A Hybrid Crow Search and Grey Wolf Optimization Technique for Enhanced Medical Data Classification in Diabetes Diagnosis System

C. Mallika¹ · S. Selvamuthukumar²

Received: 10 August 2019 / Accepted: 23 August 2021
© The Author(s) 2021

Abstract

Diabetes is an extremely serious hazard to global health and its incidence is increasing vividly. In this paper, we develop an effective system to diagnose diabetes disease using a hybrid optimization-based Support Vector Machine (SVM). The proposed hybrid optimization technique integrates a Crow Search algorithm (CSA) and Binary Grey Wolf Optimizer (BGWO) for exploiting the full potential of SVM in the diabetes diagnosis system. The effectiveness of our proposed hybrid optimization-based SVM (hereafter called CS-BGWO-SVM) approach is carefully studied on the real-world databases such as UCIPima Indian standard dataset and the diabetes type dataset from the Data World repository. To evaluate the CS-BGWO-SVM technique, its performance is related to several state-of-the-arts approaches using SVM with respect to predictive accuracy, Intersection Over-Union (IoU), specificity, sensitivity, and the area under receiver operator characteristic curve (AUC). The outcomes of empirical analysis illustrate that CS-BGWO-SVM can be considered as a more efficient approach with outstanding classification accuracy. Furthermore, we perform the Wilcoxon statistical test to decide whether the proposed cohesive CS-BGWO-SVM approach offers a substantial enhancement in terms of performance measures or not. Consequently, we can conclude that CS-BGWO-SVM is the better diabetes diagnostic model as compared to modern diagnosis methods previously reported in the literature.

Keywords Classification · Crow search algorithm · Diabetes · Grey wolf optimizer · Hybrid optimization · Support vector machine

1 Introduction

Precisemedical diagnosis and effective disease management are two significant problems in life science analysis and have a constructive effect on the health condition of individuals. At present, the majority of healthcare institutions are adopting innovative technologies for their administrative endeavor to enable an improved healthcare service in terms of techno-economic aspects [1]. Therefore, often colossal and devastating volumes of data (i.e., petabytes or terabytes)

are collected from this medical industry. The potential and usefulness of big data in the medical sector, such as disease management and providing an effective treatment are realized from its impending nature of the appropriate methods to intelligently discover knowledge and build models from data [2]. Intellectual database management systems provide efficient tools to extract knowledge from this complex and massive data and convert all extant data into useful information to develop an intelligent decision support system. Diabetes has been a serious global health issue for many years. Innumerable worldwide and nationwide epidemiological studies have witnessed the increasing number of diabetic patients globally. It is a non-communicable disease categorized by hyperglycemia (i.e., high blood glucose) and associated with disorders of carbohydrate, protein, and fat. Besides, it deteriorates nearly all organs of the body including the kidney, heart, foot, eyes, skin, and so on. The diabetes managing applications demand unremitting clinical services to decrease the jeopardy of long-term problems and to thwart deadly consequences.

✉ C. Mallika
cmallikachinna@gmail.com
S. Selvamuthukumar
smksmk@gmail.com

¹ Department of Master of Computer Applications, E.G.S. Pillay Engineering College, Nagapattinam, Tamilnadu, India

² Department of Computer Applications, A.V.C. College of Engineering, Mannampandal, Mayiladuthurai, Tamilnadu, India

In the last few decades, health care professionals pay much more attention to an early diagnosis of diabetes disease and design several approaches to resolve these diagnosis issues. Exploiting machine learning techniques in the Diabetes Diagnosis System (DDS) will help the medical professional to provide timely care for patients by diagnosing the disease at the initial stage. Researchers have employed artificial intelligence techniques to design some DDS, which enhance the performance of the diabetes management system. Recently, several research works apply the SVM method to find out diabetes [3–5]. SVM is a classifier based on a supervised machine learning algorithm and capable of differentiating the ‘intrinsic features’ of different data samples for nonlinear problems. Nevertheless, the effectiveness of the SVM classification is closely associated with the feature of the selected parameters. Hence, parameter optimization techniques perform a vital role in SVM performance and affect the classification accuracy of DDS. The SVM-based classifier consists of two processes namely, training and testing. This approach recognizes the metrics for classification using known datasets in training phase. In the testing phase, it carries out classification on the basis of those metrics and labels them to the equivalent class. In the present work, we aim to achieve the diabetes classification using a hybrid optimization technique-based SVM classification.

While developing and implementing a meta-heuristic, it should be remembered that a suitable balance between exploitation and exploration must be realized effectively. By the way, one important alternate solution is to design a hybrid approach in which (at least) two algorithms are assimilated together to increase the system performance. The key objective of this study is to develop and design a cohesive optimizer aiming at the assimilation of CSA and BGWO to establish higher classification accuracy on real-world problems. CSA is a meta-heuristic optimizer that imitates the intellectual behaviour of crows. In this study, we utilize the cleverness of crows to store food in a hiding place and retrieve it whenever needed. The merit of CSA lies in the aptitude to evade the trapping in the local optimum efficiently when coping with multimodal optimization problems in a more complex searching space. Nevertheless, the exploitation phase of CSA is not much effective.

Grey wolf optimizer (GWO) is also a meta-heuristic optimizer that imitates the social behaviour of grey wolves such as leadership hierarchy and the group hunting. GWO has the superior abilities to circumvent algorithms getting stuck into local optimal value [6]. It also has a fast convergence speed in finding the optimum solution. Generally, GWO is good at exploitation. But, it suffers from the deprived global searching ability. Therefore, in certain circumstances, it cannot handle the problem effectively and flops to calculate the global optimum solution. This study develops a binary grey

wolf optimization technique for handling parameter optimization problems, and then SVM is used to achieve diagnostic processes according to the selected features recognized by the BGWO algorithm.

Considering the virtues of CSA and BGWO, these optimization techniques are perfect for hybridization. The hybrid optimizer exploits the strengths of both CSA and BGWO techniques successfully to provide promising candidate solutions and realize global optima effectively. The proposed model adopts CSA to select the better initial position and then implements BGWO to appraise the position of each searching agent for obtaining the optimum parameters with better accuracy. Hence, the proposed approach integrates these two algorithms to balance exploitation and exploration appropriately and provide improved performance than the traditional CSA and GWO with respect to accuracy, IoU, specificity, AUC, and sensitivity of the classification. The key contributions of our study are three-fold.

1. A hybrid optimization technique based on CSA and BGWO is introduced.
2. The hybridization of CSA-BGWO is implemented with SVM to diagnose diabetes.
3. The proposed CS-BGWO-SVM classification approach uses datasets including the Pima Indian and the diabetes type dataset from the Data World repository for experimentation and the results are validated against several advanced SVM approaches.

The remainder sections of the article are organized as: We explore related hybrid optimization methods in Sect. 2. We describe the fundamental notions of SVM, BGWO, and CSA in Sect. 3. A comprehensive depiction of the presented method is presented in Sect. 4. The experimental setup, evaluation metrics and numerical results are given in Sect. 5. Finally, the research is concluded at Sect. 6.

2 Related Work

Recently, several different hybrid metaheuristics approaches have been proposed and studied to resolve various complicated optimization issues [7]. Hybrid algorithms have become very popular due to their potential to manage the high-dimensional nonlinear problems of optimization [8]. Hitherto, extensive investigations have been carried out by several researchers on hybrid optimization approaches. Kao and Zahara proposed a novel optimization technique in which the Genetic Algorithm (GA) has been integrated with Particle Swarm Optimization (PSO) for hybridization [9]. To generate individuals, this hybrid technique not only used cross-over and mutation operators but also used searching operators of PSO. Tsai et al. introduced Taguchi-genetic

algorithm-based hybridization to reach a global solution [10]. In this study, the Taguchi method is combined with GA to select higher quality genes to achieve better performance.

Jitkongchuen presented an enhanced mutation method using hybrid differential evolution approach. In this study, a differential evolution algorithm is combined with GWO for handling continuous global optimization problems [11]. Nabil proposed an improved version of the flower pollination algorithm (FPA). This work combines FPA with a clonal selection approach to achieve better accuracy than the traditional FPA [12]. The effectiveness of the modified FPA is evaluated via different benchmark datasets. Tawhid and Ali developed an innovative cohesive method based on GWO and GA. The authors applied this approach to reduce potential energy function [13]. Jayabarathi et al. developed a combined GWO approach by as simulating the original GWO with GA for getting a better efficiency for handling economic dispatch problems [14]. Singh and Singh hybridized the local search facility of PSO with the global search facility of GWO to improve efficiency and computational efforts [15]. Gaidhane and Nigam proposed a new hybrid approach to optimize the constraints for increasing the performance of complicated problems. This work integrates the bee's communication tactic from Artificial Bee Colony (ABC) algorithm, to realize effective exploration, with the leadership hierarchy of GWO retaining its exploitation ability [16].

Zhang et al. hybridized biogeography-based optimization (BBO) and GWO to make a suitable trade-off between local search and global search to achieve superior results than the original implementation of BBO and GWO. Hassanien et al. introduced the CSA to provide the global optimum result precisely. The proposed approach can effectively handle the roughness and impreciseness of the prevailing statistics about the global optimum and eventually increase the efficiency of the original CSA [17]. The aforementioned research works proved that the hybrid algorithms have higher performance related to other optimization methods. In the domain of feature extraction, several combined methods are developed. In the year 2004, Oh et al. presented the first cohesive metaheuristic method for attribute selection which intended to hybrid local search methods and GA to control the searching process [18]. Talbi et al. proposed another hybrid approach for feature selection which intended to integrate PSO, GA, and SVM classification model and applied to microarray data classification [19].

Hybridization in ant colony optimization using the cuckoo search [20], and using GA [21, 22] are also developed for effective feature selection process. Mafarja and Mirjalili developed a cohesive method to combine simulated annealing methods with the whale optimization approach for identifying the ideal set of features [23]. In recent times, a technique using evolutionary population dynamics [24] and

grasshopper optimization [25] has been effectively implemented in [26] for handling optimal feature selection problems. For further information regarding feature selection and optimization approaches, interested readers can refer to [27] and [28].

3 Background Analysis

This section presents some basic concepts of SVM, BGWO, and CSA that have been utilized to construct and implement the proposed CS-BGWO-SVM classification model.

3.1 Support Vector machine

It is a discriminative classifier developed by Vapnik to find anomalies of biomedical signals owing to its excellence and strong ability to deal with the non-linear and high-dimensional dataset in the field of the healthcare industry [29]. The key notion behind this classifier is to differentiate the unknown validation set of data in to their appropriate classes based on the training set of some well-known data.

In binary classification, SVM constructs a hyperplane that optimally differentiates samples into two classes. Let $\{x_i, y_i\}_{i=1}^n$ is the training dataset. Here x denotes the input sample, and $y \in \{+1, -1\}$ is the class label. Then the hyperplane is denoted as

$$w^T \cdot x + b = 0 \quad (1)$$

In above Eq. (1), a coefficient vector w is orthogonal to the hyperplane. The term b denotes the distance between the point in the dataset and origin. The key objective of the SVM is to determine b and w values. To get the ideal hyperplane $\|w^2\|$ should be minimized under the condition of $y_i(w^T \cdot x_i + b) \geq 1$ as shown in Fig. 1. Hence, the optimization problem is formulated as

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|w^2\| \\ &\text{subject to } y_i(w^T \cdot x_i + b) \geq 1, i = 1, 2, \dots, n \end{aligned} \quad (2)$$

In a linear problem, Lagrange multipliers are used to calculate w . The data points lay on the decision boundary are named as support vector. Therefore, the value of w can be calculated as

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (3)$$

Fig. 1 The classification process of SVM

In Eq. (3), α_i denotes the Lagrange multipliers and n represents the number of support vectors. After calculating w the value of b can be calculated as follows

$$y_i(w^T x_i + b) - 1 = 0 \quad (4)$$

Then the linear discriminant function can be expressed as

$$\hat{y} = \text{sgn}\left(\sum_{i=1}^n \alpha_i y_i x_i^T x_i + b\right) \quad (5)$$

To handle a nonlinear problem, SVM employs kernel trick. Then the decision function can be given by

$$\hat{y} = \text{sgn}\left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b\right) \quad (6)$$

here $\hat{y}$ is the kernelized label for the unlabeled input x and the sgn function defines whether the anticipated classification comes out positive or negative. Usually, any positive definite kernel functions such as Gaussian function $K(x_i, x) = \exp(-\gamma \|x - x_i\|^2)$ and the polynomial function $K(x_i, x) = (x^T x_i + 1)^d$ that meet Mercer's constraint [30]. This subsection only gives a brief note on SVM. For further reading, readers can refer to [31] which give a comprehensive depiction of the SVM concepts.

3.2 Binary GWO

The conventional greywolf optimizer is a bio-inspired metaheuristic approach introduced by Mirjalili et al. [32]. GWO imitates the leadership level and hunting actions of grey wolves. Grey wolves are Canidae species and have a rigorous

social dominant hierarchy. They are famous for their swarm intelligence-based hunting mechanism. Grey wolves are social predators or pack hunters which hunt their prey (target) by working together in a group of 5 to 12 wolves (i.e., pack). Within the pack, the leadership is divided into four levels like alpha (α), beta (β), delta (δ), and omega (ω). α (dominant) wolf is at the upper most hierarchy and acts as the leader of the cluster. Alpha is liable for making all the verdicts about hunting, sleeping place, wake uptime, maintaining discipline, etc. [33]. Its verdicts are dictated to and followed by the entire group.

The wolves in the next level are beta which acts as an advisor to alpha and fortifies the orders of alpha all over the group. This means that β aids α in decision making or other tasks and dominate other lower-level wolves. Beta wolves being the sub-leaders in the group have the maximum possibility to become α wolves if one of α becomes very old or dead. The subsequent hierarchy is δ . They have to bow to α as well as β ; in chorus, they instruct next lower-level wolves (i.e., ω wolves). Hunters, sentinels, scouts, caretakers, and elders belong to this level. Hunters support α and β when hunting targets and providing foodstuff for the group. Sentinels protect and assure the security of the group. Scouts inspect the borders of the search space and alert the group in case of any hazard. Caretakers are in charge of looking after the injured, sick, and weak wolves in the group. Finally, older wolves are experts who used to be α or β .

The wolves present at the lowermost level in the dominance structure are omega. They act as a scapegoat. They always subordinate to other higher-level wolves. The omega wolves are permitted to consume food last of all. It looks like ω is not an important wolf in the group, but it has been witnessed that the entire cluster realizes some complications and in-house fighting when dropping the ω wolf. This is owing to the expelling of frustration and violence of all wolves by ω . This helps in satisfying the whole group and preserving the hierarchy. In some scenarios, ω is also the baby-sitters in the group. Besides this dominant structure, pack hunting is an additional remarkable characteristic of grey wolves. As stated by Munro et al. the basic steps of group hunting are given by [34]: (i) to track, chase and grasp the target (ii) to enfold and harass the target until it becomes stable, and (iii) to attack and kill the prey. Figure 2 illustrates these steps.

The following equations are used to define the encircling nature of the wolves [32].

$$\vec{N} = \left| \vec{K} \cdot \vec{P}_{\text{prey}}(i) - \vec{P}_{SA}(i) \right| \quad (7)$$

$$\vec{P}_{\text{wolf}}(i+1) = \vec{P}_{\text{prey}}(i) - \vec{M} \cdot \vec{N} \quad (8)$$

where i specifies the current iteration. $\vec{P}_{\text{prey}}$ and $\vec{P}_{SA}$ are the position vectors of prey and search agent correspondingly. $\vec{M}$ and $\vec{K}$ are coefficient vectors and are estimated as

Fig. 2 Hunting behavior of grey wolves: **A** tracking; **B, C** encircling **D** attacking

$$\vec{M} = 2 \cdot \vec{m} \cdot \vec{a}_1 - \vec{m} \quad (9)$$

$$\vec{K} = 2 \cdot \vec{a}_2 \quad (10)$$

where $\vec{a}_1$ and $\vec{a}_2$ has arbitrary values in $[0,1]$. $\vec{m}$ acts as a controlling element and gradually drops from 2 to 0 over the courses of iterations. Considering the capability to locate the target, the searching agents can simply surround it. The alpha wolf directs the entire task. Every single wolf in the pack performs hunting and updates its locations based on the optimum place of α , β , and δ . The encircling behavior of searching agents is modeled as follows

$$\vec{N}_\alpha = \left| \vec{K}_1 \cdot \vec{P}_\alpha - \vec{P} \right| \quad (11)$$

$$\vec{N}_\beta = \left| \vec{K}_2 \cdot \vec{P}_\beta - \vec{P} \right| \quad (12)$$

$$\vec{N}_\delta = \left| \vec{K}_3 \cdot \vec{P}_\delta - \vec{P} \right| \quad (13)$$

$$\vec{P}_1 = \left| \vec{P}_\alpha - \vec{M}_1 \cdot \vec{N}_\alpha \right| \quad (14)$$

$$\vec{P}_2 = \left| \vec{P}_\beta - \vec{M}_2 \cdot \vec{N}_\beta \right| \quad (15)$$

$$\vec{P}_3 = \left| \vec{P}_\delta - \vec{M}_3 \cdot \vec{N}_\delta \right| \quad (16)$$

$$\vec{P}_{SA}(i+1) = \frac{\vec{P}_1 + \vec{P}_2 + \vec{P}_3}{3} \quad (17)$$

Searching agents attack the target only when it becomes stable. This process is formulated according to $\vec{m}$ and calculated as follows

$$\vec{m} = 2 - \frac{2i}{I_{\max}} \quad (18)$$

Therefore, $|\vec{M}| < 1$ denotes that the searching agent will be forced to attack the target by moving in the direction of the target and if $|\vec{M}| > 1$, the searching agent will get deviated from the target and will find out for another prey. The searching process is performed by grey wolves according to the optimum position of α , β , and δ . Similarly, the values of $\vec{M}$ and $\vec{K}$ decide the exploitation and exploration phases of the optimization process. We can use the arbitrary values for $\vec{M}$ to make the searching agent converge or diverge from the target. Arbitrary values of $\vec{K}$ lie in $[0,2]$ and act as a significant role in circumventing stagnation in local solutions. It appends some arbitrary weight to the target to make it a challenge for searching agents to denote the distance between itself and the target. If $\vec{K} > 1$, then $\vec{K}$ is maximizing the impact of the target, and if $\vec{K} < 1$, its influence will get minimize in a stochastic manner. In the entire course, $\vec{M}$ and $\vec{K}$ are attuned cautiously and the exploitation or exploration processes get emphasized or deemphasized. Eventually, when a definite condition is met the algorithm will be halted and the optimum value of α will be returned.

Indeed, the original grey wolf optimizer is introduced to handle the continuous optimization problems. In the case of discrete optimization problems, a binary form of the optimizer is essential. This work implements a binary variant of GWO for the feature extraction process. Each searching agent in the proposed BGWO consists of a flag vector and its size is as same as attributes. As soon as the position of each searching agent was updated by Eq. (17), its discrete position is expressed as follows.

$$F = \begin{cases} 1 & P_{ij} > 0.5 \\ 0 & \text{Otherwise} \end{cases} \quad (19)$$

where, P_{ij} represents the j^{th} place of the i^{th} wolf. Figure 3 displays the algorithm of BGWO.

3.3 Crow Search Algorithm

CSA imitates the smart performances of crows in hiding and stealing foodstuffs. It is a metaheuristic approach presented by Askarzadeh. It is widely used in several scientific solicitations particularly in the field of optimization techniques [35]. Crows are now considered to be the world's best genius birds. Crows have the largest ratio of brain-to-bodyweight. They can recognize individuals and inform each other when a

Algorithm 1: BGWO

Initialize the population of searching agents (n)
Initialize m, M, K
Estimate the fitness of each agent
Set P_a is the location of a
Set P_β is the location of β
Set P_γ is the position of γ
While ($i < I_{\max}$)
for ($1 \leq i \leq n$)
Update the position of each agent using equation (17)
endfor
If ($P > 0.5$)
Set $\text{Flag} = 1$
else
Set $\text{Flag} = 0$
endif
Update m, M, K
Estimate the fitness of each searching agents
Apprise the location of α, β, γ
 $i = i + 1$
endwhile
Return a

Fig. 3 Algorithm of BGWO

hostile one reaching. Likewise, they can exploit techniques, exchange information in matured ways, and recollect their feeding positions (places) across seasons [36].

Crows are famous for identifying their feeding positions, monitoring other crows, and stealing the food stuffs when the proprietor leaves. After a crow has executed larceny, it will provide additional fortifications such as shifting their feeding places to evade being an imminent target. They apply their intelligence of having been a thief to analyse the actions of other birds and can discover the optimum method to defend their foodstuffs from being taken off [37]. The rudimentary conventions of this CSA are (i) crows are reside in groups; (ii) crows reminisce their feeding positions; (iii) they monitor other crows to take foodstuff; and (iv) crows safeguard their feeding positions from being theft.

Consider a d -dimensional search area with a group of an N number of crows. The hiding place of the crow x at iteration i is denoted by $P^{x,i} = (P_1^{x,i}, P_2^{x,i}, P_3^{x,i}, \dots, P_d^{x,i})$ where $x = 1, 2, \dots, N$ and $i = 1, 2, \dots, I_{\max}$. Here $I_{\max}$ is the maximum iteration. All the crows have a memory in which their optimal place has been recalled. At iteration i , the place of search agent x is defined by $P^{x,i}$. This is the optimal position that x has achieved hitherto. Furthermore, crows shift their place and search for enriched sources of food. Assume crow y desires to visit its feeding position, $P^{y,i}$. At iteration i , x

Algorithm 2: CSA

Initialize the population of searching agents (n)
Estimate the fitness function of searching agents
Initialize the memory of searching agents
While ($i < I_{\max}$)
for each agent
Define the value of PA
Set the value of rand
if $\text{rand} \geq PA$
Apprise the location of agents by Equation (21)
else
Set the location of the agent arbitrarily
endif
endfor
Calculate the possibility of new locations
Estimate the fitness of each searching agent.
Apprise the memory of agents
 $i = i + 1$
endwhile
return the value of the position of the agent

Fig. 4 Algorithm of CSA

decides to monitor y to approach $P^{y,i}$. Now two scenarios can arise.

- (i) y is not cognizant of being monitored by x . Consequently, crow x will reach the feeding place of y . Now, the new place of x is measured by Eq. (20).

$$P^{x,i+1} = P^{x,i} + \text{rand}_x \cdot F^{x,i} \cdot (P^{y,i} - P^{x,i}) \quad (20)$$

where r and is a randomly selected value with uniform distribution taking values in $[0,1]$ and $F^{x,i}$ represents the flight length of x at iteration i .

- (ii) y is aware of being monitored by x . Hence, y will cheat crow x by flying to a new place in the search space to defend its foodstuff from being stolen.

Now, cases 1 and 2 becomes

$$P^{x,i+1} = \begin{cases} P^{x,i} + \text{rand}_x \cdot F^{x,i} \cdot (P^{y,i} - P^{x,i}) & \text{if } \text{rand}_y \geq PA^{y,i} \\ \text{arandomposition} & \text{otherwise} \end{cases} \quad (21)$$

where rand_y represents an arbitrary value and $PA^{y,i}$ denotes the probability of awareness (PA) of y at iteration i . Usually, meta-heuristic algorithms should deliver an adequate trade-off between intensification and diversification [38]. In this algorithm, diversification and intensification are mostly regulated by PA . By decreasing PA , this approach attempts to perform the search on a local space. In chorus, for the lesser value of PA the intensification is augmented. On the other hand, by increasing PA , the possibility of searching the space

for obtaining better solutions decreases, and this approach is apt to global search (randomization). Accordingly, the application of higher values of PA upturns the diversification. Figure 4 shows the pseudo-code of CSA.

4 Proposed System

4.1 Proposed CS-BGWO algorithm

For any population-based approach, the exploitation and exploration phases are mutually important to achieve better performance. In BGWO, the main problem is that all the locations of searching agents are rationalized according to the value of α , β , and δ in the entire process as given in Eq. (17). Mostly, this updating procedure causes early convergence since agents are not permitted to discover the search space effectively. Furthermore, the optimization process as given in Eq. (17) offers limited local search skills in the later part of the process which causes relaxed convergence. Therefore, to overcome these restrictions in BGWO is integrated with the crow search algorithm to realize an appropriate trade-off between local and global search. CSA includes a regulating factor $F^{x,i}$ in its position updating procedure as given in Eq. (21) which enables the agents to adapt the scope of the step movement in the direction of the new agents. This factor is significant in achieving the global optimum value as the higher value of $F^{x,i}$ result to global search where as a lesser value of $F^{x,i}$ leads to local search.

As stated in the previous section, BGWO has better local search capability however deprived global search ability, hence in the CS-BGWO algorithm, a higher value of $F^{x,i}$ is used to applying with the outstanding exploration feature of CSA as given in Eq. (22). This denotes that CS-BGWO can effectively utilize the proficiencies of two algorithms and thus, it is universally applicable. In the CS-BGWO algorithm, an agent is permitted to appraise its location only by α and β rather than apprising from α , β , and δ as given in Eq. (22).

$$\vec{P}_{SA}(i+1) = \vec{P}_1 + rand_x \cdot F^{x,i} \cdot ((\vec{P}_1 - \vec{P}) + (\vec{P}_2 - \vec{P}))/2 \quad (22)$$

To preserve population diversity, instead of allowing all the search agents to update their position from α and β wolves, only the value of α is exploited in the CS-BGWO algorithm. This reduction approach facilitates the proposed algorithm to avoid local optimum effectively.

$$\vec{P}_{SA}(i+1) = \vec{P}_1 + rand_x \cdot F^{x,i} \cdot (\vec{P}_1 - \vec{P}) \quad (23)$$

Although the CS-BGWO approach has the outstanding competencies of local and global search an appropriate

trade-off between these two phases must be realized to obtain better results. In an ideal situation, an optimization technique can search a large space in the initial phases of the process to evade early convergence and use a small search space in the later phases of the iteration to effectively refine the results.

To assess the efficiency of the developed approach, more instinctively, the convergence curve of the developed approach is related to other approaches found in the literature. In Fig. 5, the X-axis designates the iteration counts, and the Y-axis indicates the best score. In an ideal situation, an optimization technique can search a large space in the initial phases of the process to evade early convergence and use

Fig. 5 **a** Convergence curves of CS-BGWO-SVM Vs other approaches. **b** Convergence curves of CS-BGWO-SVM Vs other approaches. **c** Convergence curves of CS-BGWO-SVM Vs other approaches

a small search space in the future phases of the iteration to effectively refine the results. It is illustrated by Figs. 5a, b, and c. The convergence speed of CS-BGWO is lesser than other approaches in the earlier phase, this indicates that the proposed approach has robust global search ability than other approaches in the initial phase, and the approach is difficult to drop into the local optimum result. Furthermore, the CS-BGWO algorithm still preserves a robust search aptitude in the later phase.

From the viewpoint of the complete iteration, the global and local search abilities of the CS-BGWO approach have been well balanced, which shows that the hybridization of CSA and BGWO can successfully enhance the performance of the classification process, and is more suitable for resolving diabetes diagnosis issues than other related algorithms. This indicates that to achieve the necessary global–local search ratio, a constant balance probability among Eq. (22) and Eq. (23) is not constructive. Thus the present study implements an adaptive balance probability (ρ) that enables CS-BGWO to improve the accuracy and efficiency of searches. The value of ρ can be calculated as follows

$$\rho = 1 - \left(1.01 \frac{i^3}{I_{\max}^3} \right) \quad (24)$$

where $I_{\max}$ denotes the maximum iteration counts and i signifies the present one. This adaptive balance probability for various iterations is depicted in Fig. 6.

It is worth mentioning that $\bar{M}$ acts as a regulating element to balance local and global search. The value of $\bar{M}$ depends on $\bar{m}$ which eventually regulates the direction of the searching. The higher value of $\bar{m}$ enables global search and its smaller value enable the local search. This indicates that an apt selection of $\bar{m}$ provides a good trade-off between local and global search which results in higher performance. In the BGWO approach, the value of $\bar{m}$ is gradually reduced

Fig. 6 Adaptive balance probability

to 0 from 2. Hence, it can be witnessed that superior results can be realized if the value of $\bar{m}$ is decreased non-linearly. By exploiting this notion, an enhanced approach, as given in Eq. (25), is developed to select the values of $\bar{m}$. This tactic enables the CS-BGWO approach to discover the search space efficiently as compared with the original GWO.

$$m = 2 - \cos(\text{rand}_x()) \frac{1}{I_{\max}} \quad (25)$$

Figure 7 shows the pseudo-code of the CS-BGWO algorithm.

4.2 CS-BGWO-SVM model

In this work, a novel CS-BGWO-SVM framework is developed to handle the feature selection problem in DDS. CS-BGWO-SVM model consists of two phases: (i) CSA is employed to define the initial locations of searching agents, and then BGWO is implemented to choose the optimum feature subset by exploring the search space adaptively. The optimum feature extraction process applies to the selected

Algorithm 3: CS-BGWO

```

Initialize the population of searching agents (n)
Initialize m, M, K
Estimate the fitness of searching agent
Set Pα is the location of the optimum searching agent (α)
Set Pβ is the position of the second-optimal searching agent (β)
While (i < Imax)
  for (i = 1 to n)
    if ρ > rand
      Apprise the position of each agent by equation (22)
    else
      Apprise the location of each agent by equation (23)
    endif
  endfor
  If (Position > 0.5)
    Set Flag = 1
  else
    Set Flag = 0
  endif
  Update the value of m, M, K
  Update the value of p
  Estimate the fitness of each searching agent
  Apprise the location of α, β
  i = i + 1
endwhile
return α

```

Fig. 7 Algorithm of CS-BGWO

features to increase the classification accuracy and reduce the number of designated attributes simultaneously. The mean accuracy of the classification over the tenfold cross-validation (CV) scheme is employed as a fitness function for evaluating the selected feature subset; and (ii) the SVM classification model is simulated to achieve superior predictive accuracy according to the best feature subset.

The dataset has been normalized in the range of $[-1, +1]$ before classification. To achieve a more accurate estimate of a model's performance, the k -fold CV is employed [39]. The proposed work assumes k equals to 10. This means that the entire dataset is partitioned into ten subsections. For each iteration, one subsection is considered as the testing dataset and the other subsections are integrated to construct the training dataset. Then the mean value of error across all 10 independent runs is calculated. The merit of this technique is that all testing data are independent and the consistency of the outcomes could be improved. It is noteworthy that only one replication of the tenfold CV will not produce adequate results for evaluation owing to the randomness in data splitting. Therefore, all the results are stated on an average of ten trials to gain precise estimation [40].

5 Result and Discussion

The efficacy of the proposed approach is assessed by relating the numerical results with that of nine similar approaches, including conventional SVM [32], CSA-SVM [35], GA-SVM [41], PSO-SVM [42], Improved whale optimization algorithm-based SVM (IWOA-SVM) [43], Differential Evolution algorithm-based SVM (DE-SVM) [44], GWO-SVM [32], Enhanced GWO-based SVM (EGWO-SVM) [45], Augmented GWO-based SVM (AGWO-SVM) [46]. In this study, LIBSVM model introduced by Chang and Lin [47] is used to realize SVM.

5.1 Experimental Setup

In our assessment, the experiments are carried out on 3.6 GHz with 16 GB RAM, Intel Core i7-4790 processor with Windows 10 operating system. All the algorithms are simulated using MATLAB R2009b. The proposed classification approach uses the Pima Indian diabetes patient records in the UCI and the diabetes type dataset in the Data World repository. Pima Indian diabetes patient records are collected from the National Institute of Diabetes and Digestive and Kidney Diseases. This database involves 768 medical data of Pima Indian heritage, inhabitants existing near Phoenix, Arizona, USA [48]. There are eight features related to this dataset (i.e., plasma glucose concentration, 2-h serum insulin, the number of times pregnant, diastolic blood pressure,

Table 1 Statistical analysis of Pima Indian dataset [48]

Label	Feature	Mean	Min/Max
F ₁	Number of times pregnant	3.8	0/17
F ₂	Plasma glucose level estimated from glucose tolerance test	120.9	0/199
F ₃	Diastolic blood pressure (mm Hg)	69.1	0/122
F ₄	Triceps skinfold thickness (mm)	20.5	0/99
F ₅	Two hour serum insulin (mu U/ml)	79.8	0/846
F ₆	BMI (kg/m ²)	32	0/67.1
F ₇	Function of diabetes nutrition	0.5	0.078/2.42
F ₈	Age (years)	33.2	21/81

function of diabetes nutrition, triceps skin fold thickness, body mass index (BMI), and age). Table 1 shows the statistical report of each instance in this dataset. The range of binary variables is limited to '0' or '1'. The target variable '1' represents a positive result for diabetes disease (i.e., diabetic), '0' is a negative result (i.e., non-diabetic). After preprocessing dataset, there are 392 cases with no missing values. The number of cases in label '1' is 130, and the number of cases in label '0' is 262.

The patient records in diabetes type dataset are collected from the Data World repository [49]. This database consist of 1099 records with 8 attributes including Plasma glucose test randomly taken at any time, plasma glucose test usually taken in the morning or 8 h after a meal, blood sugar while fasting, blood sugar 90 min after a meal, HbA1c (glycated hemoglobin), type, class, and age. Table 2 shows the statistical report of each instance in this dataset. After preprocessing dataset there are 1009 cases with no missing values. The number of cases in label '1' is 653, and the number of cases in label '0' is 356.

Selecting the penalty factor (C) and kernel width (γ) is extremely significant for implementing SVM models. The value of C affects the classification accuracy. If the value of C is very high then the accuracy is maximum in the training process, but it is minimum in the testing process. If C has a lesser value, then the accuracy is unacceptable, sorting the approach impractical. The value of γ has a greater impact on accuracy than C due to its influence on classification performance. An infinitesimal value of γ leads to under-fitting whereas a very high value results in over-fitting [50]. We assume both parameters are in the range of $\{2^{-5}, 2^{-4}, \dots, 2^4, 2^5\}$. This study assumes C and γ as 32 and 0.125, respectively (i.e., $C=2^5$ (32) and $\gamma=2^{-3}$ (0.125)). The values of both parameters are selected through the trial and error technique. Table 3 shows the detail of the parameter setting.

Our proposed algorithm encounters challenging problems, such as outlier data and noise, which can reduce accuracy. The outlier detection method can be utilized in the pre-processing step to identify inconsistencies in data/outliers; thus,

Table 2 Statistical analysis of diabetes type dataset [49]

Label	Feature	Mean	Min/Max
F ₁	Plasma glucose test randomly taken at any time	10.73	7.9/13.1
F ₂	Plasma glucose test usually taken in the morning or 8 h after a meal	6.14	3.9/9.1
F ₃	Blood sugar while fasting	12.57	0/54
F ₄	Blood sugar 90 min after a meal	6.65	4.2/8.8
F ₅	HbA1c	43.48	28/66
F ₆	Type	Normal, Type 1, Type 2	
F ₇	Class	0, 1	
F ₈	Age	33.39	21/81

Table 3 Parameter settings

Algorithm	Parameters	Value(s)
CSA	Flight length	2
	Awareness probability	0.1
GA	Mutation ratio	0.1
	Cross over ratio	0.2
	Selection Mechanism	Roulette Wheel
PSO	Inertia	[0.9, 0.6]
	Acceleration constant	[2.0, 2.0]
GWO	M	Linearly decreased from 2 to 0

Table 4 The result of DBSCAN-based outlier detection

Dataset	# Instance (Original)	# Instances (After Data Cleaning)	# Outlier Data	# Normal Data
Pima Indian	768	437	45	392
Diabetes type	1099	1082	73	1009

a good classifier can be generated for better decision-making. Eliminating the outliers and noise from the training dataset will enhance the classification accuracy. Past literature showed that by removing noise the quality of real datasets is enhanced [51]. In this study, DBSCAN is employed to identify the outlier data in a diabetes dataset [52]. The goal is to find objects that close to a given point to create dense regions. The points that are located outside dense regions are treated as outliers. The resultant dataset after removing outlier is given in Table 4.

5.2 Evaluation MEASURES

To assess the efficacy of DDS using the CS-GWO-SVM model, we studied four important performance measures: classification accuracy, AUC [53], specificity, and sensitivity. These evaluation parameters are essential to be greater to

increase the efficiency of the presented approach. The utility is calculated with respect to the accuracy as

$$\text{Accuracy (ACC)} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \quad (26)$$

The intersection over-union (IoU) is a performance measure used to calculate the performance of any data classification technique. Given a data set, the IoU metric provides the similarity between the predicted data and the ground truth for a data existing in the dataset. The IoU measure can consider the class imbalance issue generally present in such a problem setting. It is defined as follows

$$\text{Intersection over – union (IoU)} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \quad (27)$$

Sensitivity and specificity denote in what manner the classification approach differentiates negative and positive cases. Sensitivity defines the rate of disease prediction and specificity denotes the rate of false alarm.

$$\text{Sensitivity (SEN)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (28)$$

$$\text{Specificity (SPE)} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (29)$$

In the above equations, true positive (TP) denote the number of persons properly categorized as diabetic patients; false negative (FN) denotes the number of diabetic patients wrongly identified as the non-diabetic. The term true negative (TN) denotes that the number of persons properly categorized as non-diabetic and false positive (FP) defines the number of non-diabetic person wrongly categorized as the diabetic one. The receiver operator characteristic (ROC) is a pictorial representation to define the classification accuracy. The curve shows the rate of TP and FP. The AUC represents the area under the ROC curve. This is an important measure for relating binary classification problems.

The Wilcoxon statistical test is carried out to decide whether the proposed CS-BGWO-SVM algorithm delivers

a noteworthy improvement related to other approaches or not [54]. The test was performed using the effects of the proposed CS-BGWO-SVM and related to each of the other approaches at a level of 5% statistical significance. Table 10 depicts the p values achieved through the test, where the p value < 0.05 indicate that the null hypothesis is precluded, i.e., there is a considerable variance at a level of 5% significance. In contrast, the p values > 0.05 signify that there is no notable variance between the compared values. It can be evaluated from the results depicted in Table 10. From the results, it is observed that most of the p values are less than 0.05 which confirms that the improvement achieved by the CS-BGWO-SVM is statistically significant.

5.3 Experimental Results and Discussion

The proposed CS-GWO-SVM classification model is implemented. Table 5 shows the comprehensive results achieved by the proposed model.

To facilitate the effectiveness of the proposed CS-GWO-SVM classification model, we relate the performance of the proposed approach to other prominent algorithms from the literature. The experimental results of various approaches are reported in Table 6. From Table 6, we can observe that the conventional SVM approach has realized nominal classification performance with 81.1% ACC, 86% IoU, 0.829 AUC, 81.4% sensitivity, and 85.4% specificity. With

Table 5 The comprehensive outcomes acquired by CS-BGWO-SVM for various designated feature subset of Pima Indian dataset

Fold	Selected feature subset	ACC	IoU	AUC	SEN	SPE	ρ value
#1	{F1, F3, F5, F6}	0.923	0.935	0.922	0.890	0.950	0.033
#2	{F2, F3, F5, F6, F8,}	0.948	0.951	0.956	0.912	0.968	0.042
#3	{F1, F2, F3, F7}	0.934	0.931	0.915	0.921	0.957	0.048
#4	{F1, F2, F5, F6, F8}	0.945	0.928	0.922	0.963	0.955	0.039
#5	{F2, F3, F8}	0.939	0.964	0.939	0.939	0.939	0.014
#6	{F2, F2, F5}	0.939	0.927	0.895	0.870	0.950	0.072
#7	{F1, F2, F6}	0.938	0.957	0.929	0.895	0.910	0.028
#8	{F1, F2, F5, F7, F8}	0.933	0.964	0.923	0.905	0.957	0.090
#9	{F1, F3, F5, F7}	0.919	0.928	0.895	0.870	0.945	0.022
#10	{F2, F3, F5, F6, F7}	0.943	0.942	0.932	0.892	0.962	0.034
Mean	N/A	0.937	0.943	0.923	0.906	0.949	0.042
SD	N/A	0.009	0.014	0.018	0.021	0.016	0.022

Table 6 The results obtained from Pima Indian dataset in terms of performance measures

Algorithm		ACC	IoU	AUC	SEN	SPE	ρ value
SVM	Mean	0.811	0.860	0.829	0.814	0.854	0.087
	SD	0.048	0.026	0.068	0.084	0.083	0.050
CSA-SVM	Mean	0.871	0.826	0.869	0.853	0.883	0.086
	SD	0.043	0.025	0.069	0.070	0.055	0.050
GA-SVM	Mean	0.857	0.874	0.853	0.825	0.862	0.091
	SD	0.109	0.022	0.086	0.085	0.071	0.101
PSO-SVM	Mean	0.862	0.867	0.859	0.829	0.891	0.174
	SD	0.038	0.016	0.036	0.051	0.052	0.129
DE-SVM	Mean	0.866	0.864	0.863	0.842	0.890	0.135
	SD	0.037	0.019	0.035	0.048	0.057	0.099
IWOA-SVM	Mean	0.874	0.877	0.869	0.861	0.892	0.145
	SD	0.036	0.024	0.034	0.045	0.064	0.107
GWO-SVM	Mean	0.861	0.876	0.856	0.835	0.877	0.141
	SD	0.031	0.011	0.033	0.040	0.072	0.104
EGWO-SVM	Mean	0.890	0.893	0.861	0.842	0.897	0.170
	SD	0.025	0.010	0.049	0.053	0.064	0.132
AGWO-SVM	Mean	0.880	0.890	0.877	0.857	0.910	0.198
	SD	0.018	0.019	0.024	0.029	0.034	0.193
CS-BGWO-SVM	Mean	0.937	0.952	0.923	0.904	0.949	0.042
	SD	0.009	0.008	0.018	0.021	0.016	0.022

the intention of further increase in efficiency of the SVM approach, we combined optimization algorithms such as CSA, GA, PSO, GWO, EGWO, AGWO, and CS-BGWO with the original SVM. These models are implemented and examined rigorously on the same data. In the CSA-SVM classifier, PA is directly employed to regulate the algorithm diversity. Hence, this classifier has generated reasonable outcomes related to the original SVM.

This algorithm has realized classification results with 87.1% ACC, 82.6% IoU, 0.869 AUC, 85.3% sensitivity, and 88.3% specificity. The CSA-SVM and GA-SVM approaches provide similar results; however, the CSA-SVM approach provides improved standard deviation as compared to GA-SVM. Since these classifiers hinge on the arbitrary generation of individuals and always there is a possibility to create a zero variable vector. Figure 8 illustrates the comprehensive outcomes acquired by CS-BGWO-SVM for various designated feature subset.

As compared with CSA-SVM, GA-SVM, and PSO-SVM, WOA has higher data classification accuracy better accuracy (87.4%), IoU (87.7%), AUC (0.869), sensitivity (86.1%), and specificity (89.2%), and minimum prediction SD by selecting whales from the present generation and records while implementing an arbitrary search procedure.

The benefits like fewer constraints and random walks in GWO enable an effective optimization approach for SVM classification. GWO-SVM realized classification results with 86.1% ACC, 87.6% IoU, 0.856 AUC, 83.5% sensitivity, and 87.7% specificity. EGWO-SVM with a better hunting mechanism has realized better performance with 89% ACC, 89.3% IoU, 0.861 AUC, 84.2% sensitivity, and 89.7% specificity by developing a superior trade-off between local and global search that provides improved outcomes. AGWO-SVM hit better results with 88% ACC, 89% IoU, 0.877 AUC, 85.7% sensitivity, and 91% specificity due to its exploitation ability. It is possible to conclude that the CS-BGWO-SVM model has achieved much better performance with 93.7% ACC, 95.2% IoU, 0.923 AUC, 90.4% sensitivity, and 94.9% specificity. As compared with the original SVM model, CS-GWO-SVM has boosted 12.6%, 9.4%, 9.0%, and 9.5% with

Fig. 8 The comprehensive outcomes acquired by CS-BGWO-SVM for various designated feature subset

respect to ACC, AUC, SEN, and SPE correspondingly. Also, it is stimulating to observe that the SD gained by the CS-GWO-SVM is lesser than that of all most all other classification methods which means that the CS-GWO-SVM can provide more robust diagnose results.

The mean value of performance measures and SD gained from Pima Indian dataset by each approach is illustrated in Figs. 9 and 10. It can be found that the CS-GWO-SVM outdoes all other approaches with respect to the performance metrics. The key reason behind the superior performance of CS-GWO-SVM is that the CSA-based initialization can increase the efficiency of the feature selection process to a certain level. From Fig. 10, it can be observed that the SD of the CS-BGWO-SVM was less than all other approaches in terms of the performance measures. Therefore, CS-BGWO-based SVM model provides much more consistent results for diagnosing diabetes than the others. In other words, the CS-BGWO can not only enhance the enactment of SVM, but also enable better results for diagnosing diabetes disease. The comparative study demonstrates that CS-BGWO-SVM is a very competitive approach for diagnosing diabetes disease.

Fig. 9 Comparison of results obtained from Pima Indian dataset in terms of the mean value

Fig. 10 Comparison of results obtained from Pima Indian dataset in terms of SD

Table 7 The results obtained from diabetes type dataset in terms of performance measures

Algorithm		ACC	IoU	AUC	SEN	SPE	ρ value
SVM	Mean	0.813	0.820	0.839	0.851	0.856	0.093
	SD	0.048	0.025	0.069	0.088	0.050	0.136
CSA-SVM	Mean	0.873	0.788	0.879	0.892	0.885	0.091
	SD	0.043	0.024	0.070	0.073	0.050	0.090
GA-SVM	Mean	0.859	0.834	0.863	0.862	0.864	0.100
	SD	0.109	0.021	0.087	0.089	0.101	0.117
PSO-SVM	Mean	0.864	0.827	0.869	0.867	0.893	0.136
	SD	0.038	0.015	0.036	0.053	0.129	0.085
DE-SVM	Mean	0.868	0.824	0.873	0.880	0.892	0.172
	SD	0.037	0.018	0.035	0.050	0.099	0.094
IWOA-SVM	Mean	0.876	0.836	0.879	0.900	0.894	0.188
	SD	0.036	0.023	0.034	0.047	0.107	0.105
GWO-SVM	Mean	0.863	0.836	0.866	0.873	0.879	0.182
	SD	0.031	0.010	0.033	0.042	0.104	0.118
EGWO-SVM	Mean	0.892	0.852	0.871	0.880	0.899	0.129
	SD	0.025	0.010	0.050	0.055	0.132	0.105
AGWO-SVM	Mean	0.882	0.849	0.887	0.896	0.912	0.175
	SD	0.018	0.018	0.024	0.030	0.193	0.056
CS-BGWO-SVM	Mean	0.939	0.908	0.934	0.945	0.951	0.069
	SD	0.018	0.009	0.018	0.022	0.026	0.002

Fig. 11 Comparison of results obtained from diabetes type dataset in terms of the mean value

We can obtain similar results when we apply our proposed algorithm on diabetes type dataset. The experimental results obtained from diabetes type dataset using various approaches are reported in Table 7. The mean value of performance measures and SD gained from diabetes type dataset by each approach is illustrated in Figs. 11 and 12. From this table, we can observe that the conventional SVM approach has realized nominal classification performance with 81.3% ACC, 82% IoU, 0.839 AUC, 85.1% sensitivity, and 85.6% specificity. In the CSA-SVM classifier, PA is used to control the algorithm diversity. Therefore, it has produced reasonable outcomes related to the original SVM. This algorithm approach has realized classification

Fig. 12 Comparison of results obtained from diabetes type dataset in terms of the SD value

results with 87.3% ACC, 78.8% IoU, 0.879 AUC, 89.2% sensitivity, and 88.5% specificity. The CSA-SVM approach provides improved standard deviation as compared to GA-SVM. Since these classifiers hinge on the arbitrary generation of individuals and always there is a possibility to create a zero variable vector.

As compared with CSA-SVM, GA-SVM, and PSO-SVM, IWOA has higher data classification accuracy better accuracy (87.6%), IoU (83.6%), AUC (0.879), sensitivity (90%), and specificity (89.4%), with minimum prediction SD. GWO-SVM realized classification results with 86.1% ACC, 87.6% IoU, 0.856 AUC, 83.5% sensitivity, and 87.7% specificity. EGWO-SVM with a better hunting mechanism has realized better performance with 89.2% ACC, 85.2%

Table 8 Accuracy of the CS-IGWO-SVM approach Vs other approaches for different folding

Fold	SVM	CSA-SVM	GA-SVM	PSO-SVM	DE-SVM	IWOA-SVM	GWO-SVM	EGWO-SVM	AGWO-SVM	CS-IGWO-SVM
#1	0.727	0.792	0.786	0.864	0.858	0.868	0.862	0.872	0.891	0.923
#2	0.753	0.812	0.942	0.778	0.842	0.856	0.865	0.896	0.904	0.948
#3	0.849	0.853	0.956	0.905	0.899	0.874	0.852	0.915	0.869	0.934
#4	0.851	0.877	0.863	0.913	0.911	0.845	0.863	0.912	0.863	0.945
#5	0.798	0.882	0.694	0.868	0.875	0.901	0.901	0.856	0.879	0.939
#6	0.786	0.875	0.922	0.836	0.847	0.899	0.902	0.849	0.846	0.939
#7	0.786	0.892	0.912	0.848	0.856	0.871	0.891	0.877	0.877	0.938
#8	0.839	0.896	0.899	0.864	0.854	0.863	0.829	0.908	0.881	0.933
#9	0.845	0.892	0.653	0.885	0.901	0.875	0.825	0.912	0.901	0.919
#10	0.875	0.941	0.949	0.857	0.863	0.854	0.816	0.900	0.893	0.943
Mean	0.811	0.871	0.857	0.862	0.855	0.542	0.861	0.890	0.880	0.937
SD	0.048	0.043	0.109	0.038	0.034	0.032	0.031	0.025	0.018	0.009

Table 9 IoU of the proposed CS-BGWO-SVM vs other algorithms for different folding

Fold	SVM	CSA-SVM	GA-SVM	PSO-SVM	DE-SVM	IWOA-SVM	GWO-SVM	EGWO-SVM	AGWO-SVM	CS-IGWO-SVM
#1	0.835	0.871	0.854	0.874	0.835	0.894	0.872	0.901	0.912	0.935
#2	0.844	0.822	0.879	0.856	0.868	0.901	0.864	0.904	0.865	0.951
#3	0.863	0.794	0.854	0.843	0.845	0.904	0.876	0.910	0.857	0.955
#4	0.857	0.836	0.836	0.873	0.866	0.873	0.876	0.874	0.871	0.948
#5	0.823	0.814	0.878	0.864	0.874	0.854	0.895	0.876	0.893	0.964
#6	0.871	0.865	0.854	0.847	0.856	0.823	0.893	0.894	0.889	0.957
#7	0.892	0.832	0.887	0.854	0.897	0.877	0.874	0.892	0.896	0.957
#8	0.893	0.799	0.896	0.877	0.845	0.864	0.856	0.894	0.904	0.964
#9	0.897	0.801	0.902	0.883	0.864	0.891	0.875	0.895	0.905	0.948
#10	0.828	0.824	0.904	0.897	0.893	0.893	0.876	0.892	0.911	0.942
Mean	0.860	0.826	0.874	0.867	0.864	0.877	0.876	0.893	0.890	0.952
SD	0.026	0.025	0.022	0.016	0.019	0.0238	0.011	0.010	0.019	0.008

Table 10 AUC of the proposed CS-BGWO-SVM vs other for different folding

Fold	SVM	CSA-SVM	GA-SVM	PSO-SVM	DE-SVM	IWOA-SVM	GWO-SVM	EGWO-SVM	AGWO-SVM	CS-IGWO-SVM
#1	0.712	0.694	0.772	0.798	0.801	0.846	0.857	0.861	0.875	0.922
#2	0.756	0.831	0.913	0.828	0.842	0.888	0.902	0.884	0.894	0.956
#3	0.873	0.842	0.953	0.917	0.892	0.893	0.896	0.895	0.858	0.915
#4	0.752	0.895	0.866	0.903	0.897	0.866	0.851	0.862	0.852	0.922
#5	0.799	0.893	0.694	0.837	0.828	0.854	0.883	0.786	0.889	0.939
#6	0.861	0.878	0.912	0.869	0.871	0.863	0.856	0.769	0.836	0.895
#7	0.875	0.897	0.922	0.838	0.845	0.861	0.871	0.867	0.867	0.929
#8	0.899	0.893	0.899	0.854	0.862	0.844	0.826	0.858	0.921	0.923
#9	0.886	0.921	0.753	0.857	0.841	0.835	0.814	0.902	0.892	0.895
#10	0.875	0.941	0.843	0.885	0.875	0.827	0.804	0.925	0.883	0.932
Mean	0.829	0.869	0.853	0.859	0.864	0.850	0.856	0.861	0.877	0.923
SD	0.068	0.069	0.086	0.036	0.034	0.032	0.033	0.049	0.024	0.018

IoU, 0.871 AUC, 88.0% sensitivity, and 89.9% specificity by developing a superior trade-off between local and global search that provides improved outcomes. AGWO-SVM hit better results with 88.2% ACC, 84.9% IoU, 0.887 AUC, 89.6% sensitivity, and 91.2% specificity due to its exploitation ability. It is possible to conclude that the CS-BGWO-SVM model has achieved much better performance with 93.9% ACC, 90.8% IoU, 0.934 AUC, 94.5% sensitivity, and 95.1% specificity. Also, it is observed that the SD gained by the CS-GWO-SVM is lesser than that of all most all other classification methods which means that the CS-GWO-SVM can provide more robust and reliable diagnose results.

Figures 9, 10, 11, 12 also illustrate the p values of CS-BGWO-SVM in comparison with some meta-heuristic approaches achieved by the rank-sum test. This assessment is performed to decide whether the variation between the outcomes of CS-BGWO-SVM and the results of other approaches is noteworthy or not. Precisely, a p value is calculated, and if it is less than 0.05 then it designates that

the results attained by CS-BGWO-SVM have considerable deviations from those of the other approaches. The p value > 0.05 denotes that there is no substantial variation between the outputs of CS-BGWO-SVM and the other approaches used for evaluation. In most cases, it can be easily found that the p values calculated by the Wilcoxon test are less than 0.05 which reveals the outstanding performance of CS-BGWO-SVM.

Tables 8, 9, 10, 11, 12, 13 illustrate the results of all the approaches for various folding on Pima Indian dataset. The mean and SD results gained by each approach are stated in each table and the optimum numerical values are emphasized in bold. It is found that the performance measures obtained by the CS-BGWO-SVM approach are superior to all other approaches. Hence, the results show that the integration of CSA and BGWO has revealed better results as compared to the original GWO, CSA, and all other approaches used in this research work. Moreover, it is worth noting that CS-BGWO-SVM exhibits better

Table 11 Sensitivity of the proposed CS-BGWO-SVM vs other for different folding

Fold	SVM	CSA-SVM	GA-SVM	PSO-SVM	DE-SVM	IWOA-SVM	GWO-SVM	EGWO-SVM	AGWO-SVM	CS-IGWO-SVM
#1	0.698	0.686	0.754	0.693	0.723	0.741	0.756	0.799	0.825	0.890
#2	0.756	0.893	0.897	0.814	0.854	0.899	0.911	0.883	0.874	0.912
#3	0.869	0.912	0.931	0.864	0.869	0.874	0.876	0.875	0.838	0.921
#4	0.782	0.913	0.746	0.821	0.842	0.852	0.821	0.842	0.849	0.963
#5	0.699	0.775	0.674	0.844	0.846	0.851	0.853	0.766	0.856	0.939
#6	0.756	0.856	0.865	0.863	0.846	0.837	0.826	0.759	0.836	0.870
#7	0.878	0.862	0.892	0.831	0.824	0.830	0.831	0.839	0.854	0.895
#8	0.901	0.887	0.886	0.846	0.843	0.831	0.826	0.847	0.884	0.905
#9	0.899	0.865	0.757	0.854	0.842	0.832	0.824	0.902	0.882	0.870
#10	0.902	0.876	0.843	0.862	0.849	0.837	0.824	0.905	0.873	0.892
Mean	0.814	0.853	0.825	0.829	0.831	0.834	0.835	0.842	0.857	0.906
SD	0.084	0.070	0.085	0.051	0.045	0.043	0.040	0.053	0.029	0.021

Table 12 Specificity of the proposed CS-BGWO-SVM vs other for different folding

Fold	SVM	CSA-SVM	GA-SVM	PSO-SVM	DE-SVM	IWOA-SVM	GWO-SVM	EGWO-SVM	AGWO-SVM	CS-IGWO-SVM
#1	0.726	0.786	0.798	0.793	0.786	0.764	0.755	0.796	0.865	0.950
#2	0.796	0.893	0.917	0.844	0.867	0.912	0.935	0.953	0.894	0.968
#3	0.898	0.952	0.942	0.964	0.957	0.954	0.952	0.945	0.878	0.957
#4	0.821	0.953	0.957	0.921	0.928	0.935	0.941	0.942	0.886	0.955
#5	0.729	0.795	0.774	0.888	0.897	0.921	0.953	0.966	0.886	0.939
#6	0.852	0.876	0.869	0.923	0.924	0.925	0.926	0.789	0.916	0.950
#7	0.908	0.892	0.901	0.921	0.911	0.895	0.826	0.869	0.924	0.910
#8	0.931	0.897	0.886	0.936	0.921	0.874	0.831	0.867	0.924	0.957
#9	0.939	0.885	0.757	0.854	0.842	0.831	0.827	0.925	0.962	0.945
#10	0.942	0.896	0.823	0.862	0.854	0.847	0.824	0.921	0.963	0.962
Mean	0.854	0.883	0.862	0.891	0.884	0.879	0.877	0.897	0.910	0.949
SD	0.083	0.055	0.071	0.052	0.048	0.054	0.072	0.064	0.034	0.016

Table 13 p values calculated from the Wilcoxon's rank-sum test on different approaches

Fold	SVM	CSA-SVM	GA-SVM	PSO-SVM	DE-SVM	IWOA-SVM	GWO-SVM	EGWO-SVM	AGWO-SVM	CS-IGWO-SVM
#1	0.046	0.041	0.043	0.045	0.046	0.039	0.047	0.045	0.048	0.033
#2	0.065	0.058	0.043	0.067	0.054	0.065	0.068	0.058	0.062	0.042
#3	0.057	0.053	0.043	0.451	0.058	0.059	0.048	0.457	0.057	0.048
#4	0.087	0.078	0.043	0.059	0.051	0.054	0.059	0.058	0.067	0.039
#5	0.078	0.071	0.043	0.084	0.073	0.072	0.074	0.063	0.71	0.014
#6	0.227	0.223	0.043	0.257	0.205	0.272	0.282	0.251	0.248	0.072
#7	0.105	0.109	0.034	0.308	0.349	0.322	0.321	0.306	0.307	0.028
#8	0.091	0.092	0.243	0.256	0.249	0.294	0.281	0.242	0.243	0.090
#9	0.074	0.089	0.335	0.124	0.178	0.188	0.136	0.141	0.149	0.022
#10	0.036	0.045	0.042	0.086	0.082	0.079	0.092	0.079	0.085	0.034
Mean	0.087	0.086	0.091	0.174	0.135	0.145	0.141	0.170	0.198	0.042
SD	0.050	0.050	0.101	0.129	0.099	0.107	0.104	0.132	0.193	0.022

results as related to EGWO and AGWO in the majority of the scenarios. This demonstrates that the integration of CSA and BGWO has considerably increased the exploration aptitude and robustness.

6 Conclusion

Data classification in the medical diagnosis system supports physicians to learn hidden data samples by training a huge amount of real-world datasets. Diabetes mellitus is a severe global health problem. As stated in the International Diabetes Foundation report, now there are 300 million diabetic patients across the world, and an additional 300 million individuals are foretelling to be at greater possibility of diabetes in the year 2030. Therefore, it is indispensable for timely prediction, analysis, and management of hyperglycaemia and its consequences. We develop an effective system to diagnose diabetes disease using a hybrid optimization-based SVM. The proposed hybrid optimization technique integrates a CSA and BGWO for exploiting the full potential of SVM in the diabetes diagnosis system. The hybrid optimizer exploits the strengths of both CSA and BGWO techniques successfully to provide promising candidate solutions and realize global optima effectively. In this work, a better initial seed is selected by the CSA and then BGWO is used to appraise the locations of the searching agent in the discrete searching area for obtaining the optimum parameters with better accuracy. The effectiveness of our proposed hybrid optimization-based SVM (CS-BGWO-SVM) approach is rigorously evaluated on the technique, its performance is real-world database. To evaluate the CS-BGWO-SVM technique, its

performance is related to many advanced metaheuristic-based SVM approaches. The empirical analysis illustrates that CS-BGWO-SVM can be considered as a promising approach with outstanding performance measures.

Funding Authors received no specific funding for this study.

Availability of Data and Material The dataset for the experimentation is obtained from the National Institute of Diabetes and Digestive and Kidney Diseases. This dataset consists of 768 clinical records of Pima Indian heritage, a population living near Phoenix, Arizona, USA. There are eight attributes associated with this database (i.e., number of times pregnant, plasma glucose concentration estimated in 2 h after consuming a 75 g oral glucose, diastolic blood pressure, Triceps skinfold thickness, 2-h serum insulin, body mass index, diabetes nutrition function, and age). The Pima Indians Diabetes problem taken from Machine Learning Repository UCI: https://github.com/LamaHamadeh/Pima-Indians-Diabetes-DataSet-UCI/blob/master/pima_indians_diabetes.txt

Declarations

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Prokosch, H.-U., Ganslandt, T.: Perspectives for medical informatics: Reusing the electronic medical record for clinical research. *Methods Inf. Med.* **48**, 38–44 (2009)
- Dash, S., Shakyawar, S.K., Sharma, M., et al.: Big data in healthcare: management, analysis and future prospects. *J. Big Data* (2019). <https://doi.org/10.1186/s40537-019-0217-0>
- Nilashi, M., Ahmadi, N., Samad, S., Shahmoradi, L., Ahmadi, H., Ibrahim, O., Asadi, S., Abdullah, R., Abumalloh, R.A., Yadegaridehkordi, E.: Disease diagnosis using machine learning techniques: A review and classification. *J. Soft Comput. Decis. Support Syst.* **7**(1), 19–30 (2020)
- Dinesh, M.G., Prabha, D.: Diabetes mellitus prediction system using hybrid KPCA-GA-SVM feature selection techniques. *J. Phys.* **1767**(012001), 1–16 (2021). <https://doi.org/10.1088/1742-6596/1767/1/012001>
- Tama, B.A., Lim, S.: A Comparative performance evaluation of classification algorithms for clinical decision support systems. *Mathematics* (1814). <https://doi.org/10.3390/math8101814>
- Shaikh, M.S., Hua, C., Jatoti, M.A., Ansari, M.M., Qader, A.A.: Application of grey wolf optimisation algorithm in parameter calculation of overhead transmission line system. *IET Sci. Meas. Technol.* **15**(2), 218–231 (2021). <https://doi.org/10.1049/smt2.12023>
- Meng, Z., Li, G., Wang, X., Sait, S.M., Yıldız, A.R.: A comparative study of metaheuristic algorithms for reliability-based design optimization problems. *Arch. Comput. Methods Eng.* **28**, 1853–1869 (2021). <https://doi.org/10.1007/s11831-020-09443-z>
- Negi, G., Kumar, A., Pant, S., Pant, S., Ram, M.: GWO: A review and applications. *Int. J. Syst. Assur. Eng. Manag.* **12**, 1–8 (2021). <https://doi.org/10.1007/s13198-020-00995-8>
- Kao, Y.-T., Zahara, E.: A hybrid genetic algorithm and particle swarm optimization for multimodal functions. *Appl. Soft Comput.* **8**(2), 849–857 (2008)
- Tsai, J.-T., Liu, T.-K., Chou, J.-H.: Hybrid Taguchi-genetic algorithm for global numerical optimization. *IEEE Trans. Evol. Comput.* **8**(4), 365–377 (2004)
- Jitkongchuen, D.: A hybrid differential evolution with grey wolf optimizer for continuous global optimization. *Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)* (2015). <https://doi.org/10.1109/ICITEE.ED.2015.7408911>
- Nabil, E.: A modified flower pollination algorithm for global optimization. *Expert Syst. Appl.* **57**, 192–203 (2016)
- Tawhid, M.A., Ali, A.F.: A hybrid grey wolf optimizer and genetic algorithm for minimizing potential energy function. *Memet. Comput.* **9**(4), 347–359 (2017)
- Jayabarathi, T., Raghunathan, T., Adarsh, B., Suganthan, P.N.: Economic dispatch using hybrid grey wolf optimizer. *Energy* **111**, 630–641 (2016)
- Singh, N., Singh, S.: Hybrid algorithm of particle swarm optimization and grey wolf optimizer for improving convergence performance. *J. Appl. Math.* **2017**, 2030489 (2017). <https://doi.org/10.1155/2017/2030489>
- Gaidhane, P.J., Nigam, M.J.: A hybrid grey wolf optimizer and artificial bee colony algorithm for enhancing the performance of complex systems. *J. Comput. Sci.* **27**, 284–302 (2018)
- Hassanien, A.E., Rizk-Allah, R.M., Elhoseny, M.: A hybrid crow search algorithm based on rough searching scheme for solving engineering optimization problems. *J. Ambient Intell. Humaniz. Comput.* (2018). <https://doi.org/10.1007/s12652-018-0924-y>
- Oh, I.-S., Lee, J.-S., Moon, B.-R.: Hybrid genetic algorithms for feature selection. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(11), 1424–1437 (2004)
- Talbi, E.-G., Jourdan, L., Garcia-Nieto, J., Alba, E.: Comparison of population based metaheuristics for feature selection: Application to microarray data classification. *IEEE/ACS Int. Conf. Comput. Syst. Appl.* (2008). <https://doi.org/10.1109/AICCSA.2008.4493515>
- Panwar, D., Tomar, P., Singh, V.: Hybridization of Cuckoo-ACO algorithm for test case prioritization. *J. Stat. Manag. Syst.* **21**(4), 539–546 (2018). <https://doi.org/10.1080/09720510.2018.1466962>
- Zhao, F., Yao, Z., Luan, J., Song, X.: A novel fused optimization algorithm of genetic algorithm and ant colony optimization. *Math. Probl. Eng.* (2016). <https://doi.org/10.1155/2016/2167413>
- Babatunde, R.S., Olabiyisi, S.O., Omidiora, E.O.: Feature dimensionality reduction using a dual level metaheuristic algorithm. *Optimization* **7**(1), 49–52 (2014)
- Mafarja, M.M., Mirjalili, S.: Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* **260**, 302–312 (2017)
- Kazakov, P.: Extension for multi-objective genetic algorithms based on the dynamic population size model. *J. Phys.* **1661**, 012046 (2020). <https://doi.org/10.1088/1742-6596/1661/1/012046>
- Meraïhi, Y., Gabis, A.B., Mirjalili, S., Ramdane-Cherif, A.: Grasshopper optimization algorithm: theory variants, and applications. *IEEE Access* **9**, 50001–50024 (2021). <https://doi.org/10.1109/ACCESS.2021.3067597>
- Belmon, A.P., Auxillia, J.: An adaptive technique based blood glucose control in type-1 diabetes mellitus patients. *Int. J. Numer. Method Biomed. Eng.* **36**, e3371 (2020). <https://doi.org/10.1002/cnm.3371>
- Boussaïd, I., Lepagnot, J., Siarry, P.: A survey on optimization metaheuristics. *Inf. Sci.* **237**, 82–117 (2013)
- Chandrashekar, G., Sahin, F.: A survey on feature selection methods. *Comput. Electr. Eng.* **40**(1), 16–28 (2014)
- Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
- Scholkopf, B., Burges, C.J.C., Smola, A.J.: *Advances in Kernel Methods: Support Vector Learning*. The MIT Press, Cambridge (1998)
- Cristianini, N., Shawe-Taylor, J.: *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*. Cambridge University Press, Cambridge (2000)
- Mirjalili, S., Mirjalili, M.S., Lewis, A.: Grey wolf optimizer. *Adv. Eng. Softw.* **69**, 46–61 (2014)
- Muangkote, N., Sunat, K., Chiewchanwattana, S.: An improved grey wolf optimizer for training q-Gaussian radial basis functional link nets. *Int. Comput. Sci. Eng. Conf. (ICSEC)* **2014**, 209–214 (2014). <https://doi.org/10.1109/ICSEC.2014.6978196>
- Munro, C., Escobedo, R., Spector, L., Coppinger, R.P.: Wolf-pack (*Canis lupus*) hunting strategies emerge from simple rules in computational simulation. *Behav. Process.* **88**, 192–197 (2011)
- Askarzadeh, A.: A novel metaheuristic method for solving constrained engineering optimization problem: crow search algorithm. *Comput. Struct.* **169**, 1–12 (2016)
- Prior, H., Schwarz, A., Güntürkün, O.: Mirror-induced behavior in the magpie (*Picapica*): evidence of self-recognition. *PLoS Biol.* **6**(8), e202 (2008)
- Clayton, N., Emery, N.: Corvid cognition. *Curr. Biol.* **15**, R80–R81 (2005)
- Yang, X.S.: Metaheuristic optimization. *Scholarpedia* **6**, 11472 (2011)
- Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proc. Int. Jt. Conf. Artif. Intell.* **2**, 1137–1143 (1995)
- Alsewari, A.R.A., Zamli, K.Z.: ‘Design and implementation of a harmony-search-based variable-strength t-way testing strategy with constraints support.’ *Inf. Softw. Technol.* **54**(6), 553–568 (2012)

41. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**(1–3), 389–422 (2002)
42. Xuehao, Y., Yan, D.H., Jiabao, Y., Chao, L.: A novel SVM parameter tuning method based on advanced whale optimization algorithm. *J. Phys.* **1237**, 022140 (2019)
43. Chaabane, S.B., Kharbech, S., Belazi, A., Bouallegue, A.: Improved whale optimization algorithm for SVM model selection: Application in medical diagnosis. *Int. Conf. Softw. Telecommun. Comput. Netw. (SoftCOM)* **5**, 5 (2020). <https://doi.org/10.23919/SoftCOM50211.2020.9238265>
44. Anton, N., Dragoi, E.N., Tarcoveanu, F., Ciuntu, R.E., Lisa, C., Curteanu, S., Doroftei, B., Ciuntu, B.M., Chiseliță, D., Bogdănici, C.M.: Assessing changes in diabetic retinopathy caused by diabetes mellitus and glaucoma using support vector machines in combination with differential evolution algorithm. *Appl. Sci.* **11**(9), 3944 (2021). <https://doi.org/10.3390/app11093944>
45. Joshi, H., Arora, S.: 'Enhanced grey wolf optimization algorithm for global optimization.' *Fundam. Inform.* **153**(3), 235–264 (2017)
46. Qais, M.H., Hasanien, H.M., Alghuwainem, S.: 'Augmented grey wolf optimizer for grid-connected PMSG-based wind energy conversion systems.' *Appl. Soft Comput.* **69**, 504–515 (2018)
47. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 27 (2011)
48. UCI repository of bioinformatics Databases, Website: <http://www.ics.uci.edu/~mllearn/MLRepository.html>
49. Data World datasets repository. <https://data.world/>. Accessed 2018
50. Pardo, M., Sberveglieri, G.: Classification of electronic nose data with support vector machines. *Sens. Actuators B Chem.* **107**, 730–737 (2005)
51. Hao, S., Zhou, X., Song, H.: A new method for noise data detection based on DBSCAN and SVDD. In: *Proceedings of the 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, Shenyang, China, 8–12 June 2015; pp. 784–789
52. Ijaz, M.F., Alfian, G., Syafrudin, M., Rhee, J.: Hybrid prediction model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, synthetic minority over sampling technique (SMOTE), and random forest. *Appl. Sci.* **8**(8), 1325 (2018)
53. Fawcett, T.: An introduction to ROC analysis. *Pattern Recogn. Lett.* **27**(8), 861–874 (2006)
54. Derrac, J., García, S., Molina, D., Herrera, F.: A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm Evol. Comput.* **1**(1), 3–18 (2011)

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