Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

A Novel Chaotic Interior Search Algorithm for Global Optimization and Feature Selection

Sankalap Arora, Manik Sharma & Priyanka Anand

To cite this article: Sankalap Arora, Manik Sharma & Priyanka Anand (2020) A Novel Chaotic Interior Search Algorithm for Global Optimization and Feature Selection, Applied Artificial Intelligence, 34:4, 292-328, DOI: 10.1080/08839514.2020.1712788

To link to this article: https://doi.org/10.1080/08839514.2020.1712788

Published online: 23 Jan 2020.

APPLIED ARTIFICIAL INTELLIGENCE
2020, VOL. 34, NO. 4, 292–328
https://doi.org/10.1080/08839514.2020.1712788

A Novel Chaotic Interior Search Algorithm for Global Optimization and Feature Selection

Sankalap Arora[a], Manik Sharma[b], and Priyanka Anand[c]

[a]Department of Computer Science and Engineering, DAV University, Jalandhar, India; [b]Department of Computer Science and Applications, DAV University, Jalandhar, India; [c]Department of Computer Science and Engineering, Lovely Professional University, Jalandhar, India

ABSTRACT

Interior Search Algorithm (ISA) is a recently proposed metaheuristic inspired by the beautification of objects and mirrors. However, similar to most of the metaheuristic algorithms, ISA also encounters two problems, i.e., entrapment in local optima and slow convergence speed. In the past, chaos theory has been successfully employed to solve such problems. In this study, 10 chaotic maps are embedded to improve the convergence rate as well as the resulting accuracy of the ISA algorithms. The proposed Chaotic Interior Search Algorithm (CISA) is validated on a diverse subset of 13 benchmark functions having unimodal and multimodal properties. The simulation results demonstrate that the chaotic maps (especially tent map) are able to significantly boost the performance of ISA. Furthermore, CISA is employed as a feature selection technique in which the aim is to remove features which may comprise irrelevant or redundant information in order to minimize the classification error rate. The performance of the proposed approaches is compared with five state-of-the-art algorithms over 21 data sets and the results proved the potential of the proposed binary approaches in searching the optimal feature subsets.

Introduction

Nowadays, several metaheuristic techniques have been effectively employed to solve various optimization problems from different domains. There exist a large number of metaheuristic techniques that effectively mimics the behavior of humans, insects, birds and animals. The most widely utilized metaheuristic techniques are Particle Swarm Optimization (PSO) (Eberhart and Kennedy 1995a), Genetic Algorithm (GA) (Goldberg and Holland 1988), Flower pollination Algorithm (FPA) (Kalra and Arora 2016; Yang, Karamanoglu, and He 2014), Firefly Algorithm (FA) (Arora and Singh 2014; Yang 2010a), Artificial Bee Colony (ABC) (Arora and Singh 2017a; Karaboga and Basturk 2007) and Grey Wolf Optimization (GWO) (Joshi and

CONTACT Sankalap Arora sankalap.arora@gmail.com; Manik Sharma manik10143@davuniversity.org DAV University, Jalandhar, India; Priyanka Anand priyanka.sep3@gmail.com Lovely Professional University, Jalandhar, India

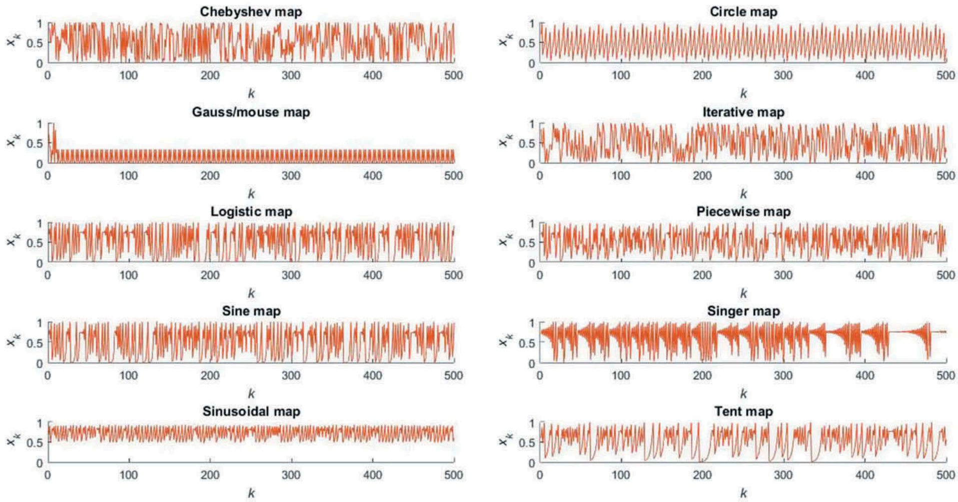

Figure 1. Visualization of chaotic maps.

Arora 2017; Mirjalili, Mirjalili, and Lewis 2014). Additionally, some of the recent nature inspired metaheuristic techniques are Whale Optimization Algorithm (WOA) (Mirjalili and Lewis 2016), Sine Cosine Algorithm (SCA) (Mirjalili 2016b), Ant Lion Optimizer (ALO) (Mirjalili 2015), Butterfly Optimization Algorithm (BOA) (Arora and Singh 2018) and more recently Interior Search Algorithm (ISA) (Gandomi 2014). For the optimization of parameters of every metaheuristic algorithm, each one has its own deficiency such as low classification accuracy, poor generalization ability, improper balance in exploration and exploitation, slow convergence speed and entrapment in local optima. Furthermore, metaheuristic algorithms need some randomness to proceed. In order to surmount all these problems, different techniques have been utilized to enhance the performance of metaheuristic algorithms in the literature. In recent years, the application of chaos theory to optimization algorithm in order to improve both exploration and exploitation has attracted more attention. The different properties of chaos like quasi-stochastic, ergodicity and sensitivity against initial conditions assist in improving the performance of different metaheuristics techniques. Chaos theory has been hybridized in various algorithms such as PSO (Alatas, Akin, and Ozer 2009), harmony search (Alatas, 2010), flower pollination algorithm (Arora and Anand 2017), artificial immune system (Jordehi 2015), FA (Wang, Li, and Ren 2010), WOA sayednew (Sayed, Darwish, and Hassanien, 2018) BOA (Arora and Singh 2017b), Salp Swarm Algorithm (SSA) (Sayed, Khoriba, and Haggag 2018) and grasshopper optimization algorithm (Arora and Anand 2018a).

Data mining is a prominent research area that combines traditional data analysis techniques with emerging computational algorithms to assist in collecting heterogeneous data from distinct sources, transform it into valuable information and utilize it in designing effective business strategies for an enterprise (Han, Pei, and Kamber 2011). It has been widely used in several classifications, clustering, association and regression-based real-life problems of different domains. Feature selection is one of the challenging tasks of data quality which assists in selecting optimal data set's features that boost the performance of classifiers. In other words, it is an extraction process which eradicates the irrelevant and redundant elements for the better understanding of data sets. The attributes that have a linear relationship with one another are called redundant attributes. The inclusion of these attributes seems to be unreasonable, as one can extract the complete information by merely incorporating one of these redundant attributes (Guyon and Elisseeff 2003). Nowadays, feature selection becomes mandatory as it is difficult to mine and transform the momentous volume of data into valuable insights.

The feature selection approaches fall into two categories: filter based methods and wrapper based methods (Balasaraswathi, Sugumaran, and Hamid 2017). The filter-based methods utilize statistical data dependency techniques to find the subset of features. The wrapper based methods employ machine learning algorithms to look for a near-optimal solution from an exponential set of feasible solutions. In filter-based methods, the selection process is entirely independent of other data mining tasks (Liu and Yu 2005). For example, one may be interested in extracting all those features where the pairwise association is too high. However, in wrapper based approach, the selection process is based upon the data mining techniques. The filter-based methods are independent of the classifier and are relatively fast but these methods do not reflect the relevance of the various dimensions while deciding the subset of the features. These techniques use methods such as information gain Yang and Pedersen (1997), principal component analysis (Han, Pei, and Kamber 2011), mutual information (Peng, Long, and Ding 2005) and employ measures such as the distance between the dimensions, correlation between the dimensions and consistency. The interaction of features and classifier is lacking in these methods which is a major limitation. Another drawback is that these methods cannot work with redundant information. On the other hand, the wrapper based techniques use classifier during their search for the optimal subset by employing the classifier's prediction accuracy. The researchers have found that the wrapper based methods obtain better results than the filter based methods (Zawbaa, Emary, and Grosan 2016). The reason for the better performance of these techniques is the use of machine learning technique (classifier) in the evaluation module.

The search space of the feature selection problems has been exponentially increased due to the rapid growth of data and the level of complexity associated with it. In conventional optimization techniques, a comprehensive list of all possible subsets is explored which makes the use of conventional approach too intractable and cost intensive for big data sets (Kohavi and John 1997). Further, due to promising computational intelligence techniques, several metaheuristic algorithms are able to solve these problems effectively whereas the conventional techniques are ineffective to provide the optimal solution (Yang 2010b). In the past, various attempts have been made to employ metaheuristic algorithms in order to solve feature selection problems. For instance, Hedar et al. utilized rough set theory for attribute reduction (Hedar, Wang, and Fukushima 2008), while Bello et al. used two-step PSO to solve the feature selection problem (Bello et al. 2007). Different initialization techniques in the PSO are also used to search feature subset space for selecting (sub)optimal feature set (Xue, Zhang, and Browne 2014). Many variants of GA have also been employed to select optimal features in different fields, such as Kabir et al. used local search method based on hybrid GA (Kabir, Shahjahan, and Murase 2011) whereas Chen et al. used GA based on the chaos theory in the field of text categorization (Chen et al. 2013). Attribute reduction algorithm using Record-to-Record Travel (RRT) algorithm has been utilized using rough set theory to evaluate the quality of the obtained solution (Mafarja and Abdullah 2013). A fuzzy-based RRT has also been employed for solving rough set attribute reduction problems (Mafarja and Abdullah 2015).

A modified GWO method in which a stochastic crossover is used to find the position of the grey wolf has been proposed as a solution for feature selection problem (Emary et al., 2016b) whereas in another research work, GWO along with kernel extreme learning machine has been used to find the optimal feature subset in the field of medical diagnosis (Li et al. 2017). The binary versions of ant lion optimizer (Emary et al., 2016a) and flower pollination based approaches (Sayed, Nabil, and Badr 2016) have been used to solve feature selection problem. In (Gu, Cheng, and Jin 2018), the feature selection problem has been solved by using a variant of the PSO known as competitive swarm optimizer. Moreover, hybrid algorithms have also been utilized for feature selection such as hybridization of differential evolution and artificial bee colony has been done (Zorarpacı & Özel, 2016) whereas hybrid algorithms based on the whale optimization algorithm and simulated annealing are also employed for feature selection (Mafarja and Mirjalili 2017). The main motivation behind these hybrid algorithms is to employ the capabilities of different algorithms for exploration and exploitation simultaneously.

Recently, chaotic maps have been employed in ant lion optimization algorithm with the aim to reduce the dimensions in high-dimensional data sets (Zawbaa, Emary, and Grosan 2016). Another feature selection method was proposed in (Mafarja and Mirjalili 2018) which has shown the use of whale optimization algorithm for feature selection based on the wrapper technique. Sayed et al. introduced the use of chaotic maps in crow search algorithm to improve the performance of the simple crow search for feature selection (Sayed, Hassanien, and Azar 2017). More recently, an approach based on the grasshopper optimization algorithm (Saremi, Mirjalili, and Lewis 2017) and binary butterfly optimization-based feature selection approaches (Arora and Anand 2019) have been successfully employed as a solution for selecting the optimal feature set.

In spite of the above-mentioned techniques which have been proposed for feature selection, still, there are many challenges unresolved such as higher error rate in case of high-dimensional data sets, higher execution time and selection of irrelevant features in the solution subset. Additionally, the No Free Lunch (NFL) theorem indicates that there is no single optimization algorithm which is adequate to provide solution for all optimization algorithms (Wolpert and Macready 1997). Therefore, this suggests that the present randomization based feature selection techniques can also suffer from degraded performance while solving some problems which influenced our attempts to analyze the efficiencies of the currently proposed Chaotic Interior Search Algorithm (CISA) (Mafarja and Mirjalili 2017, 2018). To tackle these issues, this study aims to propose a CISA for global optimization and feature selection approach. The remainder of the paper is arranged as follows: Section 2 presents the background information regarding ISA and chaotic maps. In Section 3, the improved Interior Search Algorithm (CISA) is presented. Section 4 discusses the experimental results on global benchmark problems as well as feature selection problems. Finally, conclusions and future work are given in Section 5.

Preliminaries

Interior Search Algorithm

Inspiration Analysis

Interior Search Algorithm (ISA) has been proposed to solve global optimization problems by Amir H. Gandomi (Gandomi 2014). The main inspiration of ISA is the beautification of objects and mirrors, specifically ISA deals with two groups called mirror and composition. All the available objects are initially decomposed into these two groups, i.e., mirror and composition groups, which deal with the optimal placement of the mirrors and the objects, respectively. In the composition group, the element's composition can be altered to get a more attractive view. Finally, the mirrors are placed optimally in-between the objects to improve its looks and the main aim is to find the fittest element for the placement of mirror. One of the best aspects of ISA is that it has only one parameter to control. ISA has been employed to solve engineering optimization problems (Gandomi and Roke 2014), water management (Moravej and Hosseini-Moghari 2016), digital differentiator design (Kumar et al. 2015) and automotive control problems (Yıldız, 2017).

Mathematical Model of ISA

The working principle of ISA is based upon esthetic mechanism used in interior design and decorations. The methodology of ISA has been explained in six steps which are given below:

(1) The algorithm starts with the random generation of element location where the location of elements is restricted to Lower Bound (LB) and Upper Bound (UB). The fitness of each random location is computed and on the basis of minimization or maximization problem, the fittest element is selected as global best and is denoted as x^j_{Gbest}.
(2) Further, the remaining elements are randomly partitioned into two distinct groups, i.e., composition and mirror. This division is based upon unique parameter (α) along with the following rules:

if $r_1 \leq \alpha$ **then**
element $\in$ composition
else
element $\in$ mirror
end if

Here r_1 is a random value in the range $(0, 1)$ which is associated with each element. The value of α should be chosen very carefully to avoid biased approach.

(3) Afterward, the composition of each element that lies in the composition group should be randomly changed using:

$$x^j_i = LB^j + (UB^j - LB^j) \times r_2 \tag{1}$$

Here, i and j represent the element position and iteration, respectively. LB^j and UB^j are lower and upper bound values of j-th iteration, and r_2 is the value generated randomly in the range between 0 and 1.

(4) For the optimal placement of the mirrors, first of all, a mirror object is randomly placed in between the global best element and all other elements of the problem. The position of the mirror for the i-th object and j-th iteration is given by:

$$x^j_{mir,i} = r_3 \times x^{j-1}_i + (1 - r_3) \times x^j_{Gbest} \tag{2}$$

where *mir* represents mirror object whereas r_3 represents randomly generated variable, and its value lies between 0 and 1. Additionally, the image location can be computed as follows:

$$x_i^j = 2x_{mir,i}^j - x_i^{j-1} \tag{3}$$

(5) The value of global best can be altered using random walk as mentioned below:

$$x_{Gbest}^j = x_{Gbest}^{j-1} + r_n \times \lambda \tag{4}$$

where r_n represents a normally distributed random numbers vector. The dimension of r_n and x should be same. Here, λ denotes scaling factor and its value is significantly affected by the size of search space of the problem. The real value of λ can be computed as:

$$\lambda = 0.01 \times (UB - LB) \tag{5}$$

(6) Finally, the fitness value of composition and mirror elements are calculated. Based upon the improvement, each location should be updated. The minimization problem can be formulated as

$$x_i^j = \begin{cases} x_i^j & f(x_i^j) < x_i^{j-1} \\ x_i^{j-1} & else \end{cases} \tag{6}$$

The pseudo-code of ISA is given in Algorithm 1.

Algorithm 1 Pseudo-code for the ISA.

Input: the fitness function, the upper and lower bound, number of elements, number of iterations

Output: the best solution

Initialize the elements
while stopping condition is not met do
    Find x^j_{Gbest}
    for each element do
        if x_{Gbest}
        else if $r_1 \leq ar_1 \leq a$
        else
        end if
        check for the boundary constraints
    end for

for each element **do**
Compute $f\left(x_i^j\right)$

end for
end while

Chaotic Maps

Generally, chaos can be defined as deterministic and arbitrary like strategy observed in a dynamic non-linear system which is bounded and non-converging. In mathematical terms, chaos is the arbitrariness of a basic dynamic deterministic framework and the chaotic system might be considered as a source of randomness Gandomi et al. (2013). The character of chaos is apparently random and unpredictable and it also possesses an element of regularity. Chaos utilizes chaotic variables instead of random variables Arora and Singh (2017b). Consequently, it may perform downright searches at higher speeds in comparison to the randomized searches which depend upon probabilities Kohli and Arora (2017). Simply some functions (chaotic maps) and few parameters (preliminary conditions) are required even for extremely long sequences Naanaa (2015). Additionally, a vast range of various sequences can be produced just by altering its preliminary condition. Besides, these sequences are deterministic and reproducible. Furthermore, it has an extremely sensitive dependence upon its preliminary conditions and parameter Lu et al. (2014).

A large number of chaotic maps are available in the field of optimization He et al. (2001); Arora and Singh (2017b). In this research work, 10 most extensively utilized chaotic maps have been used as shown in Figure 1 (Tavazoei and Haeri, 2007). The mathematical modulations of these employed chaotic maps are presented in the following subsections.

Chebyshev Map

Chebyshev map can be mathematically formulated as:

$$x_{k+1} = \cos(P.\cos^{-1} x_k) \tag{7}$$

Circle Map

This map was firstly defined by Andrey Colmogorov Lu et al. (2014); Zheng (1994). It is a one-dimensional map which is a member of dynamical systems on circle and is formulated as:

$$x_{k+1} = x_k + b - (\frac{P}{2\pi}) \sin(2\pi x_k) mod(1) \tag{8}$$

The values of P and b are set to 0.5 and 0.2 respectively in order to generate chaotic numbers between $(0, 1)$.

Gauss Map

Gauss map is formulated as:

$$x_{k+1} = \begin{cases} 0 & x_k = 0 \\ \frac{1}{x_k mod(1)} & otherwise \end{cases} \tag{9}$$

$$\frac{1}{x_k mod(1)} = \frac{1}{x_k} - [\frac{1}{x_k}] \tag{10}$$

This map also generates chaotic sequences in the range $(0, 1)$.

Iterative Map

The iterative chaotic map equation is defined as:

$$x_{k+1} = abs(\sin(\frac{P}{x_k})) \tag{11}$$

where P is the control parameter.

Logistic Map

This map is mathematically formulated as:

$$x_{k+1} = P.x_k(1 - x_k) \tag{12}$$

where the value of P is set to 4 in order to generate numbers in the range $(0, 1)$.

Piecewise Map

The family of piecewise map can be defined as:

$$x_{k+1} = \begin{cases} \frac{x_k}{P} & 0 \leq x_k \leq P \\ \frac{x_k - P}{0.5 - P} & P \leq x_k \leq 0.5 \\ \frac{1 - P - x_k}{0.5 - P} & 0.5 \leq x_k \leq 1 - P \\ \frac{1 - x_k}{P} & 1 - P \leq x_k \leq 1 \end{cases} \tag{13}$$

Here, the range of P, which is the control parameter is set in the range $(0, 0.5)$.

Sine Map

This map is formulated as:

$$x_{k+1} = \frac{a}{4}\sin(\pi x_k) \tag{14}$$

where P is the control parameter having values in the range $(0, 4)$.

Singer Map

This map is defined as:

$$x_{k+1} = P(7.86x_k - 23.31x_k^2 + 28.75x_k^3 - 13.302875x_k^4) \tag{15}$$

where P is the control parameter whose values lies in the range $(0.9, 1.08)$.

Sinusoidal Map

Sinusoidal map is characterized as:

$$x_{k+1} = P.x_k^2 \sin(\pi x_k) \tag{16}$$

Here P is the control parameter and a simplified equation of this map is utilized by using $P = 2.3$ and $x_0 = 0.7$ which can be formulated as:

$$x_{k+1} = \sin(\pi x_k) \tag{17}$$

Tent Map

The equation of the tent map can be characterized as:

$$x_{k+1} = \begin{cases} 2x_k & x_k < 0.5 \\ 2(1 - x_k) & x_k \geq 0.5 \end{cases} \tag{18}$$

The Proposed Chaotic Interior Search Algorithm (CISA)

In this section, a novel Interior Search Algorithm (ISA) is proposed where chaotic maps are utilized to replace the critical parameter α with chaotic variables which controls the intensification and diversification. In ISA, in order to choose between intensification and diversification, the value of α is compared with a random number. If the value of random number is greater than α, then Equation (1) is executed other Equations (2) and (3) is performed in order to achieve intensification and diversification, respectively. Basically, diversification focuses on finding new better solutions by inspecting the search space on a large scale, whereas intensification concentrates on exploiting the search space in the local region. It is clear that optimum balance of intensification and diversification phases will have considerable effect on the performance of ISA. However, in classical ISA a fixed value of α is utilized which hinders the ISA to achieve optimum results. Therefore, in this study a chaotic sequence generated from chaotic maps is utilized instead of a fixed value of α in order to properly balance the intensification and diversification with the intention to enhance the performance of ISA algorithm. Such a combination of chaos with ISA is defined as CISA. Equation (19) demonstrates the balancing of intensification and diversification using a chaotic variable.

$$\begin{array}{ll} \text{if} \quad r_1 \geq CV_t & \\ \quad \text{Apply local search as shown in Eq. (1)} & \\ \text{else} & \\ \quad \text{Apply global search as shown in Eq. (2) and Eq. (3)} & \\ \text{endif} & \end{array} \tag{19}$$

In this equation, in order to decide whether the algorithm will perform intensification and diversification, random number (r_1) of the *i-th* search agent is compared with chaotic value CV_t which is obtained from the respective chaotic map at *t-th* iteration. This integration of chaotic maps in generating new locations can improve the overall performance of ISA which will be demonstrated in the following section. The mathematical formulations of the chaotic maps are presented in the preceding section. The detailed description of the proposed chaotic version of the ISA algorithm is presented as follows:

Parameter Initialization

At the beginning of the search process, the search agents are assigned random positions in the search space. In case of global benchmark functions, the upper bound and lower bound respective to the used function is initialized whereas in case of feature selection problem, the upper bound and lower bound are set to 1 and 0 for the given data. The size of the population is fixed to 30 for global optimization problems and 7 for the feature selection problem. In case of feature selection problem, the size of the population is intentionally set low because of the complexity of the search space.

Fitness Function

The main goal of a fitness function is to assess every searching agent in terms of quality. In global benchmark problems, every function is a minimization problem, therefore the agent/solution which possesses the minimum value of fitness function is chosen as the best solution obtained so far whereas, in case of feature selection problem, the agent/solution which demonstrates lowest classification error rate and least number of features is considered as the optimal solution.

In this research work, every solution is represented as a single-dimensional vector in which the length of the vector depends on the number of features/attributes selected in a dataset. Each unit of the vector contains either 1 or 0, where the value 1 means that the corresponding feature/attribute is chosen whereas the value 0 means that the feature/attribute is not selected in the feature subset. Feature selection problem can be considered as a multi-objective optimization problem in which two opposing goals are to be accomplished; selecting a minimum number of features and achieving minimum classification error rate. In feature selection problem, that solution is considered best which contains the minimal number of features along with the lowest classification error rate. Every solution is assessed by the proposed fitness function which relies on KNN classifier Altman (1992) in order to calculate its classification error rate on the basis of selected features in the subset. Keeping in mind the end goal which is to find the balance between the number of attributes and classification performance, the fitness function in Equation (20) is employed in all the optimizers in order to evaluate the solutions.

$$Fitness = \alpha\gamma_R(D) + \beta\frac{|R|}{|N|} \tag{20}$$

where $\gamma_R(D)$ represents the classification quality of KNN classifier, $|R|$ represents the number of features selected in the subset and $|N|$ stands for the total number of features in the original dataset. The two parameters, i.e., α and β are correlated with classification performance and subset length, $\alpha \in [0, 1]$ and $\beta = (1 - \alpha)$ adopted from literature Emary et al. (2016a); Mafarja and Mirjalili (2017). In this study, CISA is employed to solve feature selection problem in a wrapper mode. The subset which comprises minimum classification error rate and lowest number of features is selected as an optimal feature subset describing the dataset.

Performance Metrics

To assess the proposed CISA, a variety of statistical measurements such as mean fitness, standard deviation, best fitness and worst fitness have been adopted on the global benchmark problems. In addition to this, in order to confirm whether the results of the proposed CISA are significantly different or not, a statistical test, i.e., Wilcoxon rank sum test is conducted. On the other hand, the following measures are utilized for feature selection problem:

Evaluation Criteria

While experimentation, for each run, five different measures, i.e., classification error rate, average selection size, mean, standard deviation, worst values and best values have been computed and compared. The brief details of these parameters are given below:

Classification Error Rate

It indicates the overall performance of the classifier. In other words, it determines that how many instances have been incorrectly classified by using selected features set. The classification error rate can be formulated as:

$$Average\ classification\ error\ rate = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N} Unmatch\left(C_i, L_i\right), \qquad (21)$$

where M is the number of times the optimization algorithm has been run, N denotes the number of test set points, C_i is output class label for particular data point i, L_i is reference class label for i, and *Unmatch* is the comparison function that gives outputs 0 when two labels are the same and outputs 1 when they are different.

Average Selection Size

It is defined as the size of selected features after each run divided by the total number of features, averaged over a number of iterations and is formulated as follows:

$$Average\ selection\ size = \frac{1}{M}\sum_{i=1}^{M} size(g^*) \qquad (22)$$

where *size* represents the number of features selected in the testing data set.

Statistical Mean

It represents the average performance of number of optimal solutions acquired during M different runs. It is formulated as:

$$Mean = \frac{1}{M}\sum_{i=1}^{M} g^{*i} \qquad (23)$$

where g^{*i} represents optimal solution during the *i-th* run.

Statistical Standard Deviation

It is a dispersion method which computes the variability of data. In other words, it represents the variation or dispersion from the central value of data series, i.e., how different outputs are concentrated around mean value or scattered from it. If the standard deviation is low, then the data points tend to be very close to the mean, whereas high standard deviation indicates that the data is spread out over an extensive range of values. It can be formulated as follows:

$$Std.\quad dev. = \sqrt{\frac{1}{M-1}\sum\left(g^{*i} - Mean\right)^2} \qquad (24)$$

where g^{*i} represents the best solution in the *i-th* run.

Statistical Best

It is the best (lowest) value out of all the fitness values acquired during the iterations and can be formulated as:

$$min_{1 \leq i \leq M}\{g^{*i}\} \tag{25}$$

Statistical Worst

It is the worst (highest) value out of all the fitness values acquired during the iterations as given in the following equation:

$$max_{1 \leq i \leq M}\{g^{*i}\} \tag{26}$$

Average Selection Size

It is defined as the number of selected features after each run divided by the total number of features, averaged over a number of iterations and is formulated as follows:

$$Mean = \frac{1}{M}\sum_{i=1}^{M} size(g^{*}) \tag{27}$$

where $size(x)$ represents the number of features selected in the testing dataset.

Nonparametric Statistical Test

In this study, two nonparametric statistical tests are employed to check whether the results of the proposed approaches are statistically superior to other algorithms. The first test is the Wilcoxon rank sum test which returns a parameter called p-value which is utilized to verify the significance level of two algorithms. In other words, Wilcoxon test determines that whether the two algorithms are statistically different or not (Derrac et al. 2011). Afterward, a second test, Friedman test is conducted which measure the performances of the algorithms based on the averages of the objective function values. Each algorithm was ranked according to their performance using an average Friedman rank competition-ranking scheme. Friedman test assigns a rank to all the values considered as a single group and consequently, ranks of every group are added. In competition ranking, algorithms are given the same rank if their performances are the same and algorithm which performs the best is assigned smaller rank, i.e., better the performance of the algorithm, smaller the rank.

Termination Criteria

Each search agent is evaluated using a fitness function and then its position is updated. This process is done iteratively until the maximum number of iterations or whenever the optimal solution is found. In this study, the termination criteria of stopping the optimization process is when the algorithm has reached the maximum number of iterations which is 500 for benchmark functions and 100 for the feature selection problem.

Experimental Results and Discussions

CISA for Global Optimization Problems

In this section, in order to evaluate the performance of the proposed CISA, various experiments on a diverse subset of global benchmark test functions is done. These functions have been chosen on the basis of dimensionality, modality and separability and moreover, these functions have been widely utilized by various authors in order to test the performance of metaheuristics Sayed, Khoriba, and Haggag (2018); Arora and Anand (2018a). As shown in Table 1, these functions can be divided into four categories. In the first category, unimodal functions (F4-F7) investigate the convergence speed of an algorithm. In the second category, multimodal functions (F1-F3 and F8-F13) assess the ability of an algorithm to find global optima when the number of local optima increases exponentially with problem dimension. In the next category, functions F1, F4 and F6-F9 are separable while the rest are non-separable. In the last category, step function (F6) which is discontinuous and has only one minima is used. The mathematical formulations, as well as properties of these benchmarks, are presented in Table 1 Arora and Anand (2018b).

The simulation results of classical ISA in comparison to different chaotic versions of ISA on different global benchmark problems are reported in Tables 2–5. The tables demonstrate the average, standard deviation, best and worst fitness values found by each algorithm. CISA1 to CISA10 utilize Chebyshev, Circle, Gauss/mouse, Iterative, Logistic, Piecewise, Sine, Singer, Sinusoidal, and Tent chaotic maps, respectively. It can be analyzed that the CISA10 (Tent chaotic map) outperforms other algorithms on nine out of 13 test functions whereas CISA1 (Chebyshev map), CISA3 (Gauss/mouse map), CISA5 (Logistic map) and CISA9 (Sinusoidal map) performed better on one benchmark function each. **Note that the best results are highlighted in bold.** Inspecting the results of chaotic ISA algorithms on the four unimodal functions in Table 2, CISA9 provides the best results on three unimodal test functions and competitive performance on fourth function which proves that Tent map improves the convergence rate of native ISA. The performance of CISA10 on multimodal function is significantly better than the classical ISA and other chaotic ISA variants which means that the local optima avoidance of CISA10 has been significantly improved by the Tent chaotic map.

Table 1. Global benchmark functions used in this study (Type: M: Multimodal, N: Non-Separable, U: Unimodal, S: Seperable, Range: Lower bound, Upper bound, f_{min}: Optima).

S. No.	Name	Formula	Type	Range	fmin
F1	Sphere	$f(x) = \sum_{i=1}^{n} x_i^2$	M, S	[−100,100]	0
F2	Schwefel 2.22	$f(x) = \sum_{i=1}^{n} \lvert x_i \rvert - \prod_{i=1}^{n} \lvert x_i \rvert$	M, N	[−10, 10]	0
F3	Schwefel 1.2	$f(x) = \sum_{i=1}^{n} \left(\sum_{j-1}^{i} x_j \right)^2$	M, N	[−100,100]	0
F4	Schwefel 2.21	$f(x) = \max(\lvert x_i \rvert, 1 \le i \le n)$	U, S	[−100,100]	0
F5	Rosenbrock	$f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	U, N	[−30,30]	0
F6	Step	$f(x) = \sum_{i=1}^{n} ([x_i + 0.5])^2$	U, S	[−100, 100]	0
F7	Quartic with noise	$f(x) = \sum_{i=1}^{n} i x_i^4 + \mathrm{rand}(0, 1)$	U, S	[−1.28, 1.28]	0
F8	Schwefel 2.26	$f(x) = \sum_{i=1}^{n} -x_i \sin\left(\sqrt{\lvert x_i \rvert}\right)$	M, S	[−500, 500]	−418.982
F9	Rastrigin	$f(x) = \sum_{i=1}^{n} (x_i^2 - 10 \cos(2\pi x_i) + 10)$	M, S	[−5.12, 5.12]	0
F10	Ackley	$f(x) = -20 \exp\left(-0.2 \times \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^{n} \cos(2\pi x_i)\right) + 20 + e$	M, N	[−32, 32]	0
F11	Griewank	$f(x) = \frac{1}{4000} \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	M, N	[−600, 600]	0

(*Continued*)

Table 1. (Continued).

S. No.	Name	Formula	Type	Range	fmin
F12	Penalty 1	$f(x) = \frac{\pi}{n}\{10\sin(\pi y_1) + \sum_{i=1}^{n-1}(y_1 - 1)^2[1 + 10\sin^2(\pi y_{i+1})] + (y_n - 1)^2\} + \sum_{i=1}^{n} u(x_i, 10, 100, 4)$ $y_i = 1 + \frac{x_i+1}{4},\ \ u(x_i, a, k, m) = \left\{\begin{matrix} k(x_i - a)^m & x_i > a \\ 0 - a & x_i < a \\ k(-x_i - a)^m & x_i < -a \end{matrix}\right\}$	M, N	[−50, 50]	0
F13	Penalty 2	$f(x) = 0.1\{\sin^2(3\pi x_i)\sum_{i=1}^{n-1}(x_i - 1)^2[1 + \sin^2(3\pi x_i + 1)] + (x_n - 1)^2[1 + sin^2(2\pi x_n)]\} + \sum_{i=1}^{n} u(x_i, 5, 100, 4)$	M, N	[−50, 50]	0

Table 2. Statistical mean fitness values of chaotic ISA algorithms on different benchmark functions.

F. No.	ISA	CISA1	CISA2	CISA3	CISA4	CISA5	CISA6	CISA7	CISA8	CISA9	CISA10
F1	4.84E+02	3.56E+02	4.68E+02	3.56E+02	4.10E+02	3.49E+02	3.56E+02	5.17E+02	4.41E+02	3.80E+02	**2.56E+02**
F2	5.85E+06	1.21E+08	1.90E+05	4.78E+06	3.11E+07	2.51E+09	2.50E+08	8.60E+06	1.76E+06	1.74E+06	**8.21E+03**
F3	1.39E+04	1.54E+04	1.75E+04	1.39E+04	1.46E+04	1.67E+04	1.84E+04	1.57E+04	1.66E+04	1.76E+04	**1.34E+04**
F4	5.38E+01	**4.55E+01**	5.13E+01	5.03E+01	4.97E+01	4.95E+01	4.76E+01	5.03E+01	4.81E+01	5.24E+01	5.16E+01
F5	1.00E+05	3.21E+04	7.79E+04	5.73E+04	6.57E+04	4.75E+04	4.76E+04	5.83E+04	6.34E+04	**2.40E+04**	4.57E+04
F6	5.99E+02	3.14E+02	3.34E+02	5.52E+02	4.10E+02	5.56E+02	4.93E+02	5.51E+02	4.29E+02	3.53E+02	**2.89E+02**
F7	2.03E+00	2.05E+00	1.45E+00	1.57E+00	1.60E+00	1.32E+00	1.60E+00	1.57E+00	1.72E+00	1.60E+00	**1.07E+00**
F8	−6.79E+03	−6.79E+03	−6.66E+03	**−7.24E+03**	−6.69E+03	−6.54E+03	−6.41E+03	−6.77E+03	−6.50E+03	−6.56E+03	−7.07E+03
F9	2.41E+02	2.14E+02	2.39E+02	2.11E+02	2.13E+02	2.26E+02	2.41E+02	2.16E+02	2.38E+02	2.13E+02	**1.81E+02**
F10	2.00E+01	1.97E+01	1.99E+01	1.95E+01	1.95E+01	1.96E+01	1.97E+01	1.95E+01	1.97E+01	1.97E+01	**1.89E+01**
F11	5.43E+00	4.63E+00	5.01E+00	4.59E+00	4.48E+00	4.99E+00	4.36E+00	4.82E+00	5.45E+00	4.57E+00	**4.30E+00**
F12	4.39E+03	4.76E+03	1.32E+03	1.31E+03	5.43E+03	**7.23E+02**	1.29E+03	1.42E+03	5.67E+03	8.27E+02	8.90E+02
F13	2.63E+04	7.13E+04	1.97E+04	1.40E+04	1.88E+04	1.03E+04	2.00E+04	1.64E+04	4.54E+04	2.91E+04	**6.61E+03**

Table 3 compares the standard deviation values of ISA with the proposed CISA algorithms. As it can be seen in these results, CISA10 performs better on six benchmark functions, maximum than any other chaotic ISA and classical ISA. Besides, high stability of the CISA with the Tent chaotic map in comparison to original ISA can be observed.

Table 4 demonstrates the best fitness values of ISA in comparison to the proposed CISA algorithms. It can be easily analyzed that CISA3 (Gauss/mouse map) and CISA10 (Tent map) have shown better performance by outperforming other algorithms on three benchmark functions each whereas native ISA, CISA1, CISA8 and CISA9 were not able to perform better on a single benchmark function. It is worthwhile to mention here that the results obtained by embedding different chaotic maps in the optimization process of ISA are better than the original ISA on most of the benchmark functions. The worst fitness values of ISA and proposed ISA are compared in Table 5. As it can be observed, CISA with the Tent map obtained better results on 6 out of the 13 benchmark functions whereas CISA with Gauss/mouse chaotic map and classical ISA outperformed other algorithms on two and one benchmark function, respectively.

A nonparametric statistical test, Wilcoxon's rank-sum test Derrac et al. (2011) is conducted in order is to determine whether the proposed CISA algorithms provide a significant improvement compared to the original ISA or not. The test was carried out using the results of the best algorithm in each benchmark function and compared with each of the other algorithms at 5% significance. Table 6 presents the p-values obtained by the test, where p-values less than 0.05 signify that the null hypothesis is rejected, i.e., there is a significant difference at a level of 5%. Here N/A means n′ot applicable,' meaning that the corresponding algorithm could not be compared with itself in the rank-sum test. The p-values in Table 6 verify that the improvement achieved by CISA10 is statistically significant on the majority of the benchmark functions. Afterward, a second statistical test, i.e., Friedman test is conducted and the results of this test are reported in Table 7. As it can be observed from CISA with Tent chaotic map attained the lowest rank which point out significant improvements of tent chaotic map over the results of the other chaotic maps. The results of Friedman test validated that the proposed CISA has strong global searching ability and stability.

In order to better compare the behavior of the proposed chaotic ISA algorithms on the global benchmark test functions in terms of convergence, Figure 2a is provided. As it can be observed in Figure 2b, CISA algorithms have shown competitive performance in comparison to original ISA. By carefully analyzing, it can be observed that the convergence of CISA with the Tent map is significantly faster than other algorithms in most of the functions, especially in F1, F2, F9, F10 and F13. This algorithm tends to outweigh other algorithms in the second quarter of optimization. From these results, it can be concluded that embedding chaotic variables during the optimization process of ISA can significantly boost the

Table 3. Statistical standard deviation values of chaotic ISA algorithms on different benchmark functions.

F. No.	ISA	CISA1	CISA2	CISA3	CISA4	CISA5	CISA6	CISA7	CISA8	CISA9	CISA10
F1	1.54E+02	1.18E+02	5.00E+02	1.58E+02	1.70E+02	9.71E+01	1.21E+02	2.69E+02	2.68E+02	2.67E+02	**5.29E+01**
F2	1.88E+07	3.87E+08	3.97E+05	1.51E+07	7.51E+07	6.29E+09	6.61E+08	2.19E+07	3.58E+06	3.01E+06	**1.24E+04**
F3	**3.16E+03**	5.55E+03	3.49E+03	6.46E+03	3.25E+03	6.61E+03	5.83E+03	4.36E+03	6.15E+03	6.97E+03	4.94E+03
F4	7.43E+00	8.84E+00	9.13E+00	9.40E+00	8.37E+00	8.93E+00	7.34E+00	**7.04E+00**	7.82E+00	7.25E+00	7.61E+00
F5	1.00E+05	**1.47E+04**	7.03E+04	4.55E+04	6.06E+04	2.14E+04	3.60E+04	7.87E+04	7.24E+04	2.27E+04	1.92E+04
F6	4.38E+02	8.51E+01	1.38E+02	2.42E+02	1.67E+02	3.36E+02	1.90E+02	2.36E+02	2.13E+02	1.67E+02	**3.88E+01**
F7	7.28E-01	1.02E+00	4.58E-01	6.85E-01	5.47E-01	4.00E-01	6.13E-01	5.94E-01	5.35E-01	4.88E-01	**2.11E-01**
F8	6.55E+02	**4.52E+02**	7.33E+02	6.25E+02	8.24E+02	7.94E+02	6.07E+02	7.33E+02	8.87E+02	6.20E+02	6.56E+02
F9	3.04E+01	4.37E+01	4.45E+01	4.30E+01	5.57E+01	4.40E+01	3.59E+01	5.44E+01	3.44E+01	4.19E+01	**2.28E+01**
F10	**5.12E-01**	7.33E-01	5.84E-01	7.98E-01	1.31E+00	6.65E-01	8.78E-01	1.27E+00	7.88E-01	7.88E-01	6.36E-01
F11	1.13E+00	1.90E+00	1.81E+00	1.52E+00	1.30E+00	1.42E+00	1.42E+00	**1.10E+00**	1.40E+00	1.70E+00	1.33E+00
F12	1.04E+04	1.69E+04	4.19E+03	2.23E+03	1.92E+04	1.86E+03	2.55E+03	3.55E+03	1.19E+04	**1.34E+03**	2.60E+03
F13	4.76E+04	1.23E+05	5.16E+04	2.20E+04	2.09E+04	1.43E+04	4.80E+04	3.33E+04	9.07E+04	3.26E+04	**1.13E+04**

The best results are highlighted in bold.

Table 4. Statistical best fitness values of chaotic ISA algorithms on different benchmark functions.

F. No.	ISA	CISA1	CISA2	CISA3	CISA4	CISA5	CISA6	CISA7	CISA8	CISA9	CISA10
F1	1.92E+02	1.95E+02	2.02E+02	1.82E+02	2.32E+02	2.17E+02	1.53E+02	2.23E+02	1.82E+02	1.72E+02	**1.48E+02**
F2	7.45E+01	2.69E+01	9.20E+01	**1.28E+01**	8.39E+01	7.68E+02	9.12E+01	2.55E+01	4.89E+01	8.12E+01	8.09E+01
F3	7.64E+03	7.30E+03	9.28E+03	**6.10E+03**	9.18E+03	6.47E+03	6.53E+03	8.30E+03	7.46E+03	9.51E+03	6.27E+03
F4	3.84E+01	3.01E+01	3.71E+01	**2.50E+01**	3.30E+01	3.20E+01	3.68E+01	3.93E+01	3.44E+01	3.84E+01	4.11E+01
F5	1.76E+04	9.05E+03	**3.26E+03**	1.12E+04	5.43E+03	2.06E+04	5.49E+03	5.82E+03	1.00E+04	6.04E+03	1.15E+04
F6	2.70E+02	1.93E+02	**1.63E+02**	2.54E+02	1.92E+02	2.57E+02	2.40E+02	2.62E+02	1.95E+02	1.89E+02	2.04E+02
F7	8.84E-01	8.19E-01	8.04E-01	5.87E-01	6.34E-01	5.67E-01	**4.93E-01**	6.46E-01	6.43E-01	6.73E-01	5.28E-01
F8	−8.10E+03	−7.65E+03	−8.30E+03	−8.26E+03	−8.03E+03	−7.54E+03	−7.71E+03	−8.10E+03	−7.84E+03	−7.36E+03	**−8.57E+03**
F9	1.74E+02	1.44E+02	1.45E+02	1.59E+02	**1.01E+02**	1.62E+02	1.87E+02	1.52E+02	1.69E+02	1.55E+02	1.29E+02
F10	1.88E+01	1.82E+01	1.87E+01	1.79E+01	**1.50E+01**	1.77E+01	1.72E+01	1.54E+01	1.76E+01	1.76E+01	1.73E+01
F11	3.87E+00	2.61E+00	2.51E+00	2.54E+00	3.08E+00	2.42E+00	2.48E+00	2.70E+00	2.52E+00	2.37E+00	**2.36E+00**
F12	4.99E+01	2.44E+01	3.65E+01	2.25E+01	3.55E+01	**1.49E+01**	3.16E+01	2.67E+01	3.84E+01	3.19E+01	4.14E+01
F13	9.13E+02	7.62E+01	6.18E+01	1.21E+02	7.76E+01	6.74E+01	3.17E+01	**2.30E+01**	7.33E+01	9.81E+02	3.29E+01

The best results are highlighted in bold.

Table 5. Statistical worst fitness values of chaotic ISA algorithms on different benchmark functions.

F. No.	ISA	CISA1	CISA2	CISA3	CISA4	CISA5	CISA6	CISA7	CISA8	CISA9	CISA10
F1	7.94E+02	5.23E+02	2.19E+03	8.25E+02	8.68E+02	5.07E+02	5.63E+02	1.19E+03	1.18E+03	1.12E+03	**3.21E+02**
F2	7.34E+07	1.51E+09	1.47E+06	5.83E+07	2.77E+08	2.25E+10	1.97E+09	8.24E+07	1.17E+07	1.01E+07	**4.00E+04**
F3	**1.95E+04**	2.38E+04	2.34E+04	2.65E+04	2.14E+04	3.30E+04	3.21E+04	2.73E+04	2.85E+04	3.21E+04	2.19E+04
F4	6.88E+01	6.20E+01	6.75E+01	6.31E+01	6.42E+01	6.50E+01	**6.06E+01**	6.42E+01	6.25E+01	6.71E+01	6.41E+01
F5	4.03E+05	**7.16E+04**	2.65E+05	1.67E+05	1.87E+05	9.20E+04	1.25E+05	3.18E+05	2.37E+05	7.87E+04	8.09E+04
F6	1.89E+03	5.00E+02	6.18E+02	9.37E+02	7.39E+02	1.56E+03	7.40E+02	1.12E+03	8.39E+02	8.27E+02	**3.36E+02**
F7	3.44E+00	5.29E+00	2.31E+00	3.07E+00	2.78E+00	1.99E+00	2.48E+00	2.55E+00	2.44E+00	2.58E+00	**1.35E+00**
F8	−6.05E+03	−6.11E+03	−5.63E+03	**−6.25E+03**	−5.36E+03	−4.69E+03	−5.75E+03	−5.39E+03	−4.71E+03	−5.22E+03	−5.95E+03
F9	2.81E+02	2.89E+02	3.12E+02	2.92E+02	3.00E+02	2.92E+02	3.16E+02	3.04E+02	3.04E+02	2.95E+02	**2.13E+02**
F10	2.05E+01	2.05E+01	2.06E+01	2.05E+01	2.04E+01	2.04E+01	2.04E+01	2.06E+01	2.06E+01	2.06E+01	**1.95E+01**
F11	7.83E+00	1.05E+01	7.85E+00	**6.98E+00**	7.72E+00	8.16E+00	8.09E+00	7.67E+00	7.95E+00	8.73E+00	7.59E+00
F12	4.10E+04	6.60E+04	1.64E+04	7.07E+03	7.46E+04	7.25E+03	8.14E+03	1.38E+04	4.19E+04	**4.73E+03**	1.03E+04
F13	1.90E+05	4.38E+05	2.01E+05	7.46E+04	6.74E+04	4.76E+04	1.88E+05	1.32E+05	3.37E+05	1.05E+05	**3.57E+04**

The best results are highlighted in bold.

Table 6. *p* -Values obtained from the rank-sum test on different benchmark functions.

F. No.	TSA	CTSA1	CTSA2	CTSA3	CTSA4	CTSA5	CTSA6	CTSA7	CTSA8	CTSA9	CTSA10
F1	1.47E-03	1.06E-02	4.35E-02	1.06E-02	3.14E-03	7.60E-03	1.06E-02	2.61E-03	6.41E-03	1.40E-01	N/A
F2	4.09E-02	2.61E-03	6.91E-02	1.25E-01	4.60E-01	1.21E-03	1.46E-02	8.20E-01	3.56E-02	2.11E-01	N/A
F3	9.10E-01	2.56E-01	2.31E-02	9.55E-01	3.94E-01	6.91E-02	2.31E-02	3.34E-01	1.40E-01	1.56E-01	N/A
F4	1.06E-02	N/A	9.95E-02	4.09E-02	2.81E-01	3.94E-01	9.55E-01	3.94E-01	3.07E-01	6.41E-03	4.09E-02
F5	4.51E-03	1.91E-01	8.98E-03	1.46E-02	4.68E-02	8.98E-03	1.71E-02	1.73E-01	2.31E-02	N/A	4.09E-02
F6	3.77E-03	4.27E-01	4.27E-01	8.05E-04	3.56E-02	1.47E-03	2.61E-03	1.79E-03	4.68E-02	3.07E-01	N/A
F7	9.87E-04	1.47E-03	3.09E-02	1.25E-02	6.41E-03	4.09E-02	2.31E-02	6.41E-03	3.77E-03	2.61E-03	N/A
F8	2.31E-02	3.09E-02	4.68E-02	N/A	1.12E-01	1.25E-02	3.77E-03	6.91E-02	1.99E-02	6.41E-03	3.94E-01
F9	8.05E-04	1.99E-02	1.21E-03	3.56E-02	5.35E-02	1.25E-02	6.55E-04	1.25E-01	8.05E-04	3.09E-02	N/A
F10	8.05E-04	8.98E-03	9.87E-04	6.91E-02	1.71E-02	2.31E-02	8.98E-03	1.25E-02	1.06E-02	1.06E-02	N/A
F11	3.56E-02	9.10E-01	1.73E-01	6.91E-01	9.55E-01	1.25E-01	7.33E-01	1.40E-01	3.56E-02	8.20E-01	N/A
F12	1.91E-01	4.27E-01	7.33E-01	4.60E-01	8.20E-01	N/A	8.20E-01	6.09E-01	3.63E-01	6.91E-01	1.00E±00
F13	6.09E-02	6.09E-02	2.33E-01	4.27E-01	3.09E-02	2.33E-01	4.27E-01	5.70E-01	5.35E-02	3.56E-02	N/A

The value where P > 0.05 are underlined.

Table 7. Friedman test results of chaotic ISA algorithms on different benchmark functions.

F. No.	ISA	CISA1	CISA2	CISA3	CISA4	CISA5	CISA6	CISA7	CISA8	CISA9	CISA10
F1	8.07	6.00	5.40	5.87	6.13	6.20	5.80	7.93	6.47	4.87	**3.27**
F2	5.87	7.53	5.20	4.93	5.93	8.80	7.07	4.53	5.93	5.87	**4.33**
F3	4.87	6.13	7.60	**4.53**	5.13	6.60	7.53	5.80	6.80	6.40	4.60
F4	7.13	**4.20**	6.53	6.60	6.07	5.87	4.87	6.40	5.20	6.67	6.47
F5	7.73	5.20	7.13	6.33	6.60	6.60	6.00	5.07	5.73	**3.07**	6.53
F6	7.00	3.87	4.40	7.80	6.00	7.80	7.00	8.33	5.67	4.53	**3.60**
F7	8.20	8.07	4.87	6.40	6.20	4.47	5.80	6.00	6.87	6.20	**2.93**
F8	5.93	5.93	6.53	**3.47**	6.00	6.60	7.40	6.00	6.60	7.07	4.47
F9	7.40	5.93	7.53	4.80	6.07	6.27	7.40	5.33	6.87	5.53	**2.87**
F10	7.53	6.33	6.60	5.53	5.67	5.87	7.20	5.80	6.50	6.50	**2.47**
F11	7.87	**5.07**	6.07	5.60	5.13	6.40	5.27	6.47	7.87	5.20	5.07
F12	7.67	**5.53**	4.93	7.20	5.40	5.67	5.40	5.67	6.87	5.93	5.73
F13	6.60	7.20	4.80	5.67	6.73	5.47	5.40	5.53	6.40	7.80	**4.40**
Total	91.87	77.00	77.60	74.73	77.07	82.60	82.13	78.87	83.77	75.63	**56.73**

The best results are highlighted in bold.

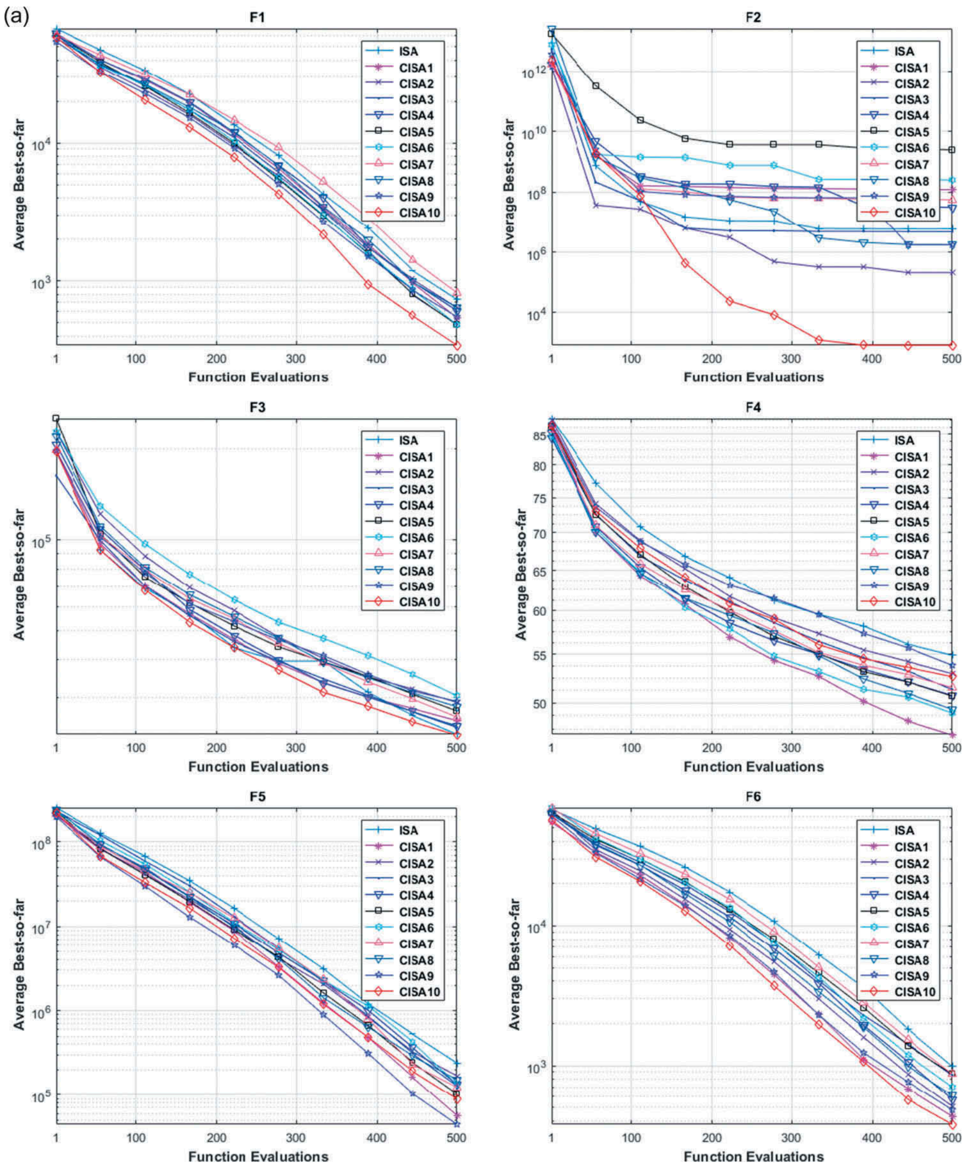

Figure 2. Convergence curves of chaotic ISA algorithms on different test functions.

performance of the original ISA. On the basis of these results, the Tent chaotic map is selected as the optimal map. Therefore, in the next section, this particular map will be further investigated in a more detailed manner.

CISA for Feature Selection

Here, 21 distinct data sets from UCI repository have been used for experimentation as shown in Table 8. These data sets have been selected so that

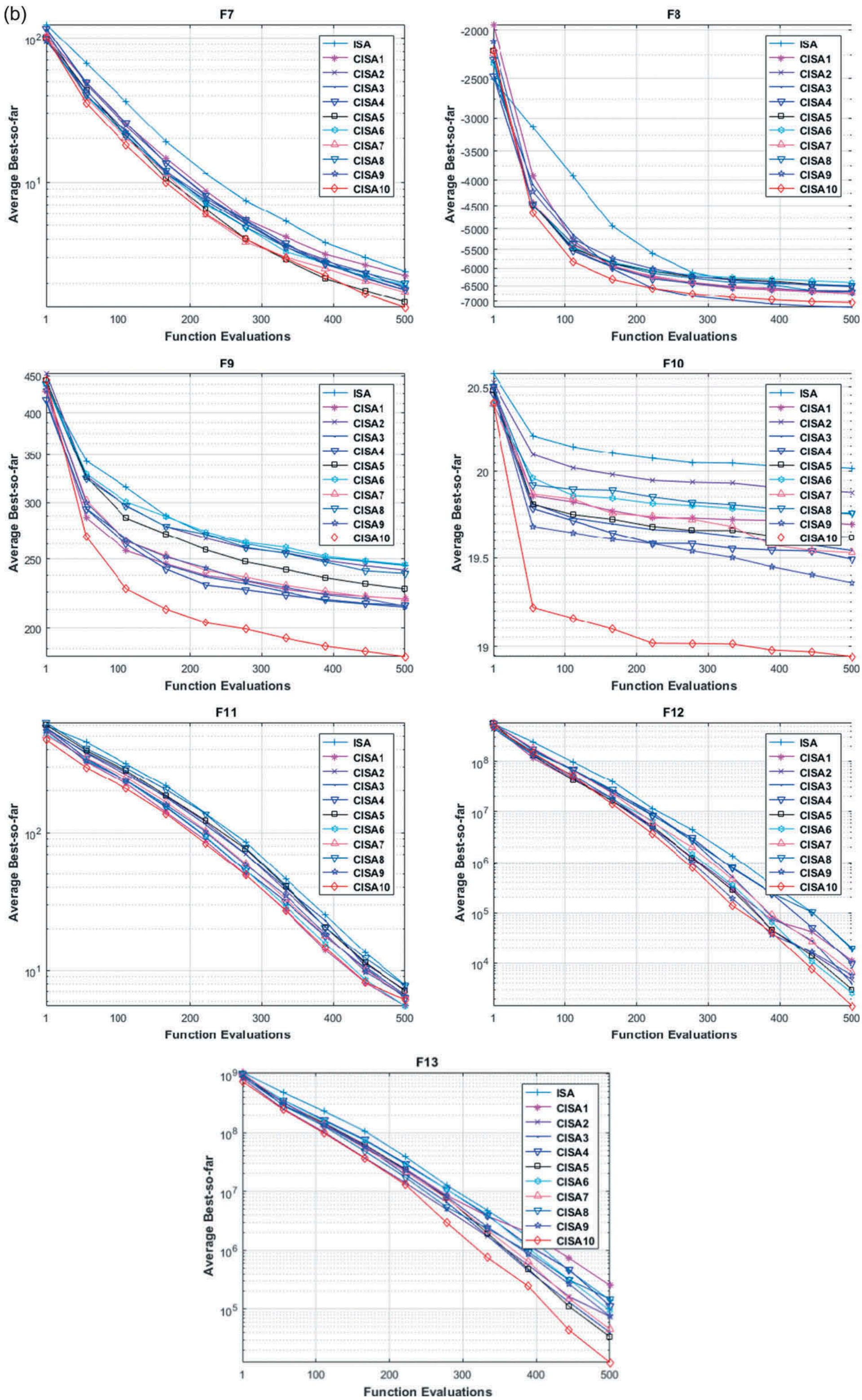

Figure 2. (Continued).

Table 8. List of data sets used in the experiments.

S. No.	Name	No. of features	No. of samples
1	Breastcancer	9	699
2	BreastEW	30	569
3	Clean1	166	476
4	Clean2	166	6598
5	CongressEW	16	435
6	Exactly	13	1000
7	Exactly2	13	1000
8	HeartEW	13	270
9	IonosphereEW	34	351
10	KrvskpEW	36	3196
11	Lymphography	18	148
12	M-of-n	13	1000
13	PenglungEW	325	73
14	Semeion	265	1593
15	SonarEW	60	208
16	SpectEW	22	267
17	Tic-tac-toe	9	958
18	Vote	16	300
19	WaveformEW	40	5000
20	WineEW	13	178
21	Zoo	16	101

they represent various number of features and tuples on which the proposed approach needs to be tested (Emary, Zawbaa, and Hassanien 2016a). Interestingly, the selected data sets have huge search space so that the testing of the optimization algorithm can be performed appropriately. Each data set is divided in a way as done in cross-validation methods (Franklin 2005).

In order to carry-out K-cross-validation, testing and validation are executed using $k-1$ folds, where the *k-th* fold is utilized for testing and each data set is evaluated for $K \times M$ times. Each data set is divided into three fractions, i.e., training, validation and testing. Using the training fraction, the classifier is trained and consequently, the validation fraction assesses the performance of the classifier. In the end, using the last fraction, i.e., testing is employed for evaluation of the features selected. During the training process, every element/search agent of the ISA is moved to select a subset of features. In order to validate the proposed CISA, its performance is compared with five state-of-the-art feature selection methods namely GA (Holland and Goldberg 1989), PSO (Eberhart and Kennedy 1995b), ALO (Emary, Zawbaa, and Hassanien 2016a), Dragonfly Algorithm (DA) (Mirjalili 2016a) and SSA (Mirjalili et al. 2017) and classical ISA. There are several parameters that should be initialized before employing these algorithms for feature selection. In this study, the number of search agents are fixed to 7 whereas the number of iterations is set to 100 and the results are averaged over 30 runs. For GA, crossover is set to 0.9, mutation ratio in set to 0.1 and roulette wheel is used as the selection mechanism. In PSO, acceleration constants are set in the range $[2, 2]$ whereas the inertia weight is set in the

Table 9. Classification accuracy of the proposed CISA vs other metaheuristics feature selection algorithms on different datasets.

S. No	Data set	ALO	GA	PSO	DA	SSA	ISA	CISA10
1	Breast Cancer	0.0409	0.0403	0.0391	0.0374	0.0400	0.0450	**0.0326**
2	BreastEW	0.0608	0.0512	0.0622	0.0615	0.0653	0.0732	**0.0470**
3	Clean1	0.1535	0.1303	0.1451	0.1459	0.1569	0.1501	**0.1176**
4	Clean2	0.0504	0.0577	0.0535	0.0513	0.0538	0.0524	**0.0429**
5	CongressEW	0.0630	0.0587	0.0765	0.0682	0.0679	0.0706	**0.0450**
6	Exactly	0.2939	0.2694	0.2529	**0.2519**	0.2909	0.3209	0.3184
7	Exactly2	0.3020	0.3060	0.3041	0.2993	0.3015	0.3179	**0.2548**
8	HeartEW	0.2227	0.2133	0.2212	0.2227	0.2405	0.2484	**0.1956**
9	IonosphereEW	0.1405	0.1062	0.1197	0.1292	0.1110	0.1299	**0.0886**
10	KrvskpEW	0.0994	0.0785	0.0800	**0.0731**	0.1071	0.1660	0.1100
11	Lymphography	0.2137	0.1836	0.2094	0.2207	0.2264	0.2542	**0.1568**
12	M-of-n	0.1816	0.2012	**0.1575**	0.1707	0.1639	0.2355	0.1728
13	PenglungEW	0.1928	0.3279	0.1860	0.1732	0.1669	0.1846	**0.0865**
14	semeion	0.0286	0.0243	0.0324	0.0279	0.0351	0.0330	**0.0236**
15	SonarEW	0.1513	**0.1250**	0.1333	0.1494	0.1551	0.1590	0.1519
16	SpectEW	0.2119	0.1903	0.2159	0.2000	0.2174	0.2199	**0.1507**
17	Tic-tac-toe	0.2413	0.2391	0.2482	0.2436	0.2443	0.2895	**0.2322**
18	Vote	0.0742	0.0667	0.0742	0.0773	0.0804	0.0778	**0.0307**
19	WaveformEW	0.2934	0.3079	**0.2808**	0.2846	0.2909	0.3141	0.2836
20	WineEW	0.0457	0.0464	0.0479	0.0449	0.0494	0.0584	**0.0315**
21	Zoo	0.0784	0.0706	0.0549	0.0641	0.0524	0.0930	**0.0314**
	Average	0.1495	0.1474	0.1426	0.1427	0.1484	0.1664	**0.1240**

The best results are highlighted in bold.

range $[0.9, 0.6]$. For ALO, the mutation rate is set in the range $[0, 0.9]$. These parameter values have been selected according to the values used in the literature (Emary, Zawbaa, and Hassanien 2016b). In this study, the same fitness function with similar population size, search boundary, number of dimensions and the same number of iterations are used in order to make a fair comparison (Emary, Zawbaa, and Hassanien 2016a).

Table 9 outlines the classification error rate for the proposed CISA in comparison to the state-of-the-art feature selection optimizers. It can be analyzed from these results that the performance of CISA is superior to ALO, GA, PSO, DA and SSA on 16 data sets in terms of classification error rate whereas DA and PSO performed better than other algorithms on two data sets each. Furthermore, it is worthwhile to mention here that the conventional ISA does not perform better than CISA over all the data sets used in this study which clearly shows the strength of the proposed chaotic ISA approach. Overall, the CISA showed lowest classification error rate, i.e., 0.1240 whereas PSO and DA showed 0.1426 and 0.1427 error rate, respectively. GA came at fourth place by demonstrating 0.1474 error rate whereas SSA has shown 0.1484 error rate.

The results of average selection size of the proposed CISA and other feature selection approaches are outlined in Table 10. A similar trend can be seen in average feature subset length where CISA exhibits much better performance by selecting less number of attributes in comparison to other approaches employed in this study. As per the results reported in this table,

Table 10. Average feature length of the proposed CISA vs other metaheuristics feature selection algorithms on different datasets.

S. No.	Data set	ALO	GA	PSO	DA	SSA	ISA	CISA10
1	Breast Cancer	7	6.1	5.7	6.27	7.2	5.27	**5.2**
2	BreastEW	24.27	12.2	18.33	20	18.27	16.33	**11.73**
3	Clean1	132	98.9	104.93	109.67	94.87	82.4	**82.09**
4	Clean2	95	94.1	109.4	100.4	90.4	86.6	**82**
5	CongressEW	9.87	7.1	10.8	10.87	8.4	8.2	**6.82**
6	Exactly	12.87	8.1	9	10.53	12.8	7.46	**7.18**
7	Exactly2	8.4	7.1	9.4	8.67	6.27	6.33	**5.73**
8	HeartEW	10.4	6.6	9.07	9.6	7.47	7.47	**6.45**
9	IonosphereEW	20.13	13.5	19.2	18	19.67	16.33	**13.18**
10	KrvskpEW	35.8	18	25.6	28.6	36	18.53	**17.8**
11	Lymphography	13.33	8.9	11.73	12.53	12.2	9.47	**8.45**
12	M-of-n	11.27	7.68	10.87	12.13	12.33	7.47	**7.36**
13	PenglungEW	172.07	153	183.33	175.2	162.33	160.67	**151.82**
14	semeion	187.8	149.4	171.6	193	161.8	136	**134.2**
15	SonarEW	48	30.3	37.6	40.6	34.13	31.53	**28.91**
16	SpectEW	13.87	7	12.07	14.67	11.33	11.33	**6.91**
17	Tic-tac-toe	8.8	5.8	6.73	7.2	8.07	5.17	**5.09**
18	Vote	8.4	5.8	9.33	8.87	8.53	8.8	**5.64**
19	WaveformEW	39.6	30.4	35.8	36	40	**20.93**	21.8
20	WineEW	11.07	6.73	10.07	9.53	9.07	7.4	**6.54**
21	Zoo	11.67	**5.35**	11.8	11.47	11.93	8.4	6.09

The best results are highlighted in bold.

CISA performed better on all data sets except WaveformEW and Zoo data set, where better performance is demonstrated by ISA and GA, respectively. The strength of the proposed CISA lies in the balanced exploration and exploitation capability which allows it to eliminate redundant attributes and then search the high-performance regions of the feature space intensively. This indicates that Tent map has improved the performance of the native ISA in terms of an average number of features/attributes selected.

The results of statistical measures (mean, standard deviation, best and worst fitness values) obtained on the different runs of the algorithms on all the data sets are presented in Tables 11–14. As shown in Table 11, CISA outperformed ISA on all the data sets. Additionally, it can be observed from Table 11 that CISA outperformed ALO, GA, GWO and PSO in mean fitness measure on 17 data sets, PSO performed better on two datasets whereas GA and DA performed better on one dataset each. It can be observed from Tables 12 and 13 that CISA has shown superior performance overall on most of the datasets in best and worst fitness measure. A summary of the statistical standard deviation measure results obtained for all the data sets is shown in Table 14. It can be analyzed from this table that GA performed better on 12 data sets whereas CISA performed superior on 6 data sets which shows the competency of CISA in comparison to the other feature selection approaches (ALO, PSO, DA and SSA).

Table 11. Statistical mean fitness measure of the proposed CISA vs other metaheuristics feature selection algorithms on different datasets.

S. No.	Data set	ALO	GA	PSO	DA	SSA	ISA	CISA10
1	Breast Cancer	4.80E-02	4.60E-02	4.50E-02	4.10E-02	4.80E-02	5.00E-02	**3.60E-02**
2	BreastEW	6.80E-02	5.50E-02	6.80E-02	6.80E-02	7.10E-02	7.80E-02	**4.90E-02**
3	Clean1	1.60E-01	1.34E-01	1.50E-01	1.51E-01	1.60E-01	1.54E-01	**1.18E-01**
4	Clean2	5.50E-02	6.20E-02	6.00E-02	5.70E-02	5.80E-02	5.70E-02	**4.40E-02**
5	CongressEW	6.90E-02	6.30E-02	8.20E-02	7.40E-02	7.20E-02	7.50E-02	**4.50E-02**
6	Exactly	3.01E-01	2.70E-01	**2.57E-01**	2.57E-01	2.98E-01	3.23E-01	3.17E-01
7	Exactly2	3.05E-01	3.08E-01	3.08E-01	3.03E-01	3.03E-01	3.20E-01	**2.54E-01**
8	HeartEW	2.28E-01	2.16E-01	2.26E-01	2.28E-01	2.44E-01	2.52E-01	**1.96E-01**
9	IonosphereEW	1.45E-01	1.09E-01	1.24E-01	1.33E-01	1.16E-01	1.33E-01	**8.90E-02**
10	KrvskpEW	1.08E-01	8.30E-02	8.60E-02	**8.00E-02**	1.16E-01	1.69E-01	1.12E-01
11	Lymphography	2.19E-01	1.87E-01	2.14E-01	2.25E-01	2.31E-01	2.57E-01	**1.58E-01**
12	M-of-n	1.88E-01	2.05E-01	**1.64E-01**	1.78E-01	1.72E-01	2.39E-01	1.75E-01
13	PenglungEW	1.96E-01	1.29E-01	1.90E-01	1.77E-01	1.70E-01	1.88E-01	**8.60E-02**
14	semeion	3.50E-02	2.90E-02	3.90E-02	3.50E-02	4.00E-02	3.80E-02	**2.60E-02**
15	SonarEW	1.58E-01	**1.28E-01**	1.38E-01	1.55E-01	1.59E-01	1.63E-01	1.52E-01
16	SpectEW	2.16E-01	1.92E-01	2.19E-01	2.05E-01	2.20E-01	2.23E-01	**1.51E-01**
17	Tic-tac-toe	2.49E-01	2.43E-01	2.53E-01	2.49E-01	2.51E-01	2.92E-01	**2.35E-01**
18	Vote	7.90E-02	7.00E-02	7.90E-02	8.20E-02	8.50E-02	8.30E-02	**3.20E-02**
19	WaveformEW	3.00E-01	3.10E-01	2.87E-01	2.91E-01	2.98E-01	3.16E-01	**2.83E-01**
20	WineEW	5.40E-02	5.10E-02	5.50E-02	5.20E-02	5.60E-02	6.40E-02	**3.40E-02**
21	Zoo	8.50E-02	7.30E-02	6.20E-02	7.10E-02	5.90E-02	9.70E-02	**3.40E-02**

The best results are highlighted in bold.

Table 12. Statistical best fitness measure of the proposed CISA vs other metaheuristics feature selection algorithms on different datasets.

S. No.	Data set	ALO	GA	PSO	DA	SSA	ISA	CISA10
1	Breast Cancer	3.80E-02	4.00E-02	3.90E-02	3.10E-02	3.80E-02	**2.70E-02**	3.10E-02
2	BreastEW	5.60E-02	4.80E-02	4.90E-02	5.10E-02	5.90E-02	6.20E-02	**4.30E-02**
3	Clean1	1.18E-01	1.22E-01	1.00E-01	1.18E-01	1.22E-01	1.26E-01	**9.80E-02**
4	Clean2	4.90E-02	6.20E-02	5.80E-02	5.00E-02	5.40E-02	5.10E-02	**4.00E-02**
5	CongressEW	4.40E-02	5.40E-02	4.80E-02	**3.50E-02**	4.10E-02	4.50E-02	3.70E-02
6	Exactly	2.67E-01	**1.50E-02**	1.38E-01	1.55E-01	2.29E-01	1.17E-01	2.68E-01
7	Exactly2	2.52E-01	2.95E-01	2.75E-01	2.38E-01	2.70E-01	2.57E-01	**2.25E-01**
8	HeartEW	1.72E-01	2.02E-01	1.78E-01	**1.59E-01**	1.94E-01	1.89E-01	1.78E-01
9	IonosphereEW	1.11E-01	9.90E-02	8.10E-02	1.04E-01	7.80E-02	7.80E-02	**6.30E-02**
10	KrvskpEW	9.30E-02	6.30E-02	**5.20E-02**	6.20E-02	1.11E-01	6.10E-02	7.10E-02
11	Lymphography	1.65E-01	1.68E-01	1.79E-01	1.66E-01	1.69E-01	1.67E-01	**1.23E-01**
12	M-of-n	1.60E-01	1.40E-01	6.40E-02	1.57E-01	**3.50E-02**	1.40E-01	1.46E-01
13	PenglungEW	8.50E-02	1.37E-01	8.60E-02	**3.50E-02**	1.12E-01	5.80E-02	5.40E-02
14	Semeion	4.10E-02	3.30E-02	4.20E-02	4.00E-02	4.50E-02	2.90E-02	**2.20E-02**
15	SonarEW	1.28E-01	1.09E-01	9.10E-02	1.13E-01	1.29E-01	1.10E-01	**7.90E-02**
16	SpectEW	1.44E-01	1.70E-01	1.66E-01	1.42E-01	1.73E-01	**5.00E-03**	1.14E-01
17	Tic-tac-toe	2.13E-01	2.32E-01	2.04E-01	2.17E-01	2.13E-01	2.44E-01	**1.92E-01**
18	Vote	4.30E-02	6.10E-02	3.90E-02	5.10E-02	5.00E-02	3.70E-02	**2.70E-02**
19	WaveformEW	2.94E-01	3.12E-01	2.71E-01	2.78E-01	2.91E-01	2.76E-01	**2.66E-01**
20	WineEW	2.90E-02	3.80E-02	2.80E-02	3.10E-02	**1.60E-02**	2.50E-02	2.60E-02
21	Zoo	2.60E-02	6.10E-02	8.00E-03	7.00E-03	2.60E-02	4.40E-02	**1.00E-03**

The best results are highlighted in bold.

Table 15 presents the p-values of CISA compared to other meta-heuristic algorithms obtained using Wilcoxon's rank sum test. It can be easily observed from this table that the p-values obtained using the rank sum test

Table 13. Statistical worst fitness measure of the proposed CISA vs other metaheuristics feature selection algorithms on different datasets.

S. No.	Data set	ALO	GA	PSO	DA	SSA	ISA	CISA10
1	Breast Cancer	5.90E-02	5.10E-02	6.00E-02	5.90E-02	5.80E-02	6.60E-02	**4.20E-02**
2	BreastEW	8.30E-02	**6.30E-02**	7.80E-02	9.00E-02	8.80E-02	9.70E-02	5.50E-02
3	Clean1	1.93E-01	**1.43E-01**	1.86E-01	1.78E-01	2.00E-01	2.01E-01	1.44E-01
4	Clean2	6.00E-02	7.10E-02	6.10E-02	5.90E-02	6.40E-02	6.20E-02	**4.70E-02**
5	CongressEW	1.10E-01	8.30E-02	1.49E-01	1.07E-01	1.20E-01	1.38E-01	**5.20E-02**
6	Exactly	3.43E-01	3.78E-01	3.84E-01	**3.19E-01**	3.35E-01	4.08E-01	3.58E-01
7	Exactly2	3.55E-01	3.31E-01	3.35E-01	3.30E-01	3.63E-01	3.65E-01	**2.97E-01**
8	HeartEW	2.89E-01	2.61E-01	2.88E-01	2.84E-01	2.99E-01	3.07E-01	**2.10E-01**
9	IonosphereEW	1.68E-01	1.34E-01	1.63E-01	1.57E-01	1.57E-01	1.63E-01	**1.20E-01**
10	KrvskpEW	1.18E-01	1.50E-01	1.64E-01	**9.70E-02**	1.21E-01	2.37E-01	1.68E-01
11	Lymphography	2.51E-01	2.20E-01	2.76E-01	3.03E-01	2.99E-01	3.25E-01	**1.91E-01**
12	M-of-n	2.24E-01	2.88E-01	2.87E-01	**2.10E-01**	2.12E-01	3.04E-01	2.53E-01
13	PenglungEW	3.00E-01	1.90E-01	3.28E-01	3.26E-01	2.73E-01	3.53E-01	**1.35E-01**
14	Semeion	4.10E-02	3.30E-02	4.20E-02	4.00E-02	4.50E-02	4.60E-02	**3.10E-02**
15	SonarEW	2.16E-01	**1.56E-01**	1.87E-01	2.17E-01	2.14E-01	2.25E-01	2.11E-01
16	SpectEW	2.71E-01	2.18E-01	2.65E-01	2.52E-01	2.62E-01	3.53E-01	**1.80E-01**
17	Tic-tac-toe	2.75E-01	**2.55E-01**	3.31E-01	2.93E-01	2.93E-01	3.85E-01	2.57E-01
18	Vote	1.13E-01	8.80E-02	1.19E-01	1.24E-01	1.18E-01	1.12E-01	**4.20E-02**
19	WaveformEW	3.04E-01	3.19E-01	3.03E-01	**2.99E-01**	3.05E-01	3.42E-01	3.08E-01
20	WineEW	7.30E-02	8.20E-02	7.50E-02	8.60E-02	7.60E-02	8.50E-02	**4.80E-02**
21	Zoo	1.58E-01	1.01E-01	1.82E-01	1.25E-01	1.07E-01	1.40E-01	**6.30E-02**

The best results are highlighted in bold.

Table 14. Statistical standard deviation measure of the proposed CISA vs other metaheuristics feature selection algorithms on different datasets.

S. No.	Data set	ALO	GA	PSO	DA	SSA	ISA	CISA10
1	Breast Cancer	8.00E-03	5.00E-03	8.00E-03	1.10E-02	8.00E-03	1.00E-02	**5.00E-03**
2	BreastEW	7.00E-03	**3.00E-03**	6.00E-03	1.00E-02	5.00E-03	1.10E-02	5.00E-03
3	Clean1	1.70E-02	**8.00E-03**	2.60E-02	2.10E-02	2.00E-02	2.40E-02	2.30E-02
4	Clean2	2.30E-02	1.36E-01	6.70E-02	3.80E-02	2.50E-02	5.00E-03	**3.00E-03**
5	CongressEW	3.00E-02	1.10E-02	1.90E-02	2.20E-02	2.50E-02	2.60E-02	**6.00E-03**
6	Exactly	3.10E-02	**2.00E-02**	2.90E-02	2.80E-02	2.80E-02	6.80E-02	3.30E-02
7	Exactly2	1.80E-02	**1.20E-02**	2.30E-02	1.50E-02	2.10E-02	2.90E-02	3.20E-02
8	HeartEW	9.00E-03	3.80E-02	4.50E-02	1.40E-02	**5.00E-03**	3.00E-02	1.20E-02
9	IonosphereEW	2.60E-02	**1.60E-02**	2.80E-02	4.10E-02	4.20E-02	2.50E-02	2.20E-02
10	KrvskpEW	2.00E-02	5.40E-02	5.80E-02	**1.80E-02**	4.90E-02	5.90E-02	5.10E-02
11	Lymphography	2.50E-02	**1.30E-02**	2.90E-02	2.30E-02	2.10E-02	4.30E-02	2.70E-02
12	M-of-n	3.50E-02	**1.60E-02**	2.70E-02	3.00E-02	2.50E-02	4.80E-02	4.50E-02
13	PenglungEW	2.00E-02	**6.00E-03**	3.40E-02	2.30E-02	2.10E-02	8.50E-02	2.90E-02
14	Semeion	1.80E-02	8.00E-03	1.90E-02	1.90E-02	2.00E-02	5.00E-03	**4.00E-03**
15	SonarEW	4.00E-03	**3.00E-03**	1.20E-02	9.00E-03	6.00E-03	3.60E-02	4.70E-02
16	SpectEW	1.20E-02	**1.10E-02**	1.20E-02	1.30E-02	1.50E-02	1.20E-01	2.80E-02
17	Tic-tac-toe	3.50E-02	**1.50E-02**	5.00E-02	3.70E-02	2.80E-02	4.20E-02	2.60E-02
18	Vote	2.10E-02	9.00E-03	2.80E-02	1.80E-02	2.20E-02	1.70E-02	**6.00E-03**
19	WaveformEW	4.00E-03	4.00E-03	**1.00E-03**	3.00E-03	4.00E-03	1.60E-02	1.80E-02
20	WineEW	7.20E-02	1.80E-02	7.70E-02	7.70E-02	5.20E-02	1.80E-02	**9.00E-03**
21	Zoo	7.00E-03	**1.00E-03**	2.00E-03	3.00E-03	4.00E-03	3.30E-02	3.00E-02

The best results are highlighted in bold.

prove that the superiority of CISA is statistically significant. Additionally, the results also indicate that the results of CISA are statistically significant compared to the native ISA. Furthermore, Table 16 demonstrates the results

Table 15. *p*-Values of the Wilcoxon test of the proposed CISA vs. other algorithms($p \geq 0.05$ are underlined).

S. No.	Data set	ALO	GA	PSO	DA	SSA	ISA	CISA10
1	Breast Cancer	5.06E-03	4.69E-02	5.06E-03	5.06E-03	9.34E-03	1.25E-02	N/A
2	BreastEW	6.53E-04	8.05E-04	3.55E-02	6.55E-04	6.55E-04	6.55E-04	N/A
3	Clean1	1.46E-03	3.14E-03	6.08E-02	2.16E-03	1.79E-03	8.05E-04	N/A
4	Clean2	6.23E-04	6.23E-04	6.23E-04	6.23E-04	6.23E-04	6.23E-04	N/A
5	CongressEW	6.55E-04	2.61E-03	6.47E-04	1.21E-03	6.55E-04	9.85E-04	N/A
6	Exactly	4.09E-02	7.76E-01	N/A	3.34E-01	4.09E-02	1.99E-02	5.39E-03
7	Exactly2	6.55E-04	6.55E-04	6.47E-04	1.79E-03	6.55E-04	1.79E-03	N/A
8	HeartEW	4.51E-03	1.47E-03	5.34E-03	1.21E-03	6.41E-03	9.87E-04	N/A
9	IonosphereEW	8.03E-04	6.55E-04	1.05E-02	6.40E-03	3.77E-03	1.20E-03	N/A
10	KrvskpEW	4.31E-02	8.93E-01	8.93E-01	N/A	2.25E-01	2.25E-01	6.86E-01
11	Lymphography	8.05E-04	8.05E-04	6.35E-03	6.55E-04	8.05E-04	9.87E-04	N/A
12	M-of-n	1.40E-01	1.98E-01	N/A	6.91E-02	3.07E-01	2.16E-03	4.60E-01
13	PenglungEW	6.55E-04	9.87E-04	6.55E-04	6.55E-04	6.55E-04	1.47E-03	N/A
14	Semeion	4.31E-02	4.31E-02	4.31E-02	4.31E-02	4.31E-02	4.31E-02	N/A
15	SonarEW	8.05E-04	N/A	1.40E-01	3.77E-03	2.16E-03	7.60E-03	1.12E-01
16	SpectEW	2.61E-03	1.47E-03	1.77E-03	6.55E-04	6.55E-04	6.55E-04	N/A
17	Tic-tac-toe	3.06E-01	1.56E-01	5.32E-01	9.95E-02	2.11E-01	2.16E-03	N/A
18	Vote	6.55E-04	6.55E-04	6.34E-04	6.53E-04	6.53E-04	6.55E-04	N/A
19	WaveformEW	7.96E-02	3.45E-01	4.31E-02	2.25E-01	5.00E-01	4.31E-02	N/A
20	WineEW	6.45E-04	6.52E-04	1.78E-03	1.78E-03	1.47E-03	1.20E-03	N/A
21	Zoo	1.47E-03	6.31E-03	1.23E-03	2.67E-02	4.09E-02	6.53E-04	N/A

The value where $P > 0.05$ are underlined.

Table 16. Friedman test results of CISA vs other metaheuristics feature selection algorithms on different datasets.

S. No.	Data set	ALO	GA	PSO	DA	SSA	ISA	CISA10
1	Breast Cancer	4.70	3.90	4.35	4.80	3.75	5.00	1.50
2	BreastEW	4.60	4.80	1.80	5.13	4.40	5.87	1.40
3	Clean1	5.20	4.60	2.43	4.97	4.20	4.67	1.93
4	Clean2	2.80	3.80	6.60	4.80	5.00	4.00	1.00
5	CongressEW	4.23	4.77	3.43	4.23	5.27	4.80	1.27
6	Exactly	4.33	2.27	4.67	3.87	2.47	5.20	5.20
7	Exactly2	4.60	4.27	3.93	4.07	4.60	5.27	1.27
8	HeartEW	3.60	4.67	3.20	5.00	4.20	5.47	1.87
9	IonosphereEW	5.53	5.00	3.10	3.40	4.30	5.00	1.67
10	KrvskpEW	5.40	3.20	2.80	4.60	2.80	5.20	4.00
11	Lymphography	4.40	4.73	2.40	5.07	4.20	5.67	1.53
12	M-of-n	4.17	3.37	4.60	3.97	2.90	5.80	3.20
13	PenglungEW	4.80	4.73	4.00	4.27	4.60	4.47	1.13
14	Semeion	4.60	3.80	2.20	5.40	5.20	5.80	1.00
15	SonarEW	4.43	4.47	2.07	4.93	3.23	4.67	4.20
16	SpectEW	4.77	3.93	2.87	4.77	5.23	5.03	1.40
17	Tic-tac-toe	3.93	3.60	2.73	4.53	4.13	5.93	3.13
18	Vote	4.57	4.87	3.13	4.70	4.73	5.00	1.00
19	WaveformEW	4.40	3.00	6.40	3.60	2.40	5.80	2.40
20	WineEW	4.83	4.03	3.47	4.83	4.53	5.03	1.27
21	Zoo	4.70	4.17	4.70	3.60	3.27	5.77	1.80
	Total	94.60	85.97	74.88	94.53	85.42	109.43	**43.17**

obtained from Friedman's test. It can be analyzed from the results of this table that the proposed CISA obtained the best rank in comparison to ISA and feature selection approaches. This means that the results of CISA are significantly better than the other algorithms.

To sum up, all findings achieved in the computational study show that the proposed CISA is an efficient optimizer for feature selection. The underlying reason for the better performance of CISA is its very high exploration and ability to avoid local optima. The exploitation capability of the proposed CISA allows it to efficiently avoid a large number of local optimal solutions in the feature selection problems and on the same time discover a precise estimation of the optimal subset.

Conclusion

In this paper, a novel hybridization approach based on Interior Search Algorithm (ISA) and chaos theory is presented. In ISA, α parameter plays a critical role in balancing the intensification and diversification. In the proposed chaotic ISA, 10 chaotic maps are embedded in ISA in order to replace the fixed value of α with the deterministic chaotic signals. The performances of the proposed chaotic variants are compared on 13 global benchmark functions and the simulation results indicated that the chaotic maps can significantly boost the performance of the ISA in terms of balancing the exploitation and exploration. Furthermore, the best chaotic ISA variant is employed as a feature selection approach and its performance is validated on 21 benchmark data sets from the UCI repository. The results of CISA as a feature selection approach are compared against well-known feature selection methods namely Ant Lion Optimization (ALO), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Dragonfly Algorithm (DA) and Salp Swarm Algorithm (SSA). The assessment is carried out using a set of evaluation criteria to analyze various aspects of performance and the results demonstrated that the proposed CISA can adaptively explore the subset of features optimally and converge to the optimal/near optimal solution better than the other algorithms.

For future studies, more chaotic maps are worth applying to ISA and CISA can be utilized as a filter feature selection approach to evaluate the generality of the selected features. It would be interesting to hybridize ISA algorithm with another population-based metaheuristic algorithm like butterfly optimization algorithm.

References

Alatas, B. 2010. Chaotic harmony search algorithms. *Applied Mathematics and Computation* 216:2687–99. doi:10.1016/j.amc.2010.03.114.

Alatas, B., E. Akin, and A. B. Ozer. 2009. Chaos embedded particle swarm optimization algorithms. *Chaos, Solitons & Fractals* 40:1715–34. doi:10.1016/j.chaos.2007.09.063.

Altman, N. S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46:175–85.

Arora, S., and P. Anand. 2017. Chaos-enhanced flower pollination algorithms for global optimization. *Journal of Intelligent & Fuzzy Systems* 33:3853–69. doi:10.3233/JIFS-17708.

Arora, S., and P. Anand. 2018a. Chaotic grasshopper optimization algorithm for global optimization. *Neural Computing and Applications*, *31*(8):4385–4405.

Arora, S., and P. Anand. 2018b. Learning automata based butterfly optimization algorithm for engineering design problems. *International Journal of Computational Materials Science and Engineering*, *116*:147-160.

Arora, S., and P. Anand. 2019. Binary butterfly optimization approaches for feature selection. *Expert Systems with Applications* 116:147–60. doi:10.1016/j.eswa.2018.08.051.

Arora, S., and S. Singh 2014. Performance research on firefly optimization algorithm with mutation. In *International Conference, Computing & Systems,* Tamil Nadu, India.

Arora, S., and S. Singh. 2017a. An effective hybrid butterfly optimization algorithm with artificial bee colony for numerical optimization. *International Journal of Interactive Multimedia and Artificial Intelligence* 4:14–21. doi:10.9781/ijimai.2017.442.

Arora, S., and S. Singh. 2017b. An improved butterfly optimization algorithm with chaos. *Journal of Intelligent & Fuzzy Systems* 32:1079–88. doi:10.3233/JIFS-16798.

Arora, S., and S. Singh. 2018. Butterfly optimization algorithm: A novel approach for global optimization. *Soft Computing*, *23*:715–734.

Balasaraswathi, V. R., M. Sugumaran, and Y. Hamid. 2017. Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms. *Journal of Communications and Information Networks* 2:107–19. doi:10.1007/s41650-017-0033-7.

Bello, R., Y. Gomez, A. Nowe, and M. M. Garcia 2007. Two-step particle swarm optimization to solve the feature selection problem. In *Intelligent Systems Design and Applications, 2007. ISDA 2007. Seventh International Conference*, Rio de Janeiro, Brazil, 691–96. IEEE.

Chen, H., W. Jiang, C. Li, and R. Li. 2013. A heuristic feature selection approach for text categorization by using chaos optimization and genetic algorithm. *Mathematical Problems in Engineering* (2013):1–6. doi:10.1155/2013/524017.

Derrac, J., S. Garca, D. Molina, and F. Herrera. 2011. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation* 1:3–18. doi:10.1016/j.swevo.2011.02.002.

Eberhart, R., and J. Kennedy 1995a. A new optimizer using particle swarm theory. In *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium*, Perth, Western Australia, 39–43. IEEE.

Eberhart, R., and J. Kennedy 1995b. *Proceedings of IEEE international conference on neural networks*.

Emary, E., H. M. Zawbaa, and A. E. Hassanien. 2016a. Binary ant lion approaches for feature selection. *Neurocomputing* 213:54–65. doi:10.1016/j.neucom.2016.03.101.

Emary, E., H. M. Zawbaa, and A. E. Hassanien. 2016b. Binary grey wolf optimization approaches for feature selection. *Neurocomputing* 172:371–81. doi:10.1016/j.neucom.2015.06.083.

Franklin, J. 2005. The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer* 27:83–85. doi:10.1007/BF02985802.

Gandomi, A., X.-S. Yang, S. Talatahari, and A. Alavi. 2013. Firefly algorithm with chaos. *Communications in Nonlinear Science and Numerical Simulation* 18:89–98. doi:10.1016/j.cnsns.2012.06.009.

Gandomi, A. H. 2014. Interior search algorithm (isa): A novel approach for global optimization. *ISA Transactions* 53:1168–83. doi:10.1016/j.isatra.2014.03.018.

Gandomi, A. H., and D. A. Roke 2014. Engineering optimization using interior search algorithm. In *Swarm Intelligence (SIS), 2014 IEEE Symposium*, Orlando, FL, USA, 1–7. IEEE.

Goldberg, D. E., and J. H. Holland. 1988. Genetic algorithms and machine learning. *Machine Learning* 3:95–99. doi:10.1023/A:1022602019183.

Gu, S., R. Cheng, and Y. Jin. 2018. Feature selection for high-dimensional classification using a competitive swarm optimizer. *Soft Computing* 22:811–22. doi:10.1007/s00500-016-2385-6.

Guyon, I., and A. Elisseeff. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* 3:1157–82.

Han, J., J. Pei, and M. Kamber. 2011. *Data mining: Concepts and techniques*. Waltham, MA, USA: Elsevier.

He, D., C. He, L.-G. Jiang, H.-W. Zhu, and G.-R. Hu. 2001. Chaotic characteristics of a one-dimensional iterative map with infinite collapses. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions* 48:900–06. doi:10.1109/81.933333.

Hedar, A.-R., J. Wang, and M. Fukushima. 2008. Tabu search for attribute reduction in rough set theory. *Soft Computing* 12:909–18. doi:10.1007/s00500-007-0260-1.

Holland, J., and D. Goldberg. 1989. *Genetic algorithms in search, optimization and machine learning*. Massachusetts: Addison-Wesley.

Jordehi, A. R. 2015. A chaotic artificial immune system optimisation algorithm for solving global continuous optimisation problems. *Neural Computing and Applications* 26:827–33. doi:10.1007/s00521-014-1751-5.

Joshi, H., and S. Arora. 2017. Enhanced grey wolf optimization algorithm for global optimization. *Fundamenta Informaticae* 153:235–64. doi:10.3233/FI-2017-1539.

Kabir, M. M., M. Shahjahan, and K. Murase. 2011. A new local search based hybrid genetic algorithm for feature selection. *Neurocomputing* 74:2914–28. doi:10.1016/j.neucom.2011.03.034.

Kalra, S., and S. Arora 2016. Firefly algorithm hybridized with flower pollination algorithm for multimodal functions. In *International Congress on Information and Communication Technology*, Udaipur, India, 207–19. Springer.

Karaboga, D., and B. Basturk. 2007. A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *Journal of Global Optimization* 39:459–71. doi:10.1007/s10898-007-9149-x.

Kohavi, R., and G. H. John. 1997. Wrappers for feature subset selection. *Artificial Intelligence* 97:273–324. doi:10.1016/S0004-3702(97)00043-X.

Kohli, M., and S. Arora. 2017. Chaotic grey wolf optimization algorithm for constrained optimization problems. *Journal of Computational Design and Engineering* 5(4):458-472.

Kumar, M., T. K. Rawat, A. Jain, A. A. Singh, and A. Mittal. 2015. Design of digital differentiators using interior search algorithm. *Procedia Computer Science* 57:368–76. doi:10.1016/j.procs.2015.07.351.

Li, Q., H. Chen, H. Huang, X. Zhao, Z. Cai, C. Tong, W. Liu, and X. Tian. 2017. An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis. *Computational and Mathematical Methods in Medicine* 1–15. doi:10.1155/2017/9512741.

Liu, H., and L. Yu. 2005. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering* 17:491–502. doi:10.1109/TKDE.2005.66.

Lu, H., X. Wang, Z. Fei, and M. Qiu. 2014. The effects of using chaotic map on improving the performance of multiobjective evolutionary algorithms.*Mathematical Problems in Engineering 2014*:1-16, Article ID 924652.

Mafarja, M., and S. Abdullah. 2013. Record-to-record travel algorithm for attribute reduction in rough set theory. *Journal of Theoretical and Applied Information Technology* 49:507–13.

Mafarja, M., and S. Abdullah. 2015. A fuzzy record-to-record travel algorithm for solving rough set attribute reduction. *International Journal of Systems Science* 46:503–12. doi:10.1080/00207721.2013.791000.

Mafarja, M., and S. Mirjalili. 2018. Whale optimization approaches for wrapper feature selection. *Applied Soft Computing* 62:441–53. doi:10.1016/j.asoc.2017.11.006.

Mafarja, M. M., and S. Mirjalili. 2017. Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* 260:302–12. doi:10.1016/j.neucom.2017.04.053.

Mirjalili, S. 2015. The ant lion optimizer. *Advances in Engineering Software* 83:80–98. doi:10.1016/j.advengsoft.2015.01.010.

Mirjalili, S. 2016a. Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Computing and Applications* 27:1053–73. doi:10.1007/s00521-015-1920-1.

Mirjalili, S. 2016b. Sca: A sine cosine algorithm for solving optimization problems. *Knowledge-Based Systems* 96:120–33. doi:10.1016/j.knosys.2015.12.022.

Mirjalili, S., A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili. 2017. Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software* 114:163–91. doi:10.1016/j.advengsoft.2017.07.002.

Mirjalili, S., and A. Lewis. 2016. The whale optimization algorithm. *Advances in Engineering Software* 95:51–67. doi:10.1016/j.advengsoft.2016.01.008.

Mirjalili, S., S. M. Mirjalili, and A. Lewis. 2014. Grey wolf optimizer. *Advances in Engineering Software* 69:46–61. doi:10.1016/j.advengsoft.2013.12.007.

Moravej, M., and S.-M. Hosseini-Moghari. 2016. Large scale reservoirs system operation optimization: The interior search algorithm (ISA) approach. *Water Resources Management* 30:3389–407. doi:10.1007/s11269-016-1358-y.

Naanaa, A. 2015. Fast chaotic optimization algorithm based on spatiotemporal maps for global optimization. *Applied Mathematics and Computation* 269:402–11. doi:10.1016/j.amc.2015.07.111.

Peng, H., F. Long, and C. Ding. 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27:1226–38. doi:10.1109/TPAMI.2005.159.

Saremi, S., S. Mirjalili, and A. Lewis. 2017. Grasshopper optimisation algorithm: Theory and application. *Advances in Engineering Software* 105:30–47. doi:10.1016/j.advengsoft.2017.01.004.

Sayed, G. I., A. Darwish, and A. E. Hassanien. 2018. A new chaotic whale optimization algorithm for features selection. *Journal of Classification 35*:300–344.

Sayed, G. I., A. E. Hassanien, and A. T. Azar. 2017. Feature selection via a novel chaotic crow search algorithm. *Neural Computing and Applications 31*(1):171–188.

Sayed, G. I., G. Khoriba, and M. H. Haggag. 2018. A novel chaotic salp swarm algorithm for global optimization and feature selection. *Applied Intelligence 48*, 10:3462-3481..

Sayed, S. A.-F., E. Nabil, and A. Badr. 2016. A binary clonal flower pollination algorithm for feature selection. *Pattern Recognition Letters* 77:21–27. doi:10.1016/j.patrec.2016.03.014.

Tavazoei, M. S., and M. Haeri. 2007. Comparison of different one-dimensional maps as chaotic search pattern in chaos optimization algorithms. *Applied Mathematics and Computation* 187:1076–85. doi:10.1016/j.amc.2006.09.087.

Wang, J., T. Li, and R. Ren 2010. A real time IDSs based on artificial bee colony-support vector machine algorithm. In *Advanced computational intelligence (IWACI), 2010 third international workshop on*, 91–96. IEEE, Suzhou, China.

Wolpert, D. H., and W. G. Macready. 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1:67–82. doi:10.1109/4235.585893.

Xue, B., M. Zhang, and W. N. Browne. 2014. Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Applied Soft Computing* 18:261–76. doi:10.1016/j.asoc.2013.09.018.

Yang, X.-S. 2010a. Firefly algorithm, stochastic test functions and design optimisation. *International Journal of Bio-Inspired Computation* 2:78–84. doi:10.1504/IJBIC.2010.032124.

Yang, X.-S. 2010b. *Nature-inspired metaheuristic algorithms*. Frome, UK: Luniver press.

Yang, X. S., M. Karamanoglu, and X. He. 2014. Flower pollination algorithm: A novel approach for multiobjective optimization. *Engineering Optimization* 46:1222–37. doi:10.1080/0305215X.2013.832237.

Yang, Y., and J. O. Pedersen. 1997. A comparative study on feature selection in text categorization. Conference Paper published in proceedings of the Fourteenth International Conference on Machine Learning, Nashville, Tennessee, USA, vol. 97, 412–20.

Yldz, B. S. 2017. Natural frequency optimization of vehicle components using the interior search algorithm. *Materials Testing* 59:456–58. doi:10.3139/120.111018.

Zawbaa, H. M., E. Emary, and C. Grosan. 2016. Feature selection via chaotic antlion optimization. *PloS One* 11:e0150652. doi:10.1371/journal.pone.0150652.

Zheng, W.-M. 1994. Kneading plane of the circle map. *Chaos, Solitons & Fractals* 4:1221–33. doi:10.1016/0960-0779(94)90033-7.

Zorarpac, E., and S. A. Özel. 2016. A hybrid approach of differential evolution and artificial bee colony for feature selection. *Expert Systems with Applications* 62:91–103. doi:10.1016/j.eswa.2016.06.004.