm BKH applied to six datasets and presented the results. Mohammadi *et al.* [167] considered breast cancer datasets for extracting the fuzzy rules. They modified the KH algorithm as binary krill herd fuzzy rule miner (BKH-FRM) that choose the best krill as well as the local best krill among the population of krills. The results obtained by BKH-FRM was compared with other ten metaheuristic algorithms and presented high accuracy among others. Rani and Ramyachitra [168] classified the cancer types using KH algorithm and random forest classifier. They modified the algorithm by using a horizontal crossover and position mutation operator and applied to ten different gene microarray cancer datasets. For microarray data classification, Zhang *et al.* [169] improved BKH algorithm named as IG-MBKH by using a hyperbolic tangent function and adaptive transfer function. Furthermore, to initialize the population, information gain feature ranking was used to explore the search space efficiently. IG-MBKH was applied to high dimensional microarray datasets that generated relevant feature subsets. K-NN classifier used for calculating the accuracy of the selected features.

6) GREY WOLF OPTIMIZER

It is based on the hunting process of a pack of grey wolves in nature. In the first binary version of GWO, Emary *et al.* [170] used the sigmoidal transfer function to get the binary vectors (bGWO). To calculate the classification accuracy, KNN classifier was utilized and applied to eighteen different UCI datasets. Moreover, in the initialization phase, small, random and large initialization techniques were used for good exploration. Sharma *et al.* [171] proposed a modified version of GWO for identifying the symptoms of Parkinson's disease with random forest, KNN and decision tree classifier. Pathak *et al.* [172] proposed levy flight GWO which was used to select the relevant features from the original datasets. It used random forest classifier for image steg analysis and applied to Bossbase ver 1.01 datasets. The obtained results showed its great performance for achieving great convergence. Devanathan *et al.* [173] identified the optimal features for Indirect Immunofluorescence image by presenting another version of binary GWO. To diagnosis of cardiovascular disease, Al-Tashi *et al.* [174] used GWO algorithm for selecting the best features and SVM used as a classifier. The proposed approach has been applied to Cleveland dataset, which is freely available and performed greatly. Moreover, the author proposed a binary version (BMOGW-S) by using a sigmoidal function for solving multi-objective feature selection problem in which the artificial neural network was used for classification. BMOGW-S applied to fifteen benchmark datasets and compared with MOGWO with tanh transfer function [175]. Hu *et al.* [176] proposed new transfer functions and new updating scheme for parameters of GWO. Advanced GWO (ABGWO) applied to twelve datasets of UCI and showed superior results as compared to other algorithms. There are several versions of GWO are developed for classification in different fields such as medical diagnosis [177], cervical cancer [178], electromyography (EMG) signal [179], facial emotion recognition [180], text feature selection [181] etc.

7) ANT LION OPTIMIZER

ALO is a very popular algorithm and inspired from the hunting procedure of antlion insects and ants. It has been applied to various real-world problems to find out the optimal (near-optimal) solutions. To solve the feature selection problem, Zawbaa *et al.* [182] used a threshold value at continuous variables to propose the binary variant of ALO algorithm. The proposed algorithm BALO with K-NN classifier was applied to eighteen different datasets and compared with well-known metaheuristic algorithms GA and PSO. They have used different evaluation criteria such as average classification accuracy, the average number of the selected feature, average Fisher score (F-score) etc. for calculating the performance. In the next year, Emary *et al.* [183] proposed different variants of BALO in which each individual changed its position according to the crossover operator between two binary solutions. The binary solutions were obtained by applying S and V-shaped transfer functions or simply by using basic operators of ALO. Moreover, three initialization methods were adopted for a good exploration of the search space and concluded that initialization process affected the searching quality and performance of the algorithm. Although, Mafarja *et al.* [184] proposed six different binary variants of

ALO using S and V-shaped transfer functions and obtained optimal feature subsets.

To improve the efficiency of proposed binary versions, Zawbaa *et al.* [185] applied chaos theory to some parameter of ALO algorithm. Considering of chaotic maps in ALO algorithm, CALO avoided the local optima and made a proper balance between exploration and exploitation quality. CALO algorithm applied to ten biological datasets, and eight datasets form other categories. The obtained solutions proved its robustness and outperformed ALO, PSO and GA.

To save the algorithm from trapping into local optima, Emary and Zawbaa proposed a new approach of a binary variant of ALO [186]. Levy flight random walk was used to generate solutions, and five different initialization methods were utilized to generate the initial solutions. Other than, Mafarja and Mirjalili [187] embedded two rough set filter approach Quick Reduct and CEBARKCC with the ALO algorithm, which improved the initial population as well as the final optimal solution. For hyper spectral image classification, a modified version of ALO algorithm (MALO) was proposed with wavelet SVM (WSVM) classifier [188]. MALO with WSVM performed better in most of the datasets and proved that it was beneficial in solving the feature selection problems. Azar *et al.* [189] combined rough set theory with ALO algorithm and applied to different datasets.

8) DRAGONFLY ALGORITHM

It is inspired by the behaviour of dragonflies in nature and applied to various problems. Medjahed *et al.* [190] proposed a complete diagnosis procedure of cancer using binary dragonfly (BDF) algorithm with SVM. In this, SVM-RFE (SVM-recursive feature elimination) used to extract the gene from the datasets, and BDF was used to enhance the performance of SVM-RFE. The proposed algorithm was applied to six microarray datasets and presented high accuracy results. Mafarja *et al.* [191] proposed a binary version of DA using a transfer function and solved the feature selection problem with several datasets. Moreover, they also introduced a binary variant of dragonfly algorithm using time-varying transfer functions which make a proper balance between exploration and exploitation. These approaches have been applied to UCI datasets and compared with other state-of-the-art metaheuristic algorithms [192]. Karizaki and Tavassoli [193] used filter and wrapper approaches simultaneously in which BDA algorithm was used to find optimal subset of features and ReliefF algorithm was used as a filter approach. It has been applied to five datasets and obtained the results. The BDA algorithm was applied to different learning algorithm such as Naik *et al.* [194] used BDA with radial basis neural network function and selected the features from microarray gene data. Several other binary versions of DA have been proposed to solve features selection problem that can be found in [195]–[198].

9) WHALE OPTIMIZATION ALGORITHM

WOA is based on the special behaviour of the hunting method of humpback whales. To solve the binary optimization problems, Hussien *et al.* [199], [200] utilized S and V-shaped transfer function in conventional WOA and solved feature selection problem with eleven UCI datasets in 2017. For classification, KNN classifier was used, which ensured the selected features for their relevancy. The proposed approach bWOA showed its capability for obtaining the maximum accuracy and the minimum number of selected features. To enhance the proposed approach, Sayed *et al.* [201] presented a chaotic whale optimization algorithm (CWOA) with ten chaotic maps. The chaotic maps were used in place random parameters that made a better tradeoff between two main important properties of algorithm exploration and exploitation. Tubishat *et al.* [202] classified Arabic datasets for sentiment analysis by proposing improved WOA (IWOA). IWOA incorporated the evolutionary operators such as crossover, mutation and selection as in differential evolution. IWOA applied to four publicly available datasets and compared with other techniques. Mafarja and Mirjalili [203] proposed two binary variants of WOA by embedding crossover and mutation operator and by using tournament and roulette wheel selection in WOA. Twenty benchmark datasets have been utilized in this approach. Agrawal *et al.* [204] proposed a new version of WOA that is based on quantum concepts in which, quantum bit representation was used for all individuals. And the new version was applied to fourteen datasets. There are some other versions of WOA [205]–[207] which are presented in solving feature selection problems.

10) GRASSHOPPER OPTIMIZATION ALGORITHM

GOA is based on the lives of grasshopper how their behaviour of living changes. Ibrahim *et al.* [208] proposed GOA with SVM classifier (GOA+SVM) in which the parameters of SVM were also optimized by using GOA. GOA+SVM applied to biomedical datasets for Iraqui cancer patients and obtained the results. To deal with immature convergence of GOA, Mafarja *et al.* [209] proposed GOA_EPD algorithm by using evolutionary population dynamic, roulette wheel and tournament selection for guiding the agent. Twenty-two benchmark datasets were considered for evaluating the performance of the proposed approach. Hichem *et al.* [210] introduced a new transfer function Hamming distance which converted continuous variables into a binary vector. The new version of GOA (NBGOA) utilized for 20 standard datasets and compared with other versions of GOA. The presented results showed the ability to achieve great performance of NBGOA. Sigmoidal and V-shaped transfer functions were used with mutation operator to enhance the exploration quality of BGOA by Mafarja *et al.* [211].

11) SALP SWARM ALGORITHM

SSA is inspired by the salps swarming attitude. In 2017, Ibrahim *et al.* [212] first time used SSA for solving feature selection problem (SSA-FS) by applying a threshold value of 0.5 to build binary vectors. SSA-FS applied to medical datasets of breast, bladder and colon cancer datasets and compared with other algorithms. Sayed *et al.* [213] proposed chaotic SSA with ten chaotic maps and transfer function. KNN classifier has been used to evaluate the classification accuracy and applied to twenty benchmark datasets. Faris *et al.* [214] developed an efficient binary SSA using S and V-shaped transfer functions in which a crossover operator was embedded in place of the average operator to enhance the exploration quality. The proposed approach has been used with KNN classifier and applied to twenty-two well known UCI datasets and obtained that S-shaped transfer function provided best results. To get rid of trapping into local optima and enhance the exploration and exploitation of SSA, salp's position was updated using the singer's chaotic map and used local search algorithm by Tubisat *et al.* [215]. It has been applied to twenty benchmark datasets and three Hadith datasets. Hegazy *et al.* [216] improved SSA (ISSA) by inserting weight to adjust the presented best solution and classified by KNN classifier. ISSA was utilized to twenty-three UCI datasets and compared with basic SSA and four other metaheuristic algorithms. Other versions of SSA was used for feature selection problems that can be found in [217], [218].

12) EMPEROR PENGUINS OPTIMIZER

EPO is based on the huddling behaviour of emperor penguins. Baliarsingh *et al.* [219] proposed multi-objective binary EPO with chaos to apply on high-dimensional biomedical datasets. The proposed methodology used for cancer classification and obtaining the optimal feature subset. The obtained results show the efficiency of the proposed algorithm in terms of accuracy, specificity and F-score. Baliarsingh and Vipsita [220] proposed an intelligence hybrid technique for gene classification in which extreme learning machine is used with chaos EPO. The proposed hybrid technique is employed on seven well known microarray datasets and the experimental results show the efficacy of the hybrid technique. To obtain the optimal feature subset, Dhiman *et al.* [221] developed binary EPO (BEPO) with S and V-shaped transfer functions which convert the continuous search space to binary space. The BEPO is applied to different feature selection datasets and the obtained results show better results as compared to others.

C. PHYSICS-BASED ALGORITHMS

Various algorithms have been developed that are based on the rules of physics. The binary versions of physics-based algorithms which have been applied to feature selection problems are discussed in the following Table 10 and 11. In Tables 10 and 11, the first column represents the name of modified algorithms, the second column gives details of the transfer function that used for deciding the binary variables and the third column shows the name of the classifier which operated as a learning algorithm in the optimization process. Furthermore, the table describes the considered datasets, evaluation metrics (the performance of the proposed algorithm is compared with these measures), name of the compared techniques and some other information regarding the proposed approaches. We find the following algorithms such as multi-verse optimizer [222], sine-cosine algorithm, gravitational search algorithm etc. which are developed to obtain the optimal subset of features on different datasets.

D. HUMAN RELATED ALGORITHMS

It gives a summary of the human-related algorithms in solving feature selection problems. In the following description, we have discussed three algorithm brain storm optimization, teaching-learning optimization and gaining sharing knowledge-based algorithm.

1) BRAIN STORM OPTIMIZATION

The algorithm works on the mechanism of human brainstorming. BSO algorithm was also applied in data classification. Papa *et al.* [223] integrated binary BSO using different S and V-shaped transfer functions. The proposed approach ran over optimum path forest classifier and tested over several Arizona State University's datasets. Pourpanah *et al.* [224] used fuzzy min-max neural network learning model with binary BSO algorithm and applied to a real-world dataset. They also introduced a fuzzy ARTMAP model with BSO algorithm [225]. Tuba *et al.* [226] used BSO algorithm with SVM classifier for medical classification. Moreover, the parameters of SVM were optimized using BSO algorithm. Oliva and Elaziz [227] proposed a new version of BSO algorithm for its good exploration quality. They have introduced chaotic maps and opposition based learning algorithm for initialization of the solution. Disruptor operator was used to updating the initial population. The modified version was used for classification and in order to obtain the optimal features, eight datasets have been considered from UCI repository.

2) TEACHING BASED LEARNING OPTIMIZATION

TLBO algorithm is based on the influence of the teacher on the students in the class. Krishna and Vishwakarma [228] proposed an improved version of TLBO algorithm with wavelet transform function for fingerprint recognition. Jain and Bhadauria [229] selected optimal features using TLBO algorithm and SVM classifier for image retrieval datasets. Kiziloz *et al.* [230] presented a multi-objective TLBO algorithm to select the features in binary classification problems. The algorithm was tested over well-known datasets of UCI with three supervised learning algorithm logistic regression, SVM, and extreme learning machine. Among all three classification model, logistic regression with TLBO presented best results in most of the datasets. Allam and Nandhini [231] developed a binary TLBO (BTLBO) with a threshold value to restrict variables into binary form. They have used different classifier and used for the classification of breast cancer datasets. The proposed approach showed its high accuracy with the minimum number of features. The TLBO algorithm has been applied to chronic kidney disease dataset with the improved version [232]. The fitness function was evaluated using Chebyshev distance formula and obtained the results.

3) JAYA ALGORITHM

Jaya algorithm is based on the framework of TLBO algorithm with only one phase. It has been successfully applied to benchmark functions. Das *et al.* [233] modified JA to find the optimal feature subset by using a search technique which updates the worst features. The proposed approach has been tested over ten benchmark datasets with other optimizer for comparison. The results show its efficacy to find the optimal feature subset. Using the S-shaped transfer function with JA, Awadallah *et al.* [234] developed a binary JA in which adaptive mutation rate has been used. The adaptive mutation rate controls the diversification in the search space. The proposed approach BJAM is applied to twenty two benchmark datasets with KNN classifier and obtained the optimal feature subset.

4) GAINING SHARING KNOWLEDGE BASED ALGORITHM

It is based on the concept of gaining and sharing knowledge among humans. Agrawal *et al.* [235] proposed the first novel binary version of GSK algorithm for feature selection problem (FS-NBGSK) by introducing binary junior and senior gaining and sharing stages. The FS-NBGSK algorithm was tested over 23 benchmark datasets of UCI repository with KNN classifier. The approach showed the best results among the compared algorithm in terms of accuracy and a minimum number of selected features.

E. HYBRID METAHEURISTIC ALGORITHMS

Hybrid metaheuristic algorithms mean to combine the best operators from different metaheuristic algorithms and develop a new enhanced algorithm. In recent years, hybrid algorithms have been achieved great attention in solving optimization problems. In particular, for feature selection problem, many hybrid metaheuristic algorithms have been developed to obtain the relevant and optimal feature subset from the original dataset. A lot of possibilities occur to produce a more enhanced algorithm which finds the solution optimally. The enhance algorithms help to get rid of trapping into local optima, free from premature convergence, explore the search space efficiently and effectively and make good exploitation. Moreover, the enhanced algorithms obtain the optimal or near-optimal solution and make a better tradeoff between exploration and exploitation quality of an algorithm. There are some algorithms described in which the best qualities of different algorithms were combined to develop a new one.

Hafez *et al.* [236] proposed MAKHA algorithm in which the jump process was taken from monkey algorithm and evolutionary operators (mutation, crossover) were used from krill herd algorithm to find the optimal solution quickly. The MAKHA algorithm was checked over eighteen datasets of UCI with KNN classifier and obtained the classification accuracy. Simulated annealing (SA) is the most popular and very promising algorithm from physics-based category. Hence, to enhance the performance of whale optimization algorithm, Mafarja and Mirjalili [237] embedded SA into WOA. It boosted the exploitation of WOA by improving the best solution found after each iteration. The performance of the hybrid algorithm WOA-SA was tested over eighteen datasets with KNN classifier. Arora *et al.* [238] used the position updation quality of crow search algorithm in grey wolf optimizer to make a good balance between exploration and exploitation. It hybridized the algorithm as GWOCSA which applied to twenty-one well-known datasets of UCI repository. THe GWOCSA algorithm restricted to binary search space using the S-shaped transfer function. The accuracy of considered KNN classifier was compared with other state-of-the-art metaheuristic algorithms. To get rid of local optima in sine cosine algorithm, Abd Elaziz *et al.* [239] proposed a hybrid algorithm by employing the local search method of differential evolution algorithm. The enhanced version of sine cosine algorithm were tested over eight UCI datasets and presented better results in case of performance measures and statistical analysis. Tawhid and Dsouza [240] developed a hybrid algorithm using bat algorithm and enhanced version of PSO algorithm for solving feature selection problem in binary space. To transform the position of bats in binary space, a V-shaped transfer function was used, and in the same way, an S-shaped transfer function was employed to get the binary position of a particle in PSO algorithm. Hybrid algorithm HBBEPSO combined good exploration of bat algorithm and convergence characteristic of the PSO algorithm that obtained the optimal features of twenty standard datasets. The obtained results proved its ability to have high accuracy among all the compared algorithms. To select the gene for preprocessing technique in cancer classification, binary black hole algorithm was employed with PSO algorithm to enhance the exploration and exploitation efficiently and effectively [241]. Different classifiers have been used to evaluate the performance accuracy and obtained the results of two standard and three clinical microarray datasets. Neggaz *et al.* [242] presented a hybrid algorithm to boost the salp swarm algorithm with the help of sine cosine algorithm in solving the feature selection problem. Baliarsingh [243] embedded social engineering optimizer in EPO to enhance the performance of EPO. The SVM classifier is modified with memetic algorithm and it is applied with the proposed hybrid approach to medical datasets. The proposed hybrid algorithm is compared with other well known metaheuristic algorithms and outperforms the other algorithms. Another hybrid approach of EPO is proposed with the cultural algorithm for face recognition [244]. The proposed approach enhances the capability of the existing one and is applied with SVM classifier for face recognition. It presents the best results in terms of convergence, robustness etc. With the use of sine cosine algorithm, the exploration phase was enhanced and also avoided to get into premature convergence. To get the optimal gene from gene expression data, Shukla *et al.* [245] combined teaching learning based optimization with SA algorithm. SA algorithm utilized to enhance the solution quality of TLBO algorithm

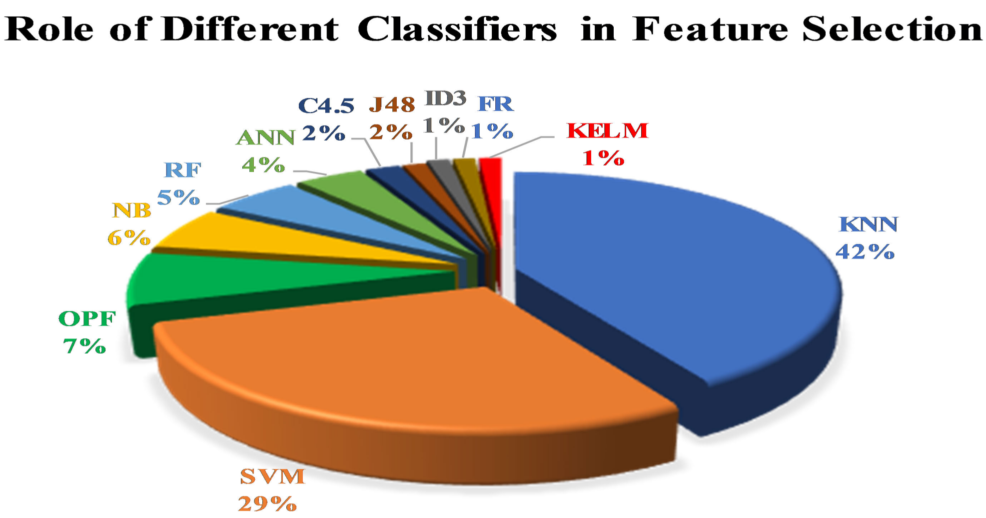

FIGURE 5. Role of different classifiers in solving feature selection problems.

and helped to find the relevant genes for the detection of cancer. Moreover, a new V-shaped transfer function was proposed to convert the variables into the binary variables. The classification accuracy was evaluated with SVM classifier and also tested over ten sets of microarray datasets. Several combinations of different metaheuristic algorithms have been developed in solving different application of feature selection problem. Jaya algorithm is used with forest optimization algorithm in gene selection [246]. The use of enhanced JA is to optimize the two parameters of forest optimization algorithm. This hybrid approach has been employed to the microarray datasets and outperformed the other optimizers. In the text feature selection, grey wolf optimizer and grasshopper optimization algorithm were employed [247], for industrial foam injection processes, PSO algorithm and gravitational search algorithm were used [248]. A combination of the grey wolf and stochastic fractal search algorithm [249] and grasshopper and cat swarm optimization algorithm are used for feature selection [250].

IV. ISSUES AND CHALLANGES

Despite achieving great success of metaheuristic algorithms in solving feature selection problems, some challenges and issues occur that will be described in the following sections:

A. SCALABILITY AND STABILITY

In real-world problems, a dataset contains thousands or even millions of features. To handle the large datasets in feature selection problem, the designed algorithm must be scalable. The designed algorithm must have a good scalable classifier which handles large dataset [251]. Therefore, scalability is an essential task for developing an algorithm to solve the feature selection problem.

Another important issue for designing an algorithm to solve the feature selection problem is stability. An algorithm is said to be stable for feature selection if it finds the same subset of features for a different sample of datasets. For trying to find the best classification, feature selection algorithm becomes unstable in most of the cases. Instability comes when there is a high correlation in the features, and they are removed because of obtaining the best classification accuracy. Therefore, stability is as important as the classification accuracy. The possible solution to make the algorithm stable for feature selection problem can be found in [252], [253].

B. CHOICE OF CLASSIFIER

To design a wrapper feature selection algorithm, the choice of a classifier has a great impact on the quality of the obtained solution. In solving feature selection problem using metaheuristic algorithm, there are different types of classifiers have been used such as K-nearest neighbour (KNN), Support Vector machine (SVM), Optimum Path Forest (OPF), Naive Bayesian (NB), Random Forest (RF), Artifical Neural Network (ANN), ID3, C4.5, Fuzzy rule based (FR), Kernel Extreme Learning Machine (KLM). The role of classifiers used in feature selection problems is shown in Fig. 5. KNN is most commonly used classifier with different datasets of UCI repository. In contrast, the SVM classifier is used frequently in intrusion detection systems and medical field datasets such as in cancer detection, artery disease etc. Additionally, Xue *et al.* [1] investigated that on medium-size datasets of features SVM algorithm was adopted to present the prominent features with the time constraints. However, KNN classifier is the most used classifier among all, and its benefits to applying for large dimensional datasets.

C. CONSTRUCTION OF OBJECTIVE FUNCTION

To select the best feature subset, a wrapper feature selection algorithm optimizes a given objective function. The construction of an objective function for feature selection varies according to the classification problem. Earlier, an objective function was formulated which contains either maximization of classification accuracy, or the minimization of number of selected features. Besides, to combine both conflicting objectives, the multi-objective function was constructed in solving the feature selection problem. The problem with the multi-objective function was converted into a single objective by applying weights to both the objectives and performed the learning algorithm. Many researchers [170], [185], [186], [191], [203], [235] have used multi-objective function and obtained the best feature subset.

Moreover, the use of multi-objective function was very effective and efficient to optimize the fitness function and find the best feature subset from the given datasets of features.

D. EVALUATION CRITERIA TO CHECK THE PERFORMANCE

In literature, there are lots of evaluation metrics that have been used to investigate the performance of the wrapper feature selection algorithm. For example, sensitivity and specificity commonly used for medical classification, precision and recall considered in computer science data classification, the area under the curve used in radar signals. In general, there are some other measures used to evaluate the performance of the algorithms. The most popular evaluation metrics are reviewed and presented in detail.

(i) True positive (tP): The actual observations are from positive class and are estimated to be positive.

(ii) True negative (tN): The actual observations are from negative class and are estimated by the model to be negative.

(iii) False positive (fP): When the model incorrectly estimates observations in the positive class.

(iv) False negative (fN): When the model incorrectly estimates the observations in the negative class.
These above measures are used in [254], [255].

(v) Sensitivity/ True positive rate/ Recall [144], [256], [257]: It is the ratio of observations that are true positive and the total number of observations that are actually positive i.e.

$$Recall = \frac{tP}{tP + fN} \tag{1}$$

(vi) Specificity / True Negative rate (TNR) [144], [257], [258]: It is the ration of the observations that are true negative and the total number of observation that are actually negative i.e.

$$Specificity = \frac{tN}{tN + fP} \tag{2}$$

(vii) False positive rate (FPR) [255]: Ratio of false positive observations and total predicted negative observations i.e.

$$FPR = \frac{fP}{tN + fN} \tag{3}$$

It can also be defined as $FPR = 1 - Specificity$.

(viii) Precision / Positive predictive value [256], [258]: Ratio of true positive observations and total predicted positive observations. i.e.

$$Precision = \frac{tP}{tP + fP} \tag{4}$$

(ix) F-score [257], [259]: It is a combination of Recall and precision measures which provides a single score. F-score is defined as a Harmonic mean of Recall and Precision measures that can be formulated as

$$F - Score = 2\frac{Precision * Recall}{Precision + Recall} \tag{5}$$

(x) Matthew's correlation coefficients (MCC) [260], [261]: It describes the quality of binary classification and mostly used in the field of bio-informatics. Based on the above measures value, it can be formulated as

$$MCC = \frac{(tP \times tN) - (fP \times fN)}{\sqrt{(tP + fP)(tP + fN)(tN + fP)(tN + fN)}} \tag{6}$$

The value of MCC lies between -1 and 1, the value 1 represents the perfect classification, 0 denotes for random prediction, and -1 represents the total disagreement of prediction and observations.

(xi) Some general measure metrics are also used for checking the performance such as average fitness value (objective function value), Worst and best fitness value, standard deviation of fitness values, average number of selected features from the original datasets. These performance measures are used in [144], [262]–[269] to evaluate the performance.

- Averge fitness value: Assume F_i^* be the optimal fitness value at i^{th} run, then the average fitness value represents the mean value of the fitness over total number of runs (T_{runs}). It can be formulated mathematically as

$$Avg_{Fitness} = \frac{1}{T_{runs}} \sum_{i=1}^{T_{runs}} F_i^* \tag{7}$$

- Average number of selected features: The selection size of features is the ratio of number of selected feauters and the total number of features in the original datast. Mathematically, it can be represented as

$$Avg_{Feature} = \frac{1}{T_{runs}} \sum_{i=1}^{T_{runs}} \frac{length(f)_i^*}{|S|} \tag{8}$$

- Average computational time: Computational time presents the running time to perform the k^{th} algorithm. The average of the running time is the mean value of the time over total number of runs. It is presented for the K^{th} algorithms as

$$Avg_{Time}^k = \frac{1}{T_{runs}} \sum_{i=1}^{T_{runs}} Time_i^k \tag{9}$$

V. CASE STUDY

This section presents applications of feature selection in different datasets of machine learning. It describes the performance of metaheuristic algorithms on feature selection datasets.

A. DATASETS

To show the performance of metaheuristic algorithms on feature selection datasets, eight datasets are adopted from the UCI repository [134]. The datasets are of different dimensions (number of features in a dataset). The datasets are considered in which dimensions vary from 9 to 856 and number of instances are from 32 to 1593. As KNN is the most preferred classifier, therefore, we consider KNN classifier to evaluate the accuracy of selected feature subset. The three equal portion of a dataset is taken for training, testing and validation in a cross validation manner with $K = 5$. The description of datasets is shown in Table 5.

TABLE 5. Datasets for the case study.

Datasets	Description	Dimension	No. of instances
D_1	Tic-Tac-Toe	9	958
D_2	House Vote	16	435
D_3	Hepatitits	19	155
D_4	Inosphere	34	351
D_5	Lung Cancer	56	32
D_6	Hillvalley	100	606
D_7	Semeion	265	1593
D_8	CNAE	856	1080

B. RESULTS AND DISCUSSION

To evaluate the performance of metaheuristic algorithms on the considered datasets, we considered five metaheuristic algorithms i.e. binary particle swarm optimization (BPSO) [270], binary differential evolution (BDE) [271], binary ant lion optimizer (BALO) [183], binary grey wolf optimizer (bGWO) [170] and binary gaining sharing knowledge based algorithm (FS-NBGSK) [235]. The detailed methodologies of the considered algorithms can be read from the corresponding references. The main aim of feature selection problem is to maximize the accuracy of the performance and minimize the number of selected features. Therefore, a fitness function is formed using the both criteria as

$$\min Z = \xi_1(\text{Accuracy}) + \xi_2 \frac{\text{Number of selected features}}{\text{Total number of features}} \tag{10}$$

where ξ_1 and ξ_2 are the coefficient of each criteria in which $\xi_2 = 1 - \xi_1$ and $\xi_1 \in [0, 1]$. The value of ξ_1 is taken as 0.99 [183], [272]. The values of the parameters used in the algorithms are as: In BDE, crossover probability = 0.95, c_1 cognitive factor = 2, c_2 social factor = 2; In BPSO, W_{max} maximum bound on inertia weight = 0.6, W_{min} minimum bound on inertia weight = 0.2; In FS-NBGSK, p probability = 0.1, k_r knowledge ratio = 0.95. To obtain the optimal feature subset, the algorithms are run on the same platform for which the following assumptions are made:

$$\text{Number of population size} = \begin{cases} 50, & \text{if Dimension} < 20 \\ 1000, & \text{if Dimension} \geq 20 \end{cases}$$

$$\text{Number of function evaluations} = \begin{cases} 5000, & \text{if Dimension} < 20 \\ 20000, & \text{if Dimension} \geq 20 \end{cases}$$

$$\text{Number of runs} = \begin{cases} 25, & \text{if Dimension} < 20 \\ 10, & \text{if Dimension} \geq 20 \end{cases}$$

The four evaluation metrices are adopted to compare the performance of the algorithms i.e. average fitness value, average classification accuracy, average number of selected features and average computational time. The obtained results are presented in Table 6-9. Table 6 presents the average fitness values of all algorithms for all datasets. Table 6 shows that FS-NBGSK algorithm obtains minimum average fitness values in most of the datasets. Specially, for large dimensional datasets i.e. D_8, it shows the commandable results.

TABLE 6. Average fitness value of all algorithms.

Datasets	BPSO	BDE	BALO	bGWO	FS-NBGSK
D_1	0.2067	0.2098	0.2067	0.2082	0.1974
D_2	0.0476	0.0570	0.0442	0.0624	0.0440
D_3	0.2526	0.2938	0.2429	0.2953	0.2400
D_4	0.1098	0.1330	0.0864	0.1281	0.0762
D_5	0.1140	0.1154	0.1124	0.1215	0.0813
D_6	0.4405	0.4409	0.3992	0.4415	0.3861
D_7	0.0271	0.0279	0.0200	0.0257	0.0134
D_8	0.0083	0.0092	0.0064	0.0086	0.0046

TABLE 7. Average accuracy of all algorithms.

Datasets	BPSO	BDE	BALO	bGWO	FS-NBGSK
D_1	0.7954	0.7968	0.7989	0.7978	0.8082
D_2	0.9530	0.9448	0.9567	0.9396	0.9563
D_3	0.7477	0.7082	0.7574	0.7062	0.7610
D_4	0.8920	0.8722	0.9148	0.8767	0.9250
D_5	0.8875	0.8875	0.8875	0.8813	0.9188
D_6	0.5597	0.5614	0.6007	0.5604	0.6139
D_7	0.9773	0.9793	0.9843	0.9812	0.9905
D_8	0.9957	0.9954	0.9965	0.9956	0.9972

TABLE 8. Average number of selected features from all algorithms.

Datasets	BPSO	BDE	BALO	bGWO	FS-NBGSK
D_1	0.76	0.86	0.76	0.81	0.75
D_2	0.11	0.23	0.13	0.26	0.08
D_3	0.28	0.49	0.27	0.44	0.34
D_4	0.30	0.64	0.20	0.60	0.19
D_5	0.13	0.40	0.11	0.39	0.09
D_6	0.46	0.67	0.39	0.63	0.38
D_7	0.47	0.74	0.44	0.71	0.39
D_8	0.41	0.46	0.29	0.42	0.18

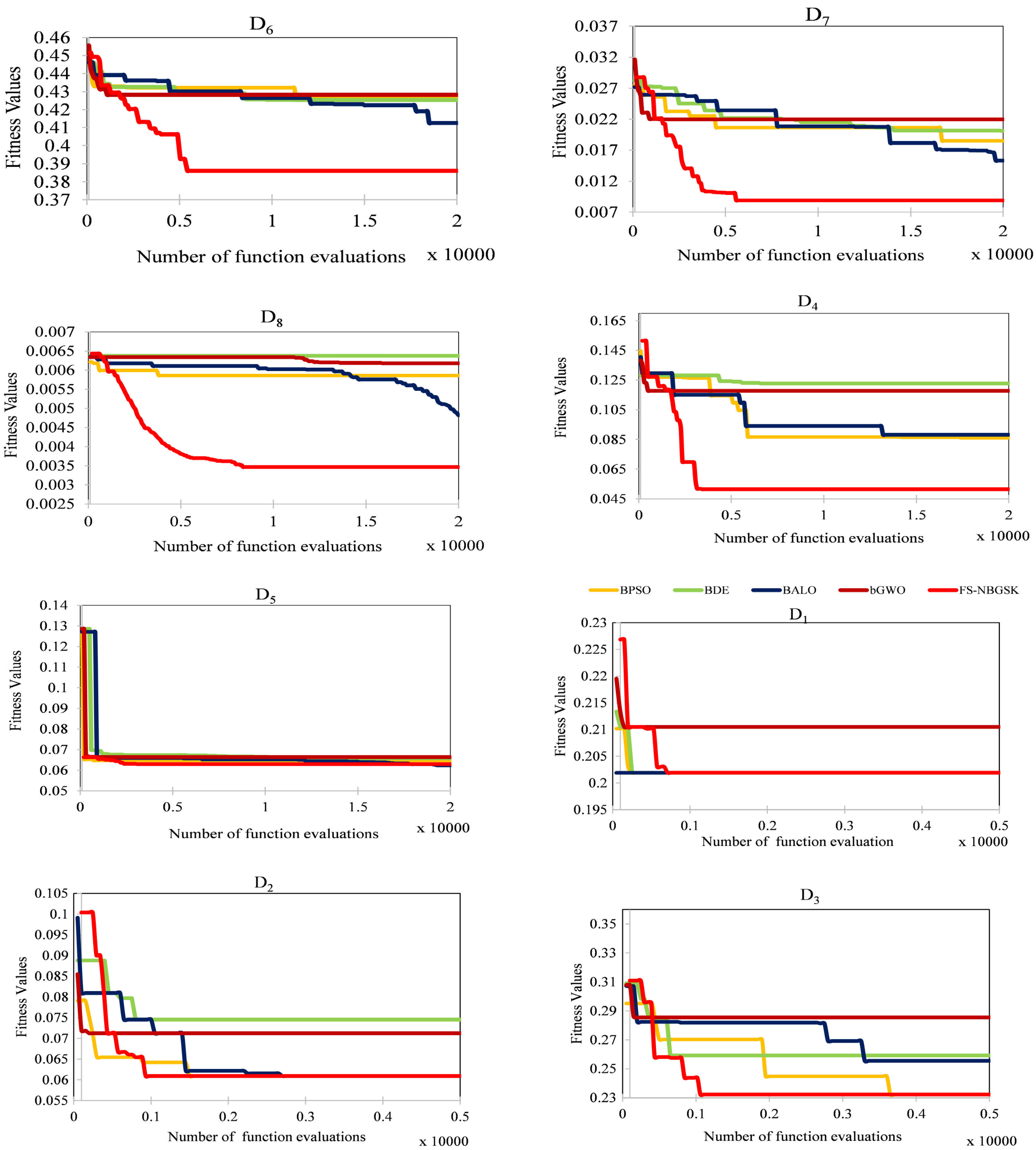

FIGURE 6. Convergence graph of all algorithms.

TABLE 9. Average computational time taken by all algorithms.

Datasets	BPSO	BDE	BALO	bGWO	FS-NBGSK
D_1	42.61	48.71	43.83	47.50	44.92
D_2	25.51	28.63	26.13	26.13	27.98
D_3	27.73	27.53	27.51	28.58	27.04
D_4	27.75	27.88	25.60	29.48	24.43
D_5	41.67	39.40	39.32	44.73	38.92
D_6	201.24	231.08	189.48	240.18	178.32
D_7	2263.87	3844.61	2687.08	5094.56	1767.84
D_8	3425.73	4704.51	3891.63	3598.61	1778.36

The average accuracy of the algorithm is one of the most important metrices which describes that how much accurate an algorithm performs over the selected features. The values of accuracy lies from 0 to 1. Table 7 presents the average accuracy of all the considered algorithms. The second main objective is to minimize the number of selected features with maximum accuracy. Thus, the average number of selected features are described in Table 8. It shows that FS-NBGSK algorithm performs better than other algorithms. To do the fair comparison, computational time is also considered as an evaluation metric. The average computational time (in sec) taken by all algorithms is given in Table 9. From the Table, it can be observed that FS-NBGSK algorithm takes less computational time as compared to others.

Moreover, convergence graph of all algorithms are also drawn in Figure 6 for fitness values over the number of

TABLE 10. Details of physics based algorithm used for feature selection.

Algorithm Name	Used transfer function	Name of classifier	Datasets	Evaluation Metrics	Compared with	Other Information
OPF-GSA [273]	$S(x) = \|\tanh x\|$	Optimum path forest (OPF)	5 datasets of different categories i.e. speaker recognition, Brazilian electric power company etc.		OPF, OPF-PSO, OPF-LDA, OPF-PCA	
GSA-BRANN [262]	$S(x) = \|\tanh x\|$	Bayesian regularization artificial neural network (BRANN)	Imidazo[4,5-b]pyridine derivatives	Avg of the mean and best fitness, the median of best and mean fitness, the standard deviation of best and mean fitness	GA	BRANN used to build a model for quantitative structure-activity relationship and anticancer potency
GSA [263]	$S(x) = \|\tanh x\|$	Rule based classifier	Electroencephalography (EEG) signal peak detection	Avg, maximum, minimum and std of fitness function	PSO and random asynchronous PSO	Parameter estimation of the classifier and selection of peak feature were done simultaneously.
FSS-MGSA [259]	$S(x) = \|\tanh x\|$	NB, SVM, ID3, KNN	15 well known UCI datasets	accuracy of classifier	BPSO, GA	Piecewise linear chaotic maps were utilized for good exploration of search space and sequential quadratic programming were used for exploitation
MI-BGSA [254]	$S(x) = \|\tanh x\|$	SVM	NSL-KDD used for intrusion detection system	True positive rate, False positive rate, accuracy, number of selected features	BGSA, BPSO	Mutual information (MI) was embedded into BGSA to extract the features
BQIGSA [256]	$S(x) = x^2$	K-NN	12 well-known datasets	precision, recall and rate of feature reduction	QBGSA, BQIPSO, IBGSA, ACOFS, ACO_H, GA, BPSO, catfishBPSO	Quantam inspired gravitational search algorithm was adopted to improve the performance of BGSA
GSA [274]	S and V-shaped transfer functions	fuzzy rule-based classifier	26 Knowledge extraction based on evolutionary learning datasets	Accuracy	non-fuzzy classifiers such as logistic regression, K-NN, SVM, Random forest etc	fuzzy rule-based classifier were formed using minimum and maximum values of features
HGSA [264]	$S(x) = \|\tanh x\|$	K-NN and decision tree classifiers	18 classical datasets	Avg fitness value, avg classification accuracy, avg number of selected features	GWO, PSO, GA, BGSA	Include evolutionary mutation and crossover operator in GSA
RBGSA [275]	$S(x) = \|\tanh x\|$	ID3, NB, K-NN, SVM	11 microarray datasets for cancer classification	classification accuracy, avg number of gene selected, computational time, Bonferroni correction p values	ReliefF-BGSA	ReliefF used to chose a relevant subset of genes and then optimization process was strated
CPBGSA [265]	$S(x) = \tanh x$	KNN	20 well known UCI datasets	Best, worst, avg, std of fitness values and accuracy	ALO, BALO, BBA, GA, PSO, GSA	For distribution of the initial population over the entire space, clustering-based population BGSA was proposed
BCSS [276]	$S(x) = \frac{1}{1+\exp^{-x}}$	Optimum path forest	4 datasets in which 3 was considered for diabites and 1 for NTL	Classification accuracy	BBA, BGSA, BHS, BPSO	NTL dataset was from Brazilian electric power for theft detection

TABLE 11. Details of physics based algorithm used for feature selection.

BBHA+C4.5 [260]	$S(x) = \|\tanh x\|$	C4.5 decision tree	2 microarray dataset and 5 different types of medical datasets	Matthew's correlation coefficient and accuracy	LDA, SVM, CART, C4.5, C5.0, SOM, NB	Used C4.5 classification method as a fitness function
BHA [266]	$S(x) = \frac{1}{1+\exp^{-x}}$	Optimum path forest	Two private commericial and industrial datasets for theft characterization	the average number of selected features, Wilcoxon statistical test	HS, DE, PSO	
BBHA [261]	$S(x) = \|\tanh x\|$	Random forest	8 publicaly available biomedical datasets	Accuracy, Matthew's correlation coefficient, area under the curve (AUC)	Bagging, C5.0, C4.5, Boosted C4.5, CART	Used classifier as a fitness function
CBBHA [277]	$S(x) = \|\tanh x\|$	K-NN	three chemical datasets that contains thousand of descriptors	Selected number of features and accuracy	BBHA	applied ten different chaotic maps to random parameters of BHA
CMVO [257]	Only use binary representation by applying threshold value	K-NN	5 classical datasets	Sensitivity, specificity, F-measure, accuracy, precision, negative predictive value	MVO, PSO, ABC	5 chaotic maps are applied to improve the performance of MVO
EMVO [278]	Used threshold value	SVM	12 classical microarrary datasets	F-measure, Avg number of selected genes	MVO	EMVO was proposed to enhance the exploitation quality of MVO
CMVO-SVM [255]	Use binary representation by applying threshold value	SVM	KDDCUP99 dataset for network intrusion detection	False negative rate, false alarm rate and detection time	PSO-SVM, GA-SVM, MVO-SVM	Principal component analysis was applied for dimension reduction and feature extraction from dataset
BMVO [267]	S and V-shaped transfer functions	K-NN	21 different types of datasets	Avg accuracy, avg number of selected features, avg fitness, best and worst fitness values, std of fitness, F-measure, avg execution time	GWO, WOA, SCA, PSO, ALO, bMVO	Used two types of S and V-shaped transfer function
SCA [279]	Used threshold value of 0.5		18 classical datasets	Best and worst fitness value, F-score, avg computational time	GA,PSO	Used both classification and feature subset size reduction
ISCA [268]	Used threshold value of 0.5	ELM with radial basis function kernel	10 well known datasets for pattern recognition	Number of selected features and classification accuracy	GA, PSO, SCA	Improved SCA by proposing elitism strategy and updation of new best solution
bSCA [269]	S and V-shaped transfer functions	K-NN	5 medical datasets	Avg classification accuracy and number of selected features	BBA, BGSA, bGWO, BDA	Named two algorithm as SBSCA and VBSCA
ISCA [258]	Used a threshold value of 0.5	Naive Bayesian	9 datasets for text classification	Avg of precision, recall and F-measure	MFO, SCA, ACO, GA	two modifications have been made for good exploration

function evaluations. The figure shows the convergence ability of the algorithms towards the optimal solution. FS-NBGSK algorithm converges to the minimum fitness value than the other algorithms. All other algorithms have premature convergence or trap into local optima.

VI. FUTURE WORKS

Based on the presented literature of metaheuristic algorithms and their binary versions for feature selection, the following research gaps are found. Interested researchers who are willing to pursue their research in this field may consider these findings:

(1) Firstly, there are very less number of evolution based and human behaviour related algorithms developed. New algorithms can be developed based on natural evolution and human activity.

(2) No research has been done to propose the binary version of the following algorithms:
From swarm-based algorithms-bumblebees algorithm [43], paddy field algorithm [44], eagle strategy [51], Hierarchical swarm optimization [52], bird mating optimizer [58], Japanese tree frogs calling algorithm [61], the great salmon run algorithm [64], Egyptian vulture optimization algorithm [67], animal migration optimization [69], shark smell optimization [70], spotted hyena optimizer [84], emperor penguins colony [88].
From physics based algorithms- galaxy based search algorithm [91], curved space optimization [97], ray optimization [98], lightning search algorithm [105], thermal exchange optimization [111], find fix finish exploit analyze [112].
From human related algorithm- league championship algorithm [38], human inspired algorithm [113], social emotional optimization [114].

(3) The algorithms mentioned above have not been applied to solve the feature selection problem. Therefore, binary variants of these algorithms can be implemented in classification problems.

(4) In addition to developing binary variants of metaheuristic algorithms, many researchers have used well known S and V-shaped transfer functions. New S and V-shaped transfer functions can be formulated and used for designing the binary version of metaheuristic algorithms.

(5) There is some less explored area(s) such as stock market prediction, short term load forecasting, weather prediction, spam detection, Parkinson disease. These problems can be investigated further using metaheuristic algorithms.

(6) In the literature mostly two objectives i.e. maximizing the accuracy and minimizing the number of selected features are considered. Besides these objectives, interested researchers can consider computational time, complexity, stability and scalability in multi-objective in feature selection.

VII. CONCLUSION

This paper presents a comprehensive survey on metaheuristic algorithms that are developed from 2009 to the 2019 year and their binary variants, which have been applied to feature selection problem. A detailed description and mathematical model of feature selection problem are given that could help researchers to understand the problem properly. Moreover, the techniques of solving feature selection problems are presented. Additionally, metaheuristic algorithms are considered in solving the feature selection problem. Therefore, basic definition, importance and the classification of metaheuristic algorithms are given. The evolution-based, swarm-based, physics-based category, human rlated algorithms have been developed and applied to feature selection problems. However, metaheuristic algorithms have some following drawbacks:

- They suffer from slow convergence rate due to random generation movement.
- They explore the search space without knowing the search direction.
- They can trap into local optima, or they have some premature convergence.
- The values of the parameters used in the metaheuristic algorithms have to be adjusted, this may also lead to pre-mature convergence.

Besides, the limitation of the metaheuristic algorithms, the modified and enhanced version of the algorithms were developed which are successfully applied to the feature selection problems. Also, a categorization is presented based on the behaviour of algorithms; evolution-based, swarm-based, physics-related and human behaviour related algorithms. This paper benefits in such a way that a list of metaheuristic algorithms is presented based on their classification. It also benefits for the application point of view as it consists of a case study. The case study presents the eight benchmark datasets and the optimal feature subsets are found by implementing different metaheuristic algorithms.

It can be observed that very few algorithms are proposed in the evolution and human-related category, but there are several algorithms have been designed in the swarm and physics-related algorithms. It implies that there is a scope to develop or propose new metaheuristic algorithms in these categories. This paper mainly focuses on solving the feature selection problem using binary variants of metaheuristic algorithms. Hence, extensive literature is presented in every class of metaheuristic algorithms. All binary variants of all reviewed algorithms regarding feature selection problems are pointed. In swarm-based category, all binary variants of Cuckoo search, Bat algorithm, Firefly algorithm, flower pollination algorithm, Krill herd algorithm, Grey wolf optimizer, Ant lion optimizer, Dragonfly algorithm, Whale optimization algorithm, Grasshopper optimization algorithm, Salp swarm algorithm are reviewed with the key factor of solving feature selection problem. Moreover, hybrid approaches are also reviewed in the process of solving the feature selection problem.

It can be concluded that there is some area(s) which are less explored, such as spam detection, theft detection and weather prediction. However, lots of research has been done on the well-known datasets of UCI repository and in medical diagnosis (cancer classification), intrusion detection systems, text classification, multimedia etc. Hence, researchers should pay great attention to explore this area with metaheuristic algorithms. Moreover, there are some algorithms in the literature for which binary variants are not developed yet such as PFA, CGS, TCO, ES, HSO, WSA, BMO, OptBees, TGSR, EVOA, VCS, EPC, GbSA, CSO, WEO, LCA, EMA, VPL. These algorithms benefit classification after developing their binary version. From the literature, it can be observed that the researcher has to face many challenges to obtain the best feature subset of the considered classification problem. A good choice of classifier has a significant impact of the quality of obtained solution such KNN classifier is the most used classifier in getting the best subset with well-known datasets of UCI repository. After that, SVM classifier used to classify in different applications such as medical diagnosis, pattern recognition, image analysis etc. There are some other classifiers which are less used in terms of classification. Hence, this another gap to use different classifiers in classification problem and compared with most used ones. Finally, researchers will get the benefit of this study as they could find all the key factors in solving the feature selection problem using metaheuristic algorithms under one roof.

ACKNOWLEDGMENT

The authors would like to acknowledge the Editors and anonymous reviewers for providing their valuable comments and suggestions.

REFERENCES

[1] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.

[2] P. Y. Lee, W. P. Loh, and J. F. Chin, "Feature selection in multimedia: The state-of-the-art review," *Image Vis. Comput.*, vol. 67, pp. 29–42, Nov. 2017.

[3] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Comput. Biol. Med.*, vol. 112, Sep. 2019, Art. no. 103375.

[4] M. Sharma and P. Kaur, "A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem," *Arch. Comput. Methods Eng.*, pp. 1–25, Feb. 2020, doi: 10.1007/s11831-020-09412-6.

[5] M. Z. Asghar, A. Khan, S. Ahmad, and F. M. Kundi, "A review of feature extraction in sentiment analysis," *J. Basic Appl. Sci. Res.*, vol. 4, no. 3, pp. 181–186, 2014.

[6] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Oct. 2007.

[7] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," *Inf. Fusion*, vol. 52, pp. 1–12, Dec. 2019.

[8] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.

[9] S. Ahmed, M. Zhang, and L. Peng, "Enhanced feature selection for biomarker discovery in LC-MS data using GP," in *Proc. IEEE Congr. Evol. Comput.*, Jun. 2013, pp. 584–591.

[10] M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri, "Text feature selection using ant colony optimization," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6843–6853, Apr. 2009.

[11] A. Ghosh, A. Datta, and S. Ghosh, "Self-adaptive differential evolution for feature selection in hyperspectral image data," *Appl. Soft Comput.*, vol. 13, no. 4, pp. 1969–1977, Apr. 2013.

[12] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, nos. 1–4, pp. 131–156, 1997.

[13] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[14] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature selection: An ever evolving frontier in data mining," in *Feature Selection in Data Mining*. Hyderabad, India, 2010, pp. 4–13.

[15] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, "MIFS-ND: A mutual information-based feature selection method," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6371–6385, Oct. 2014.

[16] Z. Xu, I. King, M. R.-T. Lyu, and R. Jin, "Discriminative semi-supervised feature selection via manifold regularization," *IEEE Trans. Neural Netw.*, vol. 21, no. 7, pp. 1033–1047, Jul. 2010.

[17] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," in *Data Classification: Algorithms and Applications*. 2014, p. 37.

[18] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *Proc. 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, 2015, pp. 1200–1205.

[19] Z. Sun, G. Bebis, and R. Miller, "Object detection using feature subset selection," *Pattern Recognit.*, vol. 37, no. 11, pp. 2165–2176, Nov. 2004.

[20] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[21] H. Liu and H. Motoda, *Feature Extraction, Construction Selection: A Data Mining Perspective*, vol. 453. USA: Springer, 1998.

[22] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[23] O. Olorunda and A. P. Engelbrecht, "Measuring exploration/exploitation in particle swarms using swarm diversity," in *Proc. IEEE Congr. Evol. Comput. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1128–1134.

[24] L. Lin and M. Gen, "Auto-tuning strategy for evolutionary algorithms: Balancing between exploration and exploitation," *Soft Comput.*, vol. 13, no. 2, pp. 157–168, Jan. 2009.

[25] A. W. Mohamed, A. A. Hadi, and A. K. Mohamed, "Gaining-sharing knowledge based algorithm for solving optimization problems: A novel nature-inspired algorithm," *Int. J. Mach. Learn. Cybern.*, vol. 11, pp. 1501–1529, 2020.

[26] J. H. Holland, "Genetic algorithms," *Sci. Amer.*, vol. 267, no. 1, pp. 66–73, 1992.

[27] I. Rechenberg, "Evolutionsstrategien," in *Simulationsmethoden der Medizin und Biologie*. Berlin, Germany: Springer, 1978, pp. 83–114.

[28] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.

[29] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Comput. Oper. Res.*, vol. 13, no. 5, pp. 533–549, Jan. 1986.

[30] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[31] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw. (ICNN)*, vol. 4, Nov./Dec. 1995, pp. 1942–1948.

[32] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 1, pp. 29–41, Feb. 1996.

[33] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Dept. Comput. Eng., Erciyes Univ., Kayseri, Turkey, Tech. Rep. 06, 2005.

[34] R.-Q. Zhao and W.-S. Tang, "Monkey algorithm for global numerical optimization," *J. Uncertain Syst.*, vol. 2, no. 3, pp. 165–176, 2008.

[35] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, ''Optimization by simulated annealing,'' *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[36] Z. Woo Geem, J. Hoon Kim, and G. V. Loganathan, ''A new heuristic optimization algorithm: Harmony search,'' *Simulation*, vol. 76, no. 2, pp. 60–68, Feb. 2001.

[37] R. V. Rao, V. J. Savsani, and D. P. Vakharia, ''Teaching–learning-based optimization: An optimization method for continuous non-linear large scale problems,'' *Inf. Sci.*, vol. 183, no. 1, pp. 1–15, Jan. 2012.

[38] A. H. Kashan, ''League championship algorithm: A new algorithm for numerical function optimization,'' in *Proc. Int. Conf. Soft Comput. Pattern Recognit.*, 2009, pp. 43–48.

[39] P. Civicioglu, ''Transforming geocentric Cartesian coordinates to geodetic coordinates by using differential search algorithm,'' *Comput. Geosci.*, vol. 46, pp. 229–247, Sep. 2012.

[40] P. Civicioglu, ''Backtracking search optimization algorithm for numerical optimization problems,'' *Appl. Math. Comput.*, vol. 219, no. 15, pp. 8121–8144, Apr. 2013.

[41] H. Salimi, ''Stochastic fractal search: A powerful metaheuristic algorithm,'' *Knowl.-Based Syst.*, vol. 75, pp. 1–18, Feb. 2015.

[42] T. T. Dhivyaprabha, P. Subashini, and M. Krishnaveni, ''Synergistic fibroblast optimization: A novel nature-inspired computing algorithm,'' *Frontiers Inf. Technol. Electron. Eng.*, vol. 19, no. 7, pp. 815–833, Jul. 2018.

[43] F. Comellas and J. Martinez-Navarro, ''Bumblebees: A multiagent combinatorial optimization algorithm inspired by social insect behaviour,'' in *Proc. 1st ACM/SIGEVO Summit Genet. Evol. Comput.*, 2009, pp. 811–814.

[44] U. Premaratne, J. Samarabandu, and T. Sidhu, ''A new biologically inspired optimization algorithm,'' in *Proc. Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2009, pp. 279–284.

[45] X.-S. Yang and S. Deb, ''Cuckoo search via Lévy flights,'' in *Proc. World Congr. Nature Biol. Inspired Comput. (NaBIC)*, 2009, pp. 210–214.

[46] S. He, Q. H. Wu, and J. R. Saunders, ''Group search optimizer: An optimization algorithm inspired by animal searching behavior,'' *IEEE Trans. Evol. Comput.*, vol. 13, no. 5, pp. 973–990, Oct. 2009.

[47] S. Iordache, ''Consultant-guided search: A new Metaheuristic for combinatorial optimization problems,'' in *Proc. 12th Annu. Conf. Genetic Evol. Comput.*, 2010, pp. 225–232.

[48] X.-S. Yang, ''A new metaheuristic bat-inspired algorithm,'' in *Nature Inspired Cooperative Strategies for Optimization*. Berlin, Germany: Springer, 2010, pp. 65–74.

[49] R. Hedayatzadeh, F. Akhavan Salmassi, M. Keshtgari, R. Akbari, and K. Ziarati, ''Termite colony optimization: A novel approach for optimizing continuous problems,'' in *Proc. 18th Iranian Conf. Electr. Eng.*, May 2010, pp. 553–558.

[50] R. Oftadeh, M. J. Mahjoob, and M. Shariatpanahi, ''A novel metaheuristic optimization algorithm inspired by group hunting of animals: Hunting search,'' *Comput. Math. Appl.*, vol. 60, no. 7, pp. 2087–2098, Oct. 2010.

[51] X.-S. Yang and S. Deb, ''Eagle strategy using lévy walk and firefly algorithms for stochastic optimization,'' in *Nature Inspired Cooperative Strategies for Optimization*. Berlin, Germany: Springer, 2010, pp. 101–111.

[52] H. Chen, Y. Zhu, K. Hu, and X. He, ''Hierarchical swarm model: A new approach to optimization,'' *Discrete Dyn. Nature Soc.*, vol. 2010, pp. 1–30, Jan. 2010.

[53] X-S. Yang, ''Firefly algorithm, stochastic test functions and design optimisation,'' *Int. J. Bio-Inspired Comput.*, vol. 2, no. 2, pp. 78–84, 2010.

[54] W.-T. Pan, ''A new fruit fly optimization algorithm: Taking the financial distress model as an example,'' *Knowl.-Based Syst.*, vol. 26, pp. 69–74, Feb. 2012.

[55] R. S. Parpinelli and H. S. Lopes, ''An eco-inspired evolutionary algorithm applied to numerical optimization,'' in *Proc. 3rd World Congr. Nature Biol. Inspired Comput.*, Oct. 2011, pp. 466–471.

[56] T. Ting, K. L. Man, S.-U. Guan, M. Nayel, and K. Wan, ''Weightless swarm algorithm (WSA) for dynamic optimization problems,'' in *Proc. IFIP Int. Conf. Netw. Parallel Comput.* Berlin, Germany: Springer, 2012, pp. 508–515.

[57] X.-S. Yang, ''Flower pollination algorithm for global optimization,'' in *Proc. Int. Conf. Unconv. Comput. Natural Comput.* Berlin, Germany: Springer, 2012, pp. 240–249.

[58] A. Askarzadeh and A. Rezazadeh, ''A new heuristic optimization algorithm for modeling of proton exchange membrane fuel cell: Bird mating optimizer,'' *Int. J. Energy Res.*, vol. 37, no. 10, pp. 1196–1204, Aug. 2013.

[59] P. Civicioglu, ''Artificial cooperative search algorithm for numerical optimization problems,'' *Inf. Sci.*, vol. 229, pp. 58–76, Apr. 2013.

[60] A. H. Gandomi and A. H. Alavi, ''Krill herd: A new bio-inspired optimization algorithm,'' *Commun. Nonlinear Sci. Numer. Simul.*, vol. 17, no. 12, pp. 4831–4845, Dec. 2012.

[61] H. Hernández and C. Blum, ''Distributed graph coloring: An approach based on the calling behavior of Japanese tree frogs,'' *Swarm Intell.*, vol. 6, no. 2, pp. 117–150, Jun. 2012.

[62] R. D. Maia, L. N. de Castro, and W. M. Caminhas, ''Bee colonies as model for multimodal continuous optimization: The OptBees algorithm,'' in *Proc. IEEE Congr. Evol. Comput.*, Jun. 2012, pp. 1–8.

[63] R. Tang, S. Fong, X.-S. Yang, and S. Deb, ''Wolf search algorithm with ephemeral memory,'' in *Proc. 7th Int. Conf. Digit. Inf. Manage. (ICDIM)*, Aug. 2012, pp. 165–172.

[64] A. Mozaffari, A. Fathi, and S. Behzadipour, ''The great salmon run: A novel bio-inspired algorithm for artificial system design and optimisation,'' *Int. J. Bio-Inspired Comput.*, vol. 4, no. 5, pp. 286–301, 2012.

[65] A. Kaveh and N. Farhoudi, ''A new optimization method: Dolphin echolocation,'' *Adv. Eng. Softw.*, vol. 59, pp. 53–70, May 2013.

[66] M. Neshat, G. Sepidnam, and M. Sargolzaei, ''Swallow swarm optimization algorithm: A new method to optimization,'' *Neural Comput. Appl.*, vol. 23, no. 2, pp. 429–454, Aug. 2013.

[67] C. Sur, S. Sharma, and A. Shukla, ''Egyptian vulture optimization algorithm—A new nature inspired meta-heuristics for knapsack problem,'' in *Proc. 9th Int. Conf. Comput. Inf. Technol. (IC2IT)*. Berlin, Germany: Springer, 2013, pp. 227–237.

[68] X. Meng, Y. Liu, X. Gao, and H. Zhang, ''A new bio-inspired algorithm: Chicken swarm optimization,'' in *Proc. Int. Conf. Swarm Intell.* Cham, Switzerland: Springer, 2014, pp. 86–94.

[69] X. Li, J. Zhang, and M. Yin, ''Animal migration optimization: An optimization algorithm inspired by animal migration behavior,'' *Neural Comput. Appl.*, vol. 24, nos. 7–8, pp. 1867–1877, Jun. 2014.

[70] O. Abedinia, N. Amjady, and A. Ghasemi, ''A new metaheuristic algorithm based on shark smell optimization,'' *Complexity*, vol. 21, no. 5, pp. 97–116, May 2016.

[71] S. Mirjalili, ''The ant lion optimizer,'' *Adv. Eng. Softw.*, vol. 83, pp. 80–98, May 2015.

[72] X.-B. Meng, X. Z. Gao, L. Lu, Y. Liu, and H. Zhang, ''A new bio-inspired optimisation algorithm: Bird swarm algorithm,'' *J. Experim. Theor. Artif. Intell.*, vol. 28, no. 4, pp. 673–687, Jul. 2016.

[73] M. D. Li, H. Zhao, X. W. Weng, and T. Han, ''A novel nature-inspired algorithm for optimization: Virus colony search,'' *Adv. Eng. Softw.*, vol. 92, pp. 65–88, Feb. 2016.

[74] S. A. Uymaz, G. Tezel, and E. Yel, ''Artificial algae algorithm (AAA) for nonlinear global optimization,'' *Appl. Soft Comput.*, vol. 31, pp. 153–171, Jun. 2015.

[75] S. Mirjalili, ''Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems,'' *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, May 2016.

[76] W. Yong, W. Tao, Z. Cheng-Zhi, and H. Hua-Juan, ''A new stochastic optimization approach—Dolphin swarm optimization algorithm,'' *Int. J. Comput. Intell. Appl.*, vol. 15, no. 2, 2016, Art. no. 1650011.

[77] A. Askarzadeh, ''A novel Metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm,'' *Comput. Struct.*, vol. 169, pp. 1–12, Jun. 2016.

[78] S. Mirjalili and A. Lewis, ''The whale optimization algorithm,'' *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.

[79] E. Jahani and M. Chizari, ''Tackling global optimization problems with a novel algorithm–mouth brooding fish algorithm,'' *Appl. Soft Comput.*, vol. 62, pp. 987–1002, Jan. 2018.

[80] X. Qi, Y. Zhu, and H. Zhang, ''A new meta-heuristic butterfly-inspired algorithm,'' *J. Comput. Sci.*, vol. 23, pp. 226–239, Nov. 2017.

[81] F. Fausto, E. Cuevas, A. Valdivia, and A. González, ''A global optimization algorithm inspired in the behavior of selfish herds,'' *Biosystems*, vol. 160, pp. 39–55, Oct. 2017.

[82] S. Saremi, S. Mirjalili, and A. Lewis, ''Grasshopper optimisation algorithm: Theory and application,'' *Adv. Eng. Softw.*, vol. 105, pp. 30–47, Mar. 2017.

[83] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, ''Salp swarm algorithm: A bio-inspired optimizer for engineering design problems,'' *Adv. Eng. Softw.*, vol. 114, pp. 163–191, Dec. 2017.

[84] G. Dhiman and V. Kumar, ''Spotted hyena optimizer: A novel bio-inspired based Metaheuristic technique for engineering applications,'' *Adv. Eng. Softw.*, vol. 114, pp. 48–70, Dec. 2017.

[85] G. Dhiman and V. Kumar, ''Emperor penguin optimizer: A bio-inspired algorithm for engineering problems,'' *Knowl.-Based Syst.*, vol. 159, pp. 20–50, Nov. 2018.

[86] M. Jain, V. Singh, and A. Rani, ''A novel nature-inspired algorithm for optimization: Squirrel search algorithm,'' *Swarm Evol. Comput.*, vol. 44, pp. 148–175, Feb. 2019.

[87] S. Arora and S. Singh, ''Butterfly optimization algorithm: A novel approach for global optimization,'' *Soft Comput.*, vol. 23, no. 3, pp. 715–734, Feb. 2019.

[88] S. Harifi, M. Khalilian, J. Mohammadzadeh, and S. Ebrahimnejad, ''Emperor penguins colony: A new Metaheuristic algorithm for optimization,'' *Evol. Intell.*, vol. 12, no. 2, pp. 211–226, Jun. 2019.

[89] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, ''GSA: A gravitational search algorithm,'' *Inf. Sci.*, vol. 179, no. 13, pp. 2232–2248, Jun. 2009.

[90] A. Kaveh and S. Talatahari, ''A novel heuristic optimization method: Charged system search,'' *Acta Mechanica*, vol. 213, nos. 3–4, pp. 267–289, Sep. 2010.

[91] H. Shah-Osseini, ''Principal components analysis by the galaxy-based search algorithm: A novel metaheuristic for continuous optimisation,'' *Int. J. Comput. Sci. Eng.*, vol. 6, nos. 1–2, pp. 132–140, 2011.

[92] E. Cuevas, D. Oliva, D. Zaldivar, M. Pérez-Cisneros, and H. Sossa, ''Circle detection using electro-magnetism optimization,'' *Inf. Sci.*, vol. 182, no. 1, pp. 40–55, Jan. 2012.

[93] B. Alatas, ''ACROA: Artificial chemical reaction optimization algorithm for global optimization,'' *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13170–13180, Sep. 2011.

[94] K. Tamura and K. Yasuda, ''Spiral optimization—A new multipoint search method,'' in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2011, pp. 1759–1764.

[95] A. Hatamlou, ''Black hole: A new heuristic optimization approach for data clustering,'' *Inf. Sci.*, vol. 222, pp. 175–184, Feb. 2013.

[96] H. Eskandar, A. Sadollah, A. Bahreininejad, and M. Hamdi, ''Water cycle algorithm—A novel Metaheuristic optimization method for solving constrained engineering optimization problems,'' *Comput. Struct.*, vols. 110–111, pp. 151–166, Nov. 2012.

[97] F. Farrahi Moghaddam, R. Farrahi Moghaddam, and M. Cheriet, ''Curved space optimization: A random search based on general relativity theory,'' 2012, *arXiv:1208.2214*. [Online]. Available: http://arxiv.org/abs/1208.2214

[98] A. Kaveh and M. Khayatazad, ''A new meta-heuristic method: Ray optimization,'' *Comput. Struct.*, vols. 112–113, pp. 283–294, Dec. 2012.

[99] A. Sadollah, A. Bahreininejad, H. Eskandar, and M. Hamdi, ''Mine blast algorithm: A new population based algorithm for solving constrained engineering optimization problems,'' *Appl. Soft Comput.*, vol. 13, no. 5, pp. 2592–2612, May 2013.

[100] M. Abdechiri, M. R. Meybodi, and H. Bahrami, ''Gases Brownian motion optimization: An algorithm for optimization (GBMO),'' *Appl. Soft Comput.*, vol. 13, no. 5, pp. 2932–2946, May 2013.

[101] G.-W. Yan and Z.-J. Hao, ''A novel optimization algorithm based on atmosphere clouds model,'' *Int. J. Comput. Intell. Appl.*, vol. 12, no. 1, Mar. 2013, Art. no. 1350002.

[102] S. Moein and R. Logeswaran, ''KGMO: A swarm optimization algorithm based on the kinetic energy of gas molecules,'' *Inf. Sci.*, vol. 275, pp. 127–144, Aug. 2014.

[103] A. Kaveh and V. R. Mahdavi, ''Colliding bodies optimization: A novel meta-heuristic method,'' *Comput. Struct.*, vol. 139, pp. 18–27, Jul. 2014.

[104] A. Baykasoğlu and Ş. Akpinar, ''Weighted superposition attraction (WSA): A swarm intelligence algorithm for optimization problems—Part 1: Unconstrained optimization,'' *Appl. Soft Comput.*, vol. 56, pp. 520–540, Jul. 2017.

[105] H. Shareef, A. A. Ibrahim, and A. H. Mutlag, ''Lightning search algorithm,'' *Appl. Soft Comput.*, vol. 36, pp. 315–333, Nov. 2015.

[106] S. Mirjalili, ''SCA: A sine cosine algorithm for solving optimization problems,'' *Knowl.-Based Syst.*, vol. 96, pp. 120–133, Mar. 2016.

[107] A. Kaveh and T. Bakhshpoori, ''Water evaporation optimization: A novel physically inspired optimization algorithm,'' *Comput. Struct.*, vol. 167, pp. 69–85, Apr. 2016.

[108] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, ''Multi-verse optimizer: A nature-inspired algorithm for global optimization,'' *Neural Comput. Appl.*, vol. 27, no. 2, pp. 495–513, Feb. 2016.

[109] A. F. Nematollahi, A. Rahiminejad, and B. Vahidi, ''A novel physical based meta-heuristic optimization method known as lightning attachment procedure optimization,'' *Appl. Soft Comput.*, vol. 59, pp. 596–621, Oct. 2017.

[110] A. Tabari and A. Ahmad, ''A new optimization method: Electro-search algorithm,'' *Comput. Chem. Eng.*, vol. 103, pp. 1–11, Aug. 2017.

[111] A. Kaveh and A. Dadras, ''A novel meta-heuristic optimization algorithm: Thermal exchange optimization,'' *Adv. Eng. Softw.*, vol. 110, pp. 69–84, Aug. 2017.

[112] A. Husseinzadeh Kashan, R. Tavakkoli-Moghaddam, and M. Gen, ''Find-fix-finish-exploit-analyze (F3EA) meta-heuristic algorithm: An effective algorithm with new evolutionary operators for global optimization,'' *Comput. Ind. Eng.*, vol. 128, pp. 192–218, Feb. 2019.

[113] L. M. Zhang, C. Dahlmann, and Y. Zhang, ''Human-inspired algorithms for continuous function optimization,'' in *Proc. IEEE Int. Conf. Intell. Comput. Intell. Syst.*, Nov. 2009, pp. 318–321.

[114] Y. Xu, Z. Cui, and J. Zeng, ''Social emotional optimization algorithm for nonlinear constrained optimization problems,'' in *Proc. Int. Conf. Swarm, Evol., Memetic Comput.* Berlin, Germany: Springer, 2010, pp. 583–590.

[115] Y. Shi, ''Brain storm optimization algorithm,'' in *Proc. Int. Conf. Swarm Intell.* Berlin, Germany: Springer, 2011, pp. 303–309.

[116] A. Ahmadi-Javid, ''Anarchic society optimization: A human-inspired method,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2011, pp. 2586–2592.

[117] N. Moosavian and B. Kasaee Roodsari, ''Soccer league competition algorithm: A novel meta-heuristic algorithm for optimal design of water distribution networks,'' *Swarm Evol. Comput.*, vol. 17, pp. 14–24, Aug. 2014.

[118] F. Ramezani and S. Lotfi, ''Social-based algorithm (SBA),'' *Appl. Soft Comput.*, vol. 13, no. 5, pp. 2837–2856, May 2013.

[119] N. Ghorbani and E. Babaei, ''Exchange market algorithm,'' *Appl. Soft Comput.*, vol. 19, pp. 177–187, Jun. 2014.

[120] M. A. Eita and M. M. Fahmy, ''Group counseling optimization,'' *Appl. Soft Comput.*, vol. 22, pp. 585–604, Sep. 2014.

[121] R. V. Rao, ''Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems,'' *Int. J. Ind. Eng. Computations*, vol. 7, no. 1, pp. 19–34, 2016.

[122] R. Moghdani and K. Salimifard, ''Volleyball premier league algorithm,'' *Appl. Soft Comput.*, vol. 64, pp. 161–185, Mar. 2018.

[123] L. Gan and H. Duan, ''Biological image processing via chaotic differential search and lateral inhibition,'' *Optik*, vol. 125, no. 9, pp. 2070–2075, May 2014.

[124] M. Negahbani, S. Joulazadeh, H. R. Marateb, and M. Mansourian, ''Coronary artery disease diagnosis using supervised fuzzy C-means with differential search algorithm-based generalized Minkowski metrics,'' *Arch. Biomed. Sci. Eng.*, vol. 1, pp. 6–14, Jun. 2015.

[125] C. Zhang, J. Zhou, C. Li, W. Fu, and T. Peng, ''A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting,'' *Energy Convers. Manage.*, vol. 143, pp. 360–376, Jul. 2017.

[126] K. G. Dhal, J. Gálvez, S. Ray, A. Das, and S. Das, ''Acute lymphoblastic leukemia image segmentation driven by stochastic fractal search,'' *Multimedia Tools Appl.*, vol. 79, pp. 12227–12255, 2020.

[127] K. M. Hosny, M. A. Elaziz, I. M. Selim, and M. M. Darwish, ''Classification of galaxy color images using quaternion polar complex exponential transform and binary stochastic fractal search,'' *Astron. Comput.*, vol. 31, Apr. 2020, Art. no. 100383.

[128] V. Tiwari, ''Face recognition based on cuckoo search algorithm,'' *Image*, vol. 7, no. 8, p. 9, 2012.

[129] D. Rodrigues, L. A. M. Pereira, T. N. S. Almeida, J. P. Papa, A. N. Souza, C. C. O. Ramos, and X.-S. Yang, ''BCS: A binary cuckoo search algorithm for feature selection,'' in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 465–468.

[130] C. Gunavathi and K. Premalatha, ''Cuckoo search optimisation for feature selection in cancer classification: A new approach,'' *Int. J. Data Mining Bioinf.*, vol. 13, no. 3, pp. 248–265, 2015.

[131] M. N. Sudha and S. Selvarajan, "Feature selection based on enhanced cuckoo search for breast cancer classification in mammogram image," *Circuits Syst.*, vol. 07, no. 04, pp. 327–338, 2016.

[132] S. Salesi and G. Cosma, "A novel extended binary cuckoo search algorithm for feature selection," in *Proc. 2nd Int. Conf. Knowl. Eng. Appl. (ICKEA)*, Oct. 2017, pp. 6–12.

[133] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Feature selection method based on hybrid data transformation and binary binomial cuckoo search," *J. Ambient Intell. Hum. Comput.*, vol. 11, no. 2, pp. 719–738, Feb. 2020.

[134] A. Frank. (2010). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml

[135] S. Sarvari, N. F. Mohd Sani, Z. Mohd Hanapi, and M. T. Abdullah, "An efficient anomaly intrusion detection method with feature selection and evolutionary neural network," *IEEE Access*, vol. 8, pp. 70651–70663, 2020.

[136] R. P. Shetty, A. Sathyabhama, and P. S. Pai, "An efficient online sequential extreme learning machine model based on feature selection and parameter optimization using cuckoo search algorithm for multi-step wind speed forecasting," *Soft Comput.*, vol. 25, pp. 1277–1295, 2021.

[137] S. Marso and M. El Merouani, "Predicting financial distress using hybrid feedforward neural network with cuckoo search algorithm," *Procedia Comput. Sci.*, vol. 170, pp. 1134–1140, 2020.

[138] R. Y. M. Nakamura, L. A. M. Pereira, K. A. Costa, D. Rodrigues, J. P. Papa, and X.-S. Yang, "BBA: A binary bat algorithm for feature selection," in *Proc. 25th SIBGRAPI Conf. Graph., Patterns Images*, Aug. 2012, pp. 291–297.

[139] M. A. Laamari and N. Kamel, "A hybrid bat based feature selection approach for intrusion detection," in *Bio-Inspired Computing-Theories and Applications*. Berlin, Germany: Springer, 2014, pp. 230–238.

[140] D. Rodrigues, L. A. M. Pereira, R. Y. M. Nakamura, K. A. P. Costa, X.-S. Yang, A. N. Souza, and J. P. Papa, "A wrapper approach for feature selection based on bat algorithm and optimum-path forest," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2250–2258, Apr. 2014.

[141] A.-C. Enache and V. Sgârciu, "A feature selection approach implemented with the binary bat algorithm applied for intrusion detection," in *Proc. 38th Int. Conf. Telecommun. Signal Process. (TSP)*, 2015, pp. 11–15.

[142] B. Yang, Y. Lu, K. Zhu, G. Yang, J. Liu, and H. Yin, "Feature selection based on modified bat algorithm," *IEICE Trans. Inf. Syst.*, vol. E100.D, no. 8, pp. 1860–1869, 2017.

[143] T. Kaur, B. S. Saini, and S. Gupta, "A novel feature selection method for brain tumor MR image classification based on the Fisher criterion and parameter-free bat optimization," *Neural Comput. Appl.*, vol. 29, no. 8, pp. 193–206, Apr. 2018.

[144] A. K. Naik, V. Kuppili, and D. R. Edla, "Efficient feature selection using one-pass generalized classifier neural network and binary bat algorithm with a novel fitness function," *Soft Comput.*, vol. 24, no. 6, pp. 4575–4587, Mar. 2020.

[145] B. Alsalibi, I. Venkat, and M. A. Al-Betar, "A membrane-inspired bat algorithm to recognize faces in unconstrained scenarios," *Eng. Appl. Artif. Intell.*, vol. 64, pp. 242–260, Sep. 2017.

[146] S.-W. Fei, "Fault diagnosis of bearing based on relevance vector machine classifier with improved binary bat algorithm for feature selection and parameter optimization," *Adv. Mech. Eng.*, vol. 9, no. 1, 2017, Art. no. 1687814016685294.

[147] S. Jeyasingh and M. Veluchamy, "Modified bat algorithm for feature selection with the Wisconsin diagnosis breast cancer (WDBC) dataset," *Asian Pacific J. Cancer Prevention*, vol. 18, no. 5, p. 1257, 2017.

[148] T. Niu, J. Wang, K. Zhang, and P. Du, "Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy," *Renew. Energy*, vol. 118, pp. 213–229, Apr. 2018.

[149] E. Emary, H. M. Zawbaa, K. K. A. Ghany, A. E. Hassanien, and B. Parv, "Firefly optimization algorithm for feature selection," in *Proc. 7th Balkan Conf. Informat. Conf.*, Sep. 2015, pp. 1–7.

[150] T. Kanimozhi and K. Latha, "An integrated approach to region based image retrieval using firefly algorithm and support vector machine," *Neurocomputing*, vol. 151, pp. 1099–1111, Mar. 2015.

[151] V. Subha and D. Murugan, "Opposition based firefly algorithm optimized feature subset selection approach for fetal risk anticipation," *Mach. Learn. Appl., Int. J.*, vol. 3, no. 2, pp. 55–64, Jun. 2016.

[152] Y. Zhang, X.-F. Song, and D.-W. Gong, "A return-cost-based binary firefly algorithm for feature selection," *Inf. Sci.*, vols. 418–419, pp. 561–574, Dec. 2017.

[153] L. Zhang, L. Shan, and J. Wang, "Optimal feature selection using distance-based discrete firefly algorithm with mutual information criterion," *Neural Comput. Appl.*, vol. 28, no. 9, pp. 2795–2808, Sep. 2017.

[154] H. Xu, S. Yu, J. Chen, and X. Zuo, "An improved firefly algorithm for feature selection in classification," *Wireless Pers. Commun.*, vol. 102, no. 4, pp. 2823–2834, Oct. 2018.

[155] B. Selvakumar and K. Muneeswaran, "Firefly algorithm based feature selection for network intrusion detection," *Comput. Secur.*, vol. 81, pp. 148–155, Mar. 2019.

[156] A. Kazem, E. Sharifi, F. K. Hussain, M. Saberi, and O. K. Hussain, "Support vector regression with chaos-based firefly algorithm for stock market price forecasting," *Appl. Soft Comput.*, vol. 13, no. 2, pp. 947–958, Feb. 2013.

[157] L. Zhang, K. Mistry, C. P. Lim, and S. C. Neoh, "Feature selection using firefly optimization for classification and regression models," *Decis. Support Syst.*, vol. 106, pp. 64–85, Feb. 2018.

[158] R. R. Chhikara, P. Sharma, and L. Singh, "An improved dynamic discrete firefly algorithm for blind image steganalysis," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 5, pp. 821–835, May 2018.

[159] S. Dash, R. Thulasiram, and P. Thulasiraman, "Modified firefly algorithm with chaos theory for feature selection: A predictive model for medical data," *Int. J. Swarm Intell. Res.*, vol. 10, no. 2, pp. 1–20, Apr. 2019.

[160] D. Rodrigues, X.-S. Yang, A. N. De Souza, and J. P. Papa, "Binary flower pollination algorithm and its application to feature selection," in *Recent Advances in Swarm Intelligence and Evolutionary Computation*. Cham, Switzerland: Springer, 2015, pp. 85–100.

[161] H. M. Zawbaa and E. Emary, "Applications of flower pollination algorithm in feature selection and knapsack problems," in *Nature-Inspired Algorithms and Applied Optimization*. Cham, Switzerland: Springer, 2018, pp. 217–243.

[162] S. P. Rajamohana, K. Umamaheswari, and B. Abirami, "Adaptive binary flower pollination algorithm for feature selection in review spam detection," in *Proc. Int. Conf. Innov. Green Energy Healthcare Technol. (IGEHT)*, 2017, pp. 1–4.

[163] H. Majidpour and F. Soleimanian Gharehchopogh, "An improved flower pollination algorithm with AdaBoost algorithm for feature selection in text documents classification," *J. Adv. Comput. Res.*, vol. 9, no. 1, pp. 29–40, 2018.

[164] S. A.-F. Sayed, E. Nabil, and A. Badr, "A binary clonal flower pollination algorithm for feature selection," *Pattern Recognit. Lett.*, vol. 77, pp. 21–27, Jul. 2016.

[165] C. Yan, J. Ma, H. Luo, G. Zhang, and J. Luo, "A novel feature selection method for high-dimensional biomedical data based on an improved binary clonal flower pollination algorithm," *Hum. Heredity*, vol. 84, no. 1, pp. 1–13, 2019.

[166] D. Rodrigues, L. A. M. Pereira, J. P. Papa, and S. A. T. Weber, "A binary krill herd approach for feature selection," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 1407–1412.

[167] A. Mohammadi, M. S. Abadeh, and H. Keshavarz, "Breast cancer detection using a multi-objective binary krill herd algorithm," in *Proc. 21th Iranian Conf. Biomed. Eng. (ICBME)*, Nov. 2014, pp. 128–133.

[168] R. R. Rani and D. Ramyachitra, "Krill herd optimization algorithm for cancer feature selection and random forest technique for classification," in *Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Nov. 2017, pp. 109–113.

[169] G. Zhang, J. Hou, J. Wang, C. Yan, and J. Luo, "Feature selection for microarray data classification using hybrid information gain and a modified binary krill herd algorithm," *Interdiscipl. Sci., Comput. Life Sci.*, vol. 12, no. 3, pp. 288–301, Sep. 2020.

[170] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.

[171] P. Sharma, S. Sundaram, M. Sharma, A. Sharma, and D. Gupta, "Diagnosis of Parkinson's disease using modified grey wolf optimization," *Cognit. Syst. Res.*, vol. 54, pp. 100–115, May 2019.

[172] Y. Pathak, K. V. Arya, and S. Tiwari, "Feature selection for image steganalysis using levy flight-based grey wolf optimization," *Multimedia Tools Appl.*, vol. 78, no. 2, pp. 1473–1494, Jan. 2019.

[173] K. Devanathan, N. Ganapathy, and R. Swaminathan, "Binary grey wolf optimizer based feature selection for nucleolar and centromere staining pattern classification in indirect immunofluorescence images," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 7040–7043.
[174] Q. Al-Tashi, H. Rais, and S. Jadid, "Feature selection method based on grey wolf optimization for coronary artery disease classification," in *Proc. Int. Conf. Reliable Inf. Commun. Technol.* Cham, Switzerland: Springer, 2018, pp. 257–266.
[175] Q. Al-Tashi, S. J. Abdulkadir, H. M. Rais, S. Mirjalili, H. Alhussian, M. G. Ragab, and A. Alqushaibi, "Binary multi-objective grey wolf optimizer for feature selection in classification," *IEEE Access*, vol. 8, pp. 106247–106263, 2020.
[176] P. Hu, J.-S. Pan, and S.-C. Chu, "Improved binary grey wolf optimizer and its application for feature selection," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105746.
[177] Q. Li, H. Chen, H. Huang, X. Zhao, Z. Cai, C. Tong, W. Liu, and X. Tian, "An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–15, Jan. 2017.
[178] A. Sahoo and S. Chandra, "Multi-objective grey wolf optimizer for improved cervix lesion classification," *Appl. Soft Comput.*, vol. 52, pp. 64–80, Mar. 2017.
[179] J. Too, A. Abdullah, N. Mohd Saad, N. Mohd Ali, and W. Tee, "A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification," *Computers*, vol. 7, no. 4, p. 58, Nov. 2018.
[180] N. P. Nirmala Sreedharan, B. Ganesan, R. Raveendran, P. Sarala, B. Dennis, and R. Boothalingam, "Grey wolf optimisation-based feature selection and classification for facial emotion recognition," *IET Biometrics*, vol. 7, no. 5, pp. 490–499, Sep. 2018.
[181] A. K. Abasi, A. T. Khader, M. A. Al-Betar, S. Naim, S. N. Makhadmeh, and Z. A. A. Alyasseri, "An improved text feature selection for clustering using binary grey wolf optimizer," in *Proc. 11th Nat. Tech. Seminar Unmanned Syst. Technol.* Singapore: Springer, 2021, pp. 503–516.
[182] H. M. Zawbaa, E. Emary, and B. Parv, "Feature selection based on antlion optimization algorithm," in *Proc. 3rd World Conf. Complex Syst. (WCCS)*, Nov. 2015, pp. 1–7.
[183] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016.
[184] M. Mafarja, D. Eleyan, S. Abdullah, and S. Mirjalili, "S-shaped vs. V-shaped transfer functions for ant lion optimization algorithm in feature selection problem," in *Proc. Int. Conf. Future Netw. Distrib. Syst.*, Jul. 2017, pp. 1–7.
[185] H. M. Zawbaa, E. Emary, and C. Grosan, "Feature selection via chaotic antlion optimization," *PLoS ONE*, vol. 11, no. 3, Mar. 2016, Art. no. e0150652.
[186] E. Emary and H. M. Zawbaa, "Feature selection via Lèvy antlion optimization," *Pattern Anal. Appl.*, vol. 22, no. 3, pp. 857–876, Aug. 2019.
[187] M. M. Mafarja and S. Mirjalili, "Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection," *Soft Comput.*, vol. 23, no. 15, pp. 6249–6265, Aug. 2019.
[188] M. Wang, C. Wu, L. Wang, D. Xiang, and X. Huang, "A feature selection approach for hyperspectral image based on modified ant lion optimizer," *Knowl.-Based Syst.*, vol. 168, pp. 39–48, Mar. 2019.
[189] A. T. Azar, N. Banu, and A. Koubaa, "Rough set based ant-lion optimizer for feature selection," in *Proc. 6th Conf. Data Sci. Mach. Learn. Appl. (CDMA)*, Mar. 2020, pp. 81–86.
[190] S. A. Medjahed, T. A. Saadi, A. Benyettou, and M. Ouali, "Kernel-based learning and feature selection analysis for cancer diagnosis," *Appl. Soft Comput.*, vol. 51, pp. 39–48, Feb. 2017.
[191] M. M. Mafarja, D. Eleyan, I. Jaber, A. Hammouri, and S. Mirjalili, "Binary dragonfly algorithm for feature selection," in *Proc. Int. Conf. New Trends Comput. Sci. (ICTCS)*, Oct. 2017, pp. 12–17.
[192] M. Mafarja, I. Aljarah, A. A. Heidari, H. Faris, P. Fournier-Viger, X. Li, and S. Mirjalili, "Binary dragonfly optimization for feature selection using time-varying transfer functions," *Knowl.-Based Syst.*, vol. 161, pp. 185–204, Dec. 2018.
[193] A. A. Karizaki and M. Tavassoli, "A novel hybrid feature selection based on ReliefF and binary dragonfly for high dimensional datasets," in *Proc. 9th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Oct. 2019, pp. 300–304.
[194] A. Naik, V. Kuppili, and D. Reddy Edla, "Binary dragonfly algorithm and Fisher score based hybrid feature selection adopting a novel fitness function applied to microarray data," in *Proc. Int. Conf. Appl. Mach. Learn. (ICAML)*, May 2019, pp. 40–43.
[195] M. Yasen, N. Al-Madi, and N. Obeid, "Optimizing neural networks using dragonfly algorithm for medical prediction," in *Proc. 8th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, Jul. 2018, pp. 71–76.
[196] R. Sawhney and R. Jain, "Modified binary dragonfly algorithm for feature selection in human papillomavirus-mediated disease treatment," in *Proc. Int. Conf. Commun., Comput. Internet Things (IC3IoT)*, Feb. 2018, pp. 91–95.
[197] J. Li, H. Kang, G. Sun, T. Feng, W. Li, W. Zhang, and B. Ji, "IBDA: Improved binary dragonfly algorithm with evolutionary population dynamics and adaptive crossover for feature selection," *IEEE Access*, vol. 8, pp. 108032–108051, 2020.
[198] X. Cui, Y. Li, J. Fan, T. Wang, and Y. Zheng, "A hybrid improved dragonfly algorithm for feature selection," *IEEE Access*, vol. 8, pp. 155619–155629, 2020.
[199] A. G. Hussien, A. E. Hassanien, E. H. Houssein, S. Bhattacharyya, and M. Amin, "S-shaped binary whale optimization algorithm for feature selection," in *Recent Trends in Signal and Image Processing*. Singapore: Springer, 2019, pp. 79–87.
[200] A. G. Hussien, E. H. Houssein, and A. E. Hassanien, "A binary whale optimization algorithm with hyperbolic tangent fitness function for feature selection," in *Proc. 8th Int. Conf. Intell. Comput. Inf. Syst. (ICICIS)*, Dec. 2017, pp. 166–172.
[201] G. I. Sayed, A. Darwish, and A. E. Hassanien, "A new chaotic whale optimization algorithm for features selection," *J. Classification*, vol. 35, no. 2, pp. 300–344, Jul. 2018.
[202] M. Tubishat, M. A. M. Abushariah, N. Idris, and I. Aljarah, "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis," *Int. J. Speech Technol.*, vol. 49, no. 5, pp. 1688–1707, May 2019.
[203] M. Mafarja and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," *Appl. Soft Comput.*, vol. 62, pp. 441–453, Jan. 2018.
[204] R. K. Agrawal, B. Kaur, and S. Sharma, "Quantum based whale optimization algorithm for wrapper feature selection," *Appl. Soft Comput.*, vol. 89, Apr. 2020, Art. no. 106092.
[205] Y. Zheng, Y. Li, G. Wang, Y. Chen, Q. Xu, J. Fan, and X. Cui, "A novel hybrid algorithm for feature selection based on whale optimization algorithm," *IEEE Access*, vol. 7, pp. 14908–14923, 2019.
[206] Q.-T. Bui, M. V. Pham, Q.-H. Nguyen, L. X. Nguyen, and H. M. Pham, "Whale optimization algorithm and adaptive neuro-fuzzy inference system: A hybrid method for feature selection and land pattern classification," *Int. J. Remote Sens.*, vol. 40, no. 13, pp. 5078–5093, Jul. 2019.
[207] M. A. Tawhid and A. M. Ibrahim, "Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 3, pp. 573–602, Mar. 2020.
[208] H. T. Ibrahim, W. J. Mazher, O. N. Ucan, and O. Bayat, "A grasshopper optimizer approach for feature selection and optimizing SVM parameters utilizing real biomedical data sets," *Neural Comput. Appl.*, vol. 31, no. 10, pp. 5965–5974, 2019.
[209] M. Mafarja, I. Aljarah, A. A. Heidari, A. I. Hammouri, H. Faris, A. M. Al-Zoubi, and S. Mirjalili, "Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems," *Knowl.-Based Syst.*, vol. 145, pp. 25–45, Apr. 2018.
[210] H. Hichem, M. Elkamel, M. Rafik, M. T. Mesaaoud, and C. Ouahiba, "A new binary grasshopper optimization algorithm for feature selection problem," *J. King Saud Univ.-Comput. Inf. Sci.*, Nov. 2019, doi: 10.1016/j.jksuci.2019.11.007.
[211] M. Mafarja, I. Aljarah, H. Faris, A. I. Hammouri, A. M. Al-Zoubi, and S. Mirjalili, "Binary grasshopper optimisation algorithm approaches for feature selection problems," *Expert Syst. Appl.*, vol. 117, pp. 267–286, Mar. 2019.
[212] H. T. Ibrahim, W. J. Mazher, O. N. Ucan, and O. Bayat, "Feature selection using salp swarm algorithm for real biomedical datasets," *IJCSNS*, vol. 17, no. 12, pp. 13–20, 2017.
[213] G. I. Sayed, G. Khoriba, and M. H. Haggag, "A novel chaotic salp swarm algorithm for global optimization and feature selection," *Int. J. Speech Technol.*, vol. 48, no. 10, pp. 3462–3481, Oct. 2018.

[214] H. Faris, M. M. Mafarja, A. A. Heidari, I. Aljarah, A. M. Al-Zoubi, S. Mirjalili, and H. Fujita, "An efficient binary salp swarm algorithm with crossover scheme for feature selection problems," *Knowl.-Based Syst.*, vol. 154, pp. 43–67, Aug. 2018.
[215] M. Tubishat, S. Ja'afar, M. Alswaitti, S. Mirjalili, N. Idris, M. A. Ismail, and M. S. Omar, "Dynamic salp swarm algorithm for feature selection," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 113873.
[216] A. E. Hegazy, M. A. Makhlouf, and G. S. El-Tawel, "Improved salp swarm algorithm for feature selection," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 32, no. 3, pp. 335–344, Mar. 2020.
[217] A. Ibrahim, A. Ahmed, S. Hussein, and A. E. Hassanien, "Fish image segmentation using salp swarm algorithm," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.* Cham, Switzerland: Springer, 2018, pp. 42–51.
[218] M. Tubishat, N. Idris, L. Shuib, M. A. M. Abushariah, and S. Mirjalili, "Improved salp swarm algorithm based on opposition based learning and novel local search algorithm for feature selection," *Expert Syst. Appl.*, vol. 145, May 2020, Art. no. 113122.
[219] S. K. Baliarsingh, S. Vipsita, K. Muhammad, and S. Bakshi, "Analysis of high-dimensional biomedical data using an evolutionary multi-objective emperor penguin optimizer," *Swarm Evol. Comput.*, vol. 48, pp. 262–273, Aug. 2019.
[220] S. K. Baliarsingh and S. Vipsita, "Chaotic emperor penguin optimised extreme learning machine for microarray cancer classification," *IET Syst. Biol.*, vol. 14, no. 2, pp. 85–95, Apr. 2020.
[221] G. Dhiman, D. Oliva, A. Kaur, K. K. Singh, S. Vimal, A. Sharma, and K. Cengiz, "BEPO: A novel binary emperor penguin optimizer for automatic feature selection," *Knowl.-Based Syst.*, vol. 211, Jan. 2021, Art. no. 106560.
[222] A. K. Abasi, A. T. Khader, M. A. Al-Betar, S. Naim, S. N. Makhadmeh, and Z. A. A. Alyasseri, "A text feature selection technique based on binary multi-verse optimizer for text clustering," in *Proc. IEEE Jordan Int. Joint Conf. Electr. Eng. Inf. Technol. (JEEIT)*, Apr. 2019, pp. 1–6.
[223] J. P. Papa, G. H. Rosa, A. N. de Souza, and L. C. S. Afonso, "Feature selection through binary brain storm optimization," *Comput. Electr. Eng.*, vol. 72, pp. 468–481, Nov. 2018.
[224] F. Pourpanah, C. P. Lim, X. Wang, C. J. Tan, M. Seera, and Y. Shi, "A hybrid model of fuzzy min–max and brain storm optimization for feature selection and data classification," *Neurocomputing*, vol. 333, pp. 440–451, Mar. 2019.
[225] F. Pourpanah, Y. Shi, C. P. Lim, Q. Hao, and C. J. Tan, "Feature selection based on brain storm optimization for data classification," *Appl. Soft Comput.*, vol. 80, pp. 761–775, Jul. 2019.
[226] E. Tuba, I. Strumberger, T. Bezdan, N. Bacanin, and M. Tuba, "Classification and feature selection method for medical datasets by brain storm optimization algorithm and support vector machine," *Procedia Comput. Sci.*, vol. 162, pp. 307–315, Jan. 2019.
[227] D. Oliva and M. A. Elaziz, "An improved brainstorm optimization using chaotic opposite-based learning with disruption operator for global optimization and feature selection," *Soft Comput.*, vol. 24, no. 18, pp. 14051–14072, 2020.
[228] M. S. Krishna and S. Vishwakarma, "Improve the performance of fingerprint recognition using feature selection and teacher learning based optimization (TLBO)," *Int. J. Master Eng. Res. Technol.*, vol. 2, no. 10, pp. 167–172, 2015.
[229] K. Jain and S. S. Bhadauria, "Enhanced content based image retrieval using feature selection using teacher learning based optimization," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 11, pp. 1052–1057, 2016.
[230] H. E. Kiziloz, A. Deniz, T. Dokeroglu, and A. Cosar, "Novel multiobjective TLBO algorithms for the feature subset selection problem," *Neurocomputing*, vol. 306, pp. 94–107, Sep. 2018.
[231] M. Allam and M. Nandhini, "Optimal feature selection using binary teaching learning based optimization algorithm," *J. King Saud Univ.-Comput. Inf. Sci.*, Dec. 2018, doi: 10.1016/j.jksuci.2018.12.001.
[232] M. Manonmani and S. Balakrishnan, "Feature selection using improved teaching learning based algorithm on chronic kidney disease dataset," *Procedia Comput. Sci.*, vol. 171, pp. 1660–1669, 2020.
[233] H. Das, B. Naik, and H. S. Behera, "A Jaya algorithm based wrapper method for optimal feature selection in supervised classification," *J. King Saud Univ.-Comput. Inf. Sci.*, May 2020, doi: 10.1016/j.jksuci.2020.05.002.
[234] M. A. Awadallah, M. A. Al-Betar, A. I. Hammouri, and O. A. Alomari, "Binary JAYA algorithm with adaptive mutation for feature selection," *Arabian J. Sci. Eng.*, vol. 45, no. 12, pp. 10875–10890, Dec. 2020.
[235] P. Agrawal, T. Ganesh, and A. W. Mohamed, "A novel binary gaining–sharing knowledge-based optimization algorithm for feature selection," *Neural Comput. Appl.*, pp. 1–20, Oct. 2020, doi: 10.1007/s00521-020-05375-8.
[236] A. I. Hafez, A. E. Hassanien, H. M. Zawbaa, and E. Emary, "Hybrid monkey algorithm with krill herd algorithm optimization for feature selection," in *Proc. 11th Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2015, pp. 273–277.
[237] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017.
[238] S. Arora, H. Singh, M. Sharma, S. Sharma, and P. Anand, "A new hybrid algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection," *IEEE Access*, vol. 7, pp. 26343–26361, 2019.
[239] M. E. Abd Elaziz, A. A. Ewees, D. Oliva, P. Duan, and S. Xiong, "A hybrid method of sine cosine algorithm and differential evolution for feature selection," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2017, pp. 145–155.
[240] M. A. Tawhid and K. B. Dsouza, "Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems," *Appl. Comput. Informat.*, vol. 16, nos. 1–2, pp. 117–136, Apr. 2018.
[241] E. Pashaei, E. Pashaei, and N. Aydin, "Gene selection using hybrid binary black hole algorithm and modified binary particle swarm optimization," *Genomics*, vol. 111, no. 4, pp. 669–686, Jul. 2019.
[242] N. Neggaz, A. A. Ewees, M. A. Elaziz, and M. Mafarja, "Boosting salp swarm algorithm by sine cosine algorithm and disrupt operator for feature selection," *Expert Syst. Appl.*, vol. 145, May 2020, Art. no. 113103.
[243] S. K. Baliarsingh, W. Ding, S. Vipsita, and S. Bakshi, "A memetic algorithm using emperor penguin and social engineering optimization for medical data classification," *Appl. Soft Comput.*, vol. 85, Dec. 2019, Art. no. 105773.
[244] J. Yang and H. Gao, "Cultural emperor penguin optimizer and its application for face recognition," *Math. Problems Eng.*, vol. 2020, pp. 1–16, Nov. 2020.
[245] A. K. Shukla, P. Singh, and M. Vardhan, "A new hybrid wrapper TLBO and SA with SVM approach for gene expression data," *Inf. Sci.*, vol. 503, pp. 238–254, Nov. 2019.
[246] S. K. Baliarsingh, S. Vipsita, and B. Dash, "A new optimal gene selection approach for cancer classification using enhanced jaya-based forest optimization algorithm," *Neural Comput. Appl.*, vol. 32, no. 12, pp. 8599–8616, Jun. 2020.
[247] R. Purushothaman, S. P. Rajagopalan, and G. Dhandapani, "Hybridizing gray wolf optimization (GWO) with grasshopper optimization algorithm (GOA) for text feature selection and clustering," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106651.
[248] E. O. Reséndiz-Flores, J. A. Navarro-Acosta, and A. Hernández-Martínez, "Optimal feature selection in industrial foam injection processes using hybrid binary particle swarm optimization and gravitational search algorithm in the Mahalanobis–Taguchi system," *Soft Comput.*, vol. 24, no. 1, pp. 341–349, Jan. 2020.
[249] E.-S.-M. El-Kenawy, M. M. Eid, M. Saber, and A. Ibrahim, "MbGWO-SFS: Modified binary grey wolf optimizer based on stochastic fractal search for feature selection," *IEEE Access*, vol. 8, pp. 107635–107649, 2020.
[250] P. Bansal, S. Kumar, S. Pasrija, and S. Singh, "A hybrid grasshopper and new cat swarm optimization algorithm for feature selection and optimization of multi-layer perceptron," *Soft Comput.*, pp. 1–27, Mar. 2020.
[251] V. Bolón-Canedo, D. Rego-Fernández, D. Peteiro-Barral, A. Alonso-Betanzos, B. Guijarro-Berdiñas, and N. Sánchez-Maroño, "On the scalability of feature selection methods on high-dimensional data," *Knowl. Inf. Syst.*, vol. 56, no. 2, pp. 395–442, 2018.
[252] R. Wald, T. M. Khoshgoftaar, and A. Napolitano, "Stability of filter- and wrapper-based feature subset selection," in *Proc. IEEE 25th Int. Conf. Tools with Artif. Intell.*, Nov. 2013, pp. 374–380.
[253] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *J. King Saud Univ.-Comput. Inf. Sci.*, Jun. 2019, doi: 10.1016/j.jksuci.2019.06.012.
[254] H. Bostani and M. Sheikhan, "Hybrid of binary gravitational search algorithm and mutual information for feature selection in intrusion detection systems," *Soft Comput.*, vol. 21, no. 9, pp. 2307–2324, May 2017.

[255] G. Liu, B. Zhang, X. Ma, and J. Wang, "Network intrusion detection based on chaotic multi-verse optimizer," in *Proc. 11th EAI Int. Conf. Mobile Multimedia Commun.* Ghent, Belgium: European Alliance for Innovation, 2018, p. 218.
[256] F. Barani, M. Mirhosseini, and H. Nezamabadi-pour, "Application of binary quantum-inspired gravitational search algorithm in feature subset selection," *Int. J. Speech Technol.*, vol. 47, no. 2, pp. 304–318, Sep. 2017.
[257] A. A. Ewees, M. A. El Aziz, and A. E. Hassanien, "Chaotic multi-verse optimizer-based feature selection," *Neural Comput. Appl.*, vol. 31, no. 4, pp. 991–1006, Apr. 2019.
[258] M. Belazzoug, M. Touahria, F. Nouioua, and M. Brahimi, "An improved sine cosine algorithm to select features for text categorization," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 32, no. 4, pp. 454–464, May 2020.
[259] X. Han, X. Chang, L. Quan, X. Xiong, J. Li, Z. Zhang, and Y. Liu, "Feature subset selection by gravitational search algorithm optimization," *Inf. Sci.*, vol. 281, pp. 128–146, Oct. 2014.
[260] E. Pashaei, M. Ozen, and N. Aydin, "An application of black hole algorithm and decision tree for medical problem," in *Proc. IEEE 15th Int. Conf. Bioinf. Bioeng. (BIBE)*, Nov. 2015, pp. 1–6.
[261] E. Pashaei and N. Aydin, "Binary black hole algorithm for feature selection and classification on biological data," *Appl. Soft Comput.*, vol. 56, pp. 94–106, Jul. 2017.
[262] B. Mohseni Bababdani and M. Mousavi, "Gravitational search algorithm: A new feature selection method for QSAR study of anticancer potency of imidazo[4,5-b]pyridine derivatives," *Chemometric Intell. Lab. Syst.*, vol. 122, pp. 1–11, Mar. 2013.
[263] A. Adam, N. Mokhtar, M. Mubin, Z. Ibrahim, M. Z. M. Tumari, and M. I. Shapiai, "Feature selection and classifier parameter estimation for EEG signal peak detection using gravitational search algorithm," in *Proc. 4th Int. Conf. Artif. Intell. with Appl. Eng. Technol.*, Dec. 2014, pp. 103–108.
[264] M. Taradeh, M. Mafarja, A. A. Heidari, H. Faris, I. Aljarah, S. Mirjalili, and H. Fujita, "An evolutionary gravitational search-based feature selection," *Inf. Sci.*, vol. 497, pp. 219–239, Sep. 2019.
[265] R. Guha, M. Ghosh, A. Chakrabarti, R. Sarkar, and S. Mirjalili, "Introducing clustering based population in binary gravitational search algorithm for feature selection," *Appl. Soft Comput.*, vol. 93, Aug. 2020, Art. no. 106341.
[266] C. C. O. Ramos, D. Rodrigues, A. N. de Souza, and J. P. Papa, "On the study of commercial losses in Brazil: A binary black hole algorithm for theft characterization," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 676–683, Mar. 2018.
[267] R. Hans and H. Kaur, "Binary multi-verse optimization (BMVO) approaches for feature selection," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 6, no. 1, p. 91, 2020.
[268] R. Sindhu, R. Ngadiran, Y. M. Yacob, N. A. H. Zahri, and M. Hariharan, "Sine–cosine algorithm for feature selection with elitism strategy and new updating mechanism," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2947–2958, Oct. 2017.
[269] S. Taghian and M. H. Nadimi-Shahraki, "Binary sine cosine algorithms for feature selection from medical data," 2019, *arXiv:1911.07805*. [Online]. Available: http://arxiv.org/abs/1911.07805
[270] L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, "Improved binary PSO for feature selection using gene expression data," *Comput. Biol. Chem.*, vol. 32, no. 1, pp. 29–38, Feb. 2008.
[271] X. He, Q. Zhang, N. Sun, and Y. Dong, "Feature selection with discrete binary differential evolution," in *Proc. Int. Conf. Artif. Intell. Comput. Intell.*, vol. 4, 2009, pp. 327–330.
[272] N. Al-Madi, H. Faris, and S. Mirjalili, "Binary multi-verse optimization algorithm for global optimization and discrete problems," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 12, pp. 3445–3465, Dec. 2019.
[273] J. P. Papa, A. Pagnin, S. A. Schellini, A. Spadotto, R. C. Guido, M. Ponti, G. Chiachia, and A. X. Falcao, "Feature selection through gravitational search algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 2052–2055.
[274] M. Bardamova, A. Konev, I. Hodashinsky, and A. Shelupanov, "A fuzzy classifier with feature selection based on the gravitational search algorithm," *Symmetry*, vol. 10, no. 11, p. 609, Nov. 2018.
[275] X. Han, D. Li, P. Liu, and L. Wang, "Feature selection by recursive binary gravitational search algorithm optimization for cancer classification," *Soft Comput.*, vol. 24, no. 6, pp. 4407–4425, Mar. 2020.
[276] D. Rodrigues, L. A. Pereira, J. P. Papa, C. C. Ramos, A. N. Souza, and L. P. Papa, "Optimizing feature selection through binary charged system search," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Berlin, Germany: Springer, 2013, pp. 377–384.
[277] O. S. Qasim, N. A. Al-Thanoon, and Z. Y. Algamal, "Feature selection based on chaotic binary black hole algorithm for data classification," *Chemometric Intell. Lab. Syst.*, vol. 204, Sep. 2020, Art. no. 104104.
[278] N. Dif and Z. Elberrichi, "Microarray data feature selection and classification using an enhanced multi-verse optimizer and support vector machine," in *Proc. 3rd Int. Conf. Netw. Adv. Syst.*, 2017.
[279] A. I. Hafez, H. M. Zawbaa, E. Emary, and A. E. Hassanien, "Sine cosine optimization algorithm for feature selection," in *Proc. Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, Aug. 2016, pp. 1–5.

PRACHI AGRAWAL received the B.Sc. degree in mathematics from the Ramjas College, Delhi University, Delhi, India, in 2012, and the M.Sc. degree in mathematics from the Indian Institute of Technology Delhi, India, in 2014. She is currently pursuing the Ph.D. degree with the Department of Mathematics and Scientific Computing, National Institute of Technology Hamirpur, India. She has published more than ten research articles in peer reviewed journals such as neural computing and applications, OPSERCH, International journal of swarm and intelligent research, Complexity and presented various papers in international conferences. Her main research interests are in the areas of stochastic optimization using computational intelligence techniques. Her research areas include optimization, metaheuristic algorithms, and evolutionary programming.

HATTAN F. ABUTARBOUSH received the B.Sc (Hons.) degree in electrical communications and electronics engineering from Greenwich University, London, U.K., in 2005, the M.Sc. degree in mobile personal and satellite communications from the Department of Electrical and Electronics Engineering, Westminster University, London, U.K., in 2007, and the Ph.D. degree in antennas and propagation from the Department of Electronics and Computer Engineering, Brunel University, London, U.K., in July 2011. He was a Research Visitor to the Hong Kong University and National Physical Laboratory (NPL), Teddington, U.K., in 2010. He worked as a Research Associate for the American University in Cairo (AUC), in 2011, where he worked on the packaging of novel mm-wave antennas. He was a Research Fellow at the King Abdullah University of Science and Technology (KAUST), working on mm-wave and inkjet-printed antennas, from 2011 to 2013. He also worked as a Research Assistant in microwave medical imaging at Bristol University, Bristol, U.K., from 2013 to 2015. In 2015, he joined the Electrical Engineering Department, Taibah University, where he is currently an Assistant Professor. He received the Vice-Chancellor's Prize Award for the Best Doctoral Research for 2011 from Brunel University, London, U.K., and the outstanding achievement award from the Saudi Arabian Cultural Bureau in London, for the outstanding overall achievement in the Ph.D. degree. Also, he was selected for the top 10 Reviewers Award by the Editorial Board for his exceptional performance in the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, in 2014 and 2015. He has authored/coauthored over 50 technical articles in international journals and conferences. He also has served as a Reviewer and an Associate Editor for different international journals and conferences in the areas of antennas and propagation. His current research interests lie in the design of reconfigurable antennas, antennas for mobile phones, RF/microwave circuit design, mm-wave antennas, biosensors, and microwave medical imaging for breast and brain cancer detection.

TALARI GANESH received the M.Sc. and Ph.D. degrees in statistics from Sri Venkateswara University, Tirupati, Andhra Pradesh. He is currently an Assistant Professor with the Department of Mathematics and Scientific Computing, National Institute of Technology, Hamirpur, Himachal Pradesh. He had worked as an Assistant Professor with the Seshadripuram Institute of Management Studies, Bangalore, and also joined as a Guest Faculty at the Department of Statistics, Pondicherry University, Puducherry, from September 2014 to August 2015. He was a Lecturer at the Dr. Jyothirmayi Degree College, from 2009 to 2010. His main research interests are in the areas of multi-criterion decision aid, operations research, statistics and big data science. He has worked on stochastic modeling, time series analysis, and forecasting by applying to the gold mineralization and rainfall data using SPSS and R tool respectively. He also worked in analytical hierarchy process and goal programming in dealing with multi criterion decision aid. Also, he has published various papers in the field of distribution theory, convolution and mixture distribution. He is currently working on classification techniques, data mining and its applications to the wide varieties of high-volume data which comes under big data analytics and data science. His current research is focused on multi-choice stochastic transportation problem involving fuzzy programming and evolutionary algorithms. He has published more than 12 articles in reputed international and national journals and presented more than ten papers in various international and national conferences.

ALI WAGDY MOHAMED received the B.Sc., M.Sc., and Ph.D. degrees from Cairo University, in 2000, 2004 and 2010, respectively. He is an Associate Professor with the Operations Research Department, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt. He is currently an Associate Professor of statistics with the Wireless Intelligent Networks Center (WINC), Faculty of Engineering and Applied Sciences, Nile University. Recently, he has recognized among the top 2% scientists according to Stanford University report 2019. He serves as a Reviewer of more than 60 international accredited top-tier journals and has been awarded the Publons Peer Review Awards 2018, for placing in the top 1% of reviewers worldwide in assorted field. He is editor in more than ten journals of information sciences, applied mathematics, engineering, system science, and operations research. He has presented and participated in more than five international conferences. He participated as a member of the reviewer committee for 35 different conferences sponsored by Springer and IEEE. He has Obtained Rank 3 in CEC'17 competition on single objective bound constrained real-parameter numerical optimization in Proc of IEEE Congress on Evolutionary Computation, IEEE-CEC 2017, San Sebastián, Spain. Besides, Obtained Rank 3 and Rank 2 in CEC'18 competitions on single objective bound constrained real-parameter numerical optimization and Competition on Large scale global optimization, in Proc of IEEE Congress on Evolutionary Computation, IEEE-CEC 2017, Sao Paolo, Brazil. He published more than 55 papers in reputed and high impact journals like *Information Sciences, Swarm and Evolutionary Computation, Computers & Industrial Engineering, Intelligent Manufacturing, Soft Computing and International Journal of Machine Learning and Cybernetics*. He is interested in mathematical and statistical modeling, stochastic and deterministic optimization, swarm intelligence and evolutionary computation. Additionally, he is also interested in real world problems such as industrial, transportation, manufacturing, education and capital investment problems. Two Ph.D. and one master have been completed under his supervision.