IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 22, 2020, accepted December 27, 2020, date of publication January 6, 2021, date of current version January 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3049547

bSSA: Binary Salp Swarm Algorithm With Hybrid Data Transformation for Feature Selection

SAYAR SINGH SHEKHAWAT[1], HARISH SHARMA[1], SANDEEP KUMAR[2], ANAND NAYYAR[3,4], (Senior Member, IEEE), AND BASIT QURESHI[5], (Member, IEEE)
[1]Department of Computer Science and Engineering, Rajasthan Technical University, Kota 324010, India
[2]Department of Computer Science and Engineering, CHRIST (Deemed to be University), Bengaluru 560074, India
[3]Graduate School, Duy Tan University, Da Nang 550000, Vietnam
[4]Faculty of Information Technology, Duy Tan University, Da Nang 550000, Vietnam
[5]Department of Computer Science, Prince Sultan University, Riyadh 11586, Saudi Arabia

Corresponding authors: Anand Nayyar (anandnayyar@duytan.edu.vn) and Basit Qureshi (qureshi@psu.edu.sa)

This work was supported by the Robotics and IoT Laboratory, Prince Sultan University.

ABSTRACT Feature selection is a technique commonly used in Data Mining and Machine Learning. Traditional feature selection methods, when applied to large datasets, generate a large number of feature subsets. Selecting optimal features within this high dimensional data space is time-consuming and negatively affects the system's performance. This paper proposes a new binary Salp Swarm Algorithm (bSSA) for selecting the best feature set from transformed datasets. The proposed feature selection method first transforms the original data-set using Principal Component Analysis (PCA) and fast Independent Component Analysis (fastICA) based hybrid data transformation methods; next, a binary Salp Swarm optimizer is used for finding the best features. The proposed feature selection approach improves accuracy and eliminates the selection of irrelevant features. We validate our technique on fifteen different benchmark data sets. We conduct an extensive study to measure the performance and feature selection accuracy of the proposed technique. The proposed bSSA is compared to Binary Genetic Algorithm (bGA), Binary Binomial Cuckoo Search (bBCS), Binary Grey Wolf Optimizer (bGWO), Binary Competitive Swarm Optimizer (bCSO), and Binary Crow Search Algorithm (bCSA). The proposed method attains a mean accuracy of 95.26% with 7.78% features on PCA-fastICA transformed datasets. The results show that bSSA outperforms the existing methods for the majority of the performance measures.

INDEX TERMS Data transformation, fast independent component analysis, feature selection, principal component analysis, salp swarm optimizer.

I. INTRODUCTION

In the age of Big Data, the Internet of Things, and pervasive communication, massive amounts of unstructured and high-dimensional data are generated using millions of ubiquitous devices every day. Feature Selection (FS) is a vital preprocessing step that reduces the data dimensionality and eliminates redundant or irrelevant features. While redundant features add complexity, the irrelevant features barely discriminate samples from class in supervised FS or clusters in unsupervised FS, consequently hindering the performance [1], [2]. FS enhances the accuracy and improves the outcome's intelligibility by eliminating the noisy, useless, and unimportant data from high dimensional datasets [3]. Furthermore, FS reduces the over-fitting problem and classifier's complexity, making it easier to elucidate [4].

FS techniques are grouped into three main classes, namely Wrapper [5], [6] Filter [7], [8], and Embedded methods [9]. The filter methods are generally used in the preprocessing phase [10]. These methods do not use any machine learning classifiers to select the optimal features subset. Alternatively, these methods employ various statistical measures to choose the relevant features. Filter methods are computationally better than other relevant methods [2]. However, these methods show poor results if the data distribution is not uniform and features are highly correlated.

In contrast, wrapper based FS techniques use some predefined machine learning classifiers to measure the selected features [2]. In the wrapper-based approach, features are included-in or eliminated from an optimal feature-set

The associate editor coordinating the review of this manuscript and approving it for publication was Emre Koyuncu.

using recursive or backward feature elimination or forward-feature-selection techniques. These methods are computationally expensive compared to the filter methods and their complexity increases unexpectedly for datasets comprising of huge set of features. These methods select the optimal features during the training phase and hence are classified as embedded methods [11].

Meta-heuristic based FS methods show low consistency over high dimensional datasets. Many feature representation (data transformation) techniques like ICA, fastICA, PCA, etc. have been presented in the literature to tackle the consistency issue. The original set of features is converted into a preferable form in feature representation, making it suitable for different data mining functionalities.

PCA [12] is a feature transformation and dimensionality reduction technique that transform interrelated features into a set of unrelated features. Uğuz [13] presented a Genetic Algorithm (GA) based hybrid approach and information gain for selecting an optimal feature set from a dataset transformed by PCA. Further, Lee *et al.* [14] examined the three PCA variants: sparse PCA, independent PCA, and sparse independent PCA. A randomized PCA has been proposed by Rokhlin *et al.* [15] to identify principal components that maximize the variance. In continuation, a randomized nonlinear component analysis based on PCA and canonical correlation analysis (CCA) has been introduced by Lopez-Paz *et al.* [16] for data transformation. However, PCA shows poor performance on the blind source separation problem due to more complex dependencies among features [17], [18].

Consequently, to mitigate the issue mentioned above, ICA that transforms original features into a set of autonomous non-Gaussian features, are used. To get the data's insightful patterns, an Independent PCA (IPCA) has been developed by Yao *et al.* [19]. Pandey *et al.* [20] used a hybrid data transformation approach to improve clustering methods' efficacy. Kaya and Pehlivan [21] used ICA to reduce the unnecessary and irrelevant features from electrocardiogram (ECG) signals. Kaya *et al.* [22] analyzed the performance of GA, PCA, and fastICA based feature selection method and found that the GA outperformed the other methods. However, the original ICA is computationally expensive. Therefore, fastICA a novel variant of conventional ICA [23] is generally used.

It has been observed from the above discussion that the meta-heuristic and traditional FS methods suffer from stability problems and are very expensive in terms of computation. Feature selection methods based on meta-heuristics also suffer from premature convergence. Moreover, the existing data transformation does not eliminate the correlation and higher-order dependencies among features concurrently.

Hence, to overcome the limitations of data transformation and metaheuristic-based FS techniques, a new PCA-fastICA-based binary SSA (bSSA) has been proposed in this paper. The anticipated bSSA first converts the original dataset using PCA-fastICA based data transformation methods into a new set of features followed by feature selection using bSSA. In literature, a number of metaheuristic algorithms have been proposed to solve various real-world optimization problems and as per NFL theorem, new algorithms are good for some set of problems. Therefore, different algorithms have been proposed to solve different sets of problems. In this paper, an improved SSA has been used to solve the feature selection problem since it has been observed from the literature that the Salp swarm algorithm shows better efficacy for the feature selection problem. The proposed bSSA first converts the original dataset using PCA-fastICA based data transformation methods into a new set of features followed by feature selection using bSSA. To appraise the effectiveness of the proposed PCA-fastICA based bSSA, ten well-known datasets have been utilized. For the performance analysis, the newly developed method has also compared with bGA [24], bCSO [25], bGWO [26], [27], bCSA [28], and bBCS [29].

The significant research contribution of this work is as follows:

- This paper proposed a new variant of SSA and named Binary SSA.
- A hybrid data transformation approach is developed to transform the datasets by combining PCA and fastICA.
- The newly developed binary salp swarm optimizer deployed for selecting the best feature set from transformed datasets

The paper's organization is as follows; Previous work on data transformation and related terminologies are discussed in Section II, while Section III describes the SSA in detail. Section IV discusses the proposed feature selection technique followed by the experimental results in Section V. Lastly, the conclusion is drawn in Section VI.

II. LITERATURE REVIEW

FS methods are grouped into three categories viz supervised [30], [31], unsupervised [30], [32], and semi-supervised methods [33]–[35], based on the class label information. The supervised approach selects the optimal feature subset using the class label information [36], while unsupervised methods find the most optimal features using compactness and variance of clusters [37]–[39]. The semi-supervised methods [40] employ both types of information (labeled and unlabeled) to find the relevant features [40], [41]. The primary purpose is to improve the classifier's performance by eliminating the redundant and unimportant features that are opposite to each other. Da Silva *et al.* [42] treated this problem as a single objective function that enhances accuracy and reduces the cost.

Optimal features help to attain the highest classification accuracy. To obtain the relevant features from n features is a difficult task since for n features, 2^n subsets are possible [26], [43]. It will be even more difficult for high dimensional datasets. In such cases, traditional and exhaustive search (brute-force search) cannot find the best features set. Therefore, random search, heuristic search, and metaheuristic-based methods [43]–[47] are used as an alternative for selecting relevant features. Particle Swarm Optimization (PSO) [48], GA [24], Cuckoo Search (CS) [49], [50], Differential Evolution (DE) [51], Gravitational Search Algorithm (GSA) [52], etc. are some of the metaheuristics employed to solve feature selection problem.

Chen and Hsiao [53] presented a binary GA that uses an Support Vector Machine (SVM) for selecting the pertinent features and a GA-based co-operative technique proposed by Derrac *et al.* [54] for the same. Besides, Genetic Programming (GP) has also been employed to solve FS problems [55]. The GP-based feature selection techniques construct a feature tree with leaf nodes containing features. Neshatian and Zhang [56] introduced a wrapper selection method in which they used Naive Bayes (NB) to assess selected features' decency. The proposed FS method returns different feature set in different runs over imbalanced datasets. Some researchers used continuous and binary PSO for electing the best features. In the continuous PSO, a threshold β is used to select the relevant features, while in binary PSO, binary bit 0 means feature will not be selected, and 1 means the corresponding features will be selected. Lane *et al.* [57] used PSO along with statistical clustering for selecting essential features. However, existing PSO based feature selection algorithms don't show satisfactory performance on datasets having heterogeneous attributes. Tran *et al.* [58] employed a resetting mechanism based on *gbest* [59] to eliminate irrelevant and redundant features. Further, Xue *et al.* [60] enhanced the PSO-based feature selection method using the forward and backward feature selection policy.

Furthermore, Ke *et al.* [61] presented a binary Ant Colony Optimization (ACO) for selecting the dataset's critical features. OBoyle *et al.* [62] used SVM and ACO for selecting significant feature subset from the dataset. Khushaba *et al.* [51] developed ACO and DE based hybrid FS method in which DE finds the optimum features from the solutions returned by ACO. The proposed ACO-DE method returns a different set of features in different runs and also shows unstable behavior on high dimensional datasets. To mitigate the same, Ghosh *et al.* [63] used an adaptive DE to select relevant feature sets. Further, Ke *et al.* [64] introduced a multi-objective ACO based on the elitism strategy for selecting the relevant features. Yang *et al.* [65] developed an FS method by combining the memetic wrapper and Relief-F in which Relief-F sorts all the features and memetic wrapper selects the features using a score. Gu *et al.* [25] presented a bCSO for selecting the best features.

Further, a bGWO has been proposed by Emary *et al.* [26] to reduce unnecessary and irrelevant features. The proposed bGWO algorithm shows poor performance on very high dimensional datasets and sometimes also trap to local optima. Therefore, Pandey and Rajpoot [27] proposed a method based on GWO and Simulated Annealing (SA). Kang *et al.* [66] developed a hybrid approach that eliminates the outliers and returns the optimal feature subset. Pashaei and Aydin [67] proposed a binary black-hole algorithm for finding the appropriate features from biological datasets. To enhance the exploitation process and to find the relevant feature set, a hybrid method has been proposed by Mafarja and Mirjalili [1].

Cuckoo search has also been used for selecting the relevant features. Kaya [68] presented a binary cuckoo search to eliminate irrelevant features and improve classification accuracy. However, bCS algorithm returns more number of features if datasets contain highly correlated features. Therefore, to eliminate this issue, Pandey *et al.* [29] introduced bBCS that selects relevant features from transformed datasets. To eliminate the correlation, the proposed bBCS first transforms datasets using PCA-fastICA followed by feature selection. Sayed *et al.* [69] introduced a chaotic dragonfly algorithm to select the relevant features. Guha *et al.* [70] introduced a deluge-based GA for selecting optimal features. Taradeh *et al.* [71] proposed an evolutionary gravitational search-based method for selecting relevant features. Agrawal *et al.* [72] anticipated a quantum-based WOA for reducing extraneous and unnecessary features from the datasets. Ouadfel and Abd Elaziz [73] used an improved CSA to select the optimal features. Jia *et al.* [74] proposed a hybrid FS method based on spotted hyena optimization (SHO) and simulated annealing (SA). The proposed hybrid FS method shows better results over high-dimensional datasets but it shows unstable results on high-dimensional biomedical datasets. Ghosh *et al.* [75] used enhanced binary sailfish optimizer and β-hill climbing for finding relevant features from the dataset. Alweshah *et al.* [76] presented a hybrid mine blast FS method to enhance the quality of selected features. The proposed FS method is not tested on datasets that consist of more than 400 features. Alweshah *et al.* [77] introduced a binary monarch butterfly optimization algorithm (bMBO) for selecting optimal features from high dimensional datasets. The proposed bMBO could show inconsistent results on gigabyte-size datasets.

Recently, several multi-objective feature selection algorithms have also been proposed. Ke *et al.* [64] introduced a multi-objective ACO based on elitism strategy for selecting the relevant features. Li *et al.* [78] presented a multi-objective FS method using Bat algorithm and rebalancing strategy. The proposed algorithm shows satisfactory results on high dimensional datasets; however, it is computationally very expensive. Zhang *et al.* [79] introduced a multi-objective PSO for cost-based feature selection problems. The proposed FS method generates a Pareto front of nondominated solutions and finds a good set of features; however, it is not known that the achieved Pareto fronts could be improved further or not. Further, a cost-sensitive multi-objective feature selection approach based on artificial bee colony has been presented to group the features into non-dominated and dominated feature subsets [80]. The efficacy of the proposed FS approach may decrease if datasets contain a large number of features and are unbalanced. Hu *et al.* [81] proposed a multiobjective PSO for selecting relevant features using fuzzy cost. In general FS method give equal weightage to all the features; however, the proposed FS method creates a fuzzy dominance

TABLE 1. Considered datasets.

Sr.No.	Dataset	Number of Features	Number of Instances
1.	Breastcancer	9	699
2.	BreastEW	30	569
3.	CongressEW	16	435
4.	HeartEW	13	270
5.	IonosphereEW	34	351
6.	SonarEW	60	208
7.	WineEW	13	178
8.	Vote	16	300
9.	Tic-tac-toe	9	958
10.	Zoo	16	101
11.	SpectEW	22	267
12.	PenglungEW	325	73
13.	Prostate Tumor	10509	102
13.	11 Tumors	12533	174
15.	Lung Cancer	12600	203

relationship to compare the decency of candidate particles and defines a fuzzy crowding distance to discard the elitist archive.

Recently Elhariri *et al.* [82] proposed a multi-objective variant of the SSA and deployed with a hybrid of filter-wrapper based FS approach. The new approach used for recognition of crack severity and achieved significant improvement in accuracy. Chen *et al.* [83] developed a new variant of SSA with enhanced global search capability by deploying concept of spark programming and tested for three poplar data-sets.

A. PRINCIPAL COMPONENT ANALYSIS

PCA [12] is a popular feature transformation approach that converts correlated features into unrelated features, also called principal components. PCA always returns the principal components equal to or less than the total features in the data set. In general, PCA reduces the data dimensionality and maximizes the variance. The maximum part of the variance covered by the largest component is the PCA'a first principal component. To calculate the principal components, a covariance matrix is created. This paper used a PCA based on covariance matrix eigen-decomposition that discovers the principal component using four steps.

- Create the covariance matrix.
- Find the Eigenvalues.
- Find the Eigen-vector that represents the direction of principal components.
- Get the coordinates of each data in the principal component's direction.

Suppose a dataset D_x has n rows and k columns, where n and k denotes the number of instances and features in the dataset, accordingly. To transform D_x, into transformed matrix D_y, a covariance matrix C is created. In this matrix, covariances are placed at off-diagonal locations, and variances are placed at diagonal places. PCA reduces the correlation between the transformed matrix variables, which means that the correlation of different characteristics in the matrix, C, should be close to zero. Diagonal entries in the C matrix must be maximized and off-diagonal values minimized.

In addition, Eigenvalues are computed by solving the equation $|C - \lambda I| = 0$. In the subsequent step, Eigenvectors are computed according to Eq. (1). Finally, the Eigenvector matrix is multiplied with the original matrix D_x to find the transformed dataset.

$$[A - \lambda_j I] \times [x] = [0], \quad \text{for j} = 1, 2 \ldots, \text{n}. \tag{1}$$

B. FAST INDEPENDENT COMPONENT ANALYSIS

The fastICA [84] converts dependent multivariate signals into non-Gaussian independent signals. As such, PCA only reduces the correlation between features, not the higher-order dependence between attributes. Therefore, to reduce higher-order dependencies among features, fastICA is employed. In the fastICA, each component is treated with equal importance, whereas in PCA, some attributes are measured with high importance [18]. The considered elements' statistical autonomy is maximized for finding the independent components of fastICA [17]. Independence in fastICA is achieved by either maximizing the non-gaussianity or Minimizing Mutual Information (MMI) [85]. The fastICA is a competent approach for evaluating ICA, and it pertains to the family of fix-point ICA. A fixed point iteration policy is used by fastICA to measure the rotating components of data. The fastICA is very effective and computes independent elements 10 to 100 times quicker than the original ICA [23]. Furthermore, the fastICA also converges quickly than the original ICA.

The proposed approach uses PCA and fastICA to eliminate the correlation and higher-order data dependencies among features followed by feature selection using bSSA. This section presented an overview of PCA, fastICA, and SSA.

III. SALP SWARM ALGORITHM

Salps are cylindrical marine tunics that usually live in groups or in chains and floats deliberately. It has been confirmed in the literature that the movement of the salp swarm attains a better locomotion and find food easily with this approach. SSA is introduced by Mirjalili *et al.* [86] to solve various optimization problems. All the salp form a salp chain in SSA to find the target or food source in search space [87]. To mathematically model the chains of the salp, the swarm is slashed into two parts, leader and follower. The leading salp is always in front and guides the other follower in the chain.

Consider $Y = \{Y_1, Y_2, \ldots, Y_i, \ldots, Y_n\}$ is a group of n salps and F_s denotes the target vector or source of food. Every salp in salp chain is represented in d dimensional vector $(Y_i = y_1, y_2, \ldots, y_d)$. The position of leader salp in the chain is updated using Eq. (2)

$$Y_1 = \begin{cases} F_s + \alpha_1((Y_{max} - Y_{min})\alpha_2 + Y_{min}) & \alpha_3 \geq 0 \\ F_s - \alpha_1((Y_{max} - Y_{min})\alpha_2 + Y_{min}) & \alpha_3 < 0 \end{cases} \tag{2}$$

where, α_1, α_2, and α_3 are the random values, Y_1 is the location of a front salp (leader), Y_{max}, and Y_{min} are the upper and lower boundary values for each salp. In SSA, balancing between exploration and exploitation controlled by α_1 and its value is

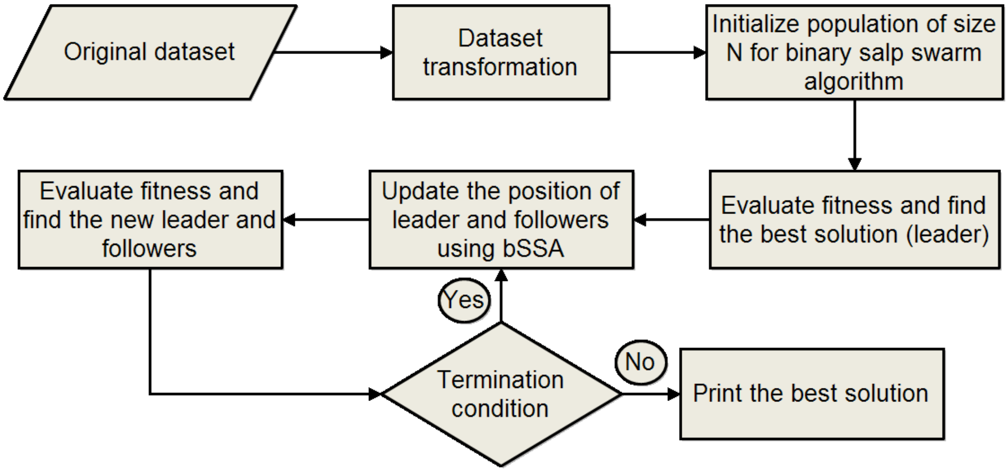


FIGURE 1. Flowchart of the proposed bSSA.

updated in every cycle using Eq. (3).

$$\alpha_1 = 2e^{-\left(\frac{4*c_{iter}}{Max_{iter}}\right)^2} \tag{3}$$

where, Max_{iter} and c_{iter} represents the total and current iteration, respectively.

The follower salps' positions except Y_1 are improved using Newton's law of motion as stated by Eq. (4).

$$Y_j(i) = \frac{1}{2}at^2 + v_0 t \tag{4}$$

where $j = 2, 3, \ldots, n$ and $Y_j(i)$ indicates the i^{th} dimension of j^{th} salp, v_0, t, and a is the initial velocity, time, and acceleration respectively, which are computed by the following equation.

$$a = \frac{v_{end}}{v_0} \quad where\ v = \frac{y - y_0}{t}$$

For the optimization problems, time denotes the iteration, and the initial velocity is 0. Thus the positions of followers are updated using modified equation given in Eq. (5).

$$Y_j(i) = \frac{1}{2}\left(Y_j(i) + Y_{j-1}(i)\right) \tag{5}$$

The detailed steps of SSA are discussed in Algorithm 1. In SSA, the initial population is generated arbitrarily, and the best fitted salp is treated as a food source, and other salps move towards the food source. A food source (F_s) updates its position in every iteration.

IV. PROPOSED BINARY SALP SWARM FEATURE SELECTION METHOD

The proposed binary SSA uses two steps for finding the relevant features from datasets. In the first step, the dataset is transformed using PCA-fastICA based feature transformation method followed by binary SSA in the second step for electing optimal features. This section discusses each step of the new feature selection approach. The proposed system model of bSSA is illustrated in Fig. 1.

Algorithm 1 Salp Swarm Algorithm

```
Input: Engender initial population ($Y_i$), where $i$ = $1, 2, \ldots, N$ randomly
Output: The solution vector $F_s$
while stopping criterion do
    Compute fitness and find the current best salp
    Calculate $\alpha_1$ using to Eq. (3)
    for each salp ($X_j$) in salp chain do
        if ($j$ == 1) then
            Modify position of leader salp using Eq. (2)
        else
            Modify locations of follower salps according to Eq. (5)
        end if
    end for
end while
Print the best solution $F_s$
```

A. FEATURE TRANSFORMATION

This phase transforms the original features into a new feature set through PCA and fastICA data transformation method. The process of data transformation uses the following two steps to convert the original dataset.

1) In the first phase, PCA transforms the original features into new features, also known as principal components. Generally, dimensionality reduction performed by using PCA.
2) The second phase get input from PCA in form of principal components and convert them into new feature set using fastICA. The features returned by FastICA are input into the proposed bSSA to find related features.

PCA only reduces correlation among features and does not eliminate high order dependencies between features.

TABLE 2. Parameter settings of all the considered algorithm.

Sr.No.	Parameter	bCSA	bCSO	bGA	bGWO	bBCS	bSSA
1.	Population Size (N)	50	50	50	50	50	50
2.	Number of iterations (itr)	1000	1000	1000	1000	1000	1000
3.	Number of search agents	8	8	8	8	8	8
4.	Probability (P_a)	–	–	–	–	$[0.05, 0.5]$	–
5.	Step scaling factor (α)	–	–	–	–	$[0.01, 0.5]$	–
6.	Number of iterations	70	70	70	70	70	70
7.	Search domain	[0, 1]	[0, 1]	[0, 1]	[0, 1]	[0, 1]	[0, 1]
8.	Number of repeatative runs	20	20	20	20	20	20
9.	α parameter	–	–	–	0.99	–	–
10.	β parameter	–	–	–	0.01	–	–
11.	Crossover Probability	–	–	0.75	–	–	–
12.	Mutation Probability	–	–	0. 02	–	–	–
13.	Social factor (ϕ)	–	0.2	–	–	–	–
14.	Max Velocity	–	6	–	–	–	–
15.	Flight length(fl)	2	–	–	–	–	–
16.	Awareness probability(AP)	0.1	–	–	–	–	–
17.	Control parameter C_1	–	–	–	–	–	[0, 1]
18.	Control parameter C_2	–	–	–	–	–	[0, 1]

Thus, to reduce the dependencies among features, fastICA is used. However, the fastICA does not remove the correlation between two features. Therefore, to eliminate the correlation and higher-order dependencies among features, a combined data transformation method based on PCA-fastICA has been used. After converting the data set, bSSA is used to select what is important and relevant from the transformed data sets.

B. BINARY SALP SWARM ALGORITHM

In standard SSA, solution individuals are generated in continuous search space and they contain continuous values while in FS problem search region is binary or discrete (0 and 1). Hence, a binary variant of SSA has been proposed. The anticipated bSSA convert the continuous values into binary after each iteration using a threshold δ. The value of δ is empirically set to 0.6 by testing its efficacy for different δ between, 0 to 1. Consider $X_i, (i = 1, 2, 3, \cdots, d)$ is search individual having d dimension. In this approach, a specific feature or dimension will only be included in the selected feature list and (the value is set to 1), if $X_i \geq \delta$ else set to 0. Suppose $X_i = (0.35, 0.63, 0.55, 0.64, 0.74, 0.15)$ is a search agent with 6 features. The proposed bSSA converts each dimension of search individual in binary using threshold $\delta = 0.6$. The first feature is 0.35 which is smaller than 0.6 so, it is not selected while the second is 0.63 which is larger hence it is selected. The final search agent after converting all the dimension into binary will be $X_i = (0, 1, 0, 1, 1, 0)$. The Algorithm 2 demonstrate the pseudo-code of bSSA.

Algorithm 2 Binary Salp Swarm Algorithm

```
Input: Engender initial population ($Y_i$), where $i = 1, 2, \ldots, N$ randomly
Output: Selected feature List $F$
Convert initial population into binary using threshold $\delta$
while stopping criterion do
    Compute fitness and find the current best
    Calculate $\alpha_1$ using to Eq. (3)
    for each salp ($X_j$) in salp chain do
        if ($j == 1$) then
            Modify position of leader salp using Eq. (2) and convert them into binary using threshold $\delta$
        else
            Update locations of follower salps according to Eq. (5) and convert them into binary using threshold $\delta$
        end if
    end for
end while
Print the optimal features
```

C. FEATURE SELECTION USING BINARY SALP SWARM

This paper introduces a wrapper based binary SSA for selecting the relevant features. The wrapper feature selection approach uses predefined machine learning classifiers to elect the relevant features from high-dimensional datasets. SVM, k-Neural Network (KNN), Artificial Neural Network (ANN) are some of the commonly used classifiers. KNN classifier is generally employed to solve the FS problem because of its simplicity. In this article, a KNN classifier has been used to measure the goodness of identified characteristics. The objective of FS is to improve/maximize the accuracy and minimize

TABLE 3. The validation accuracy of bSSA on datasets with feature selection.

Dataset	Mean Accuracy					
	bGA	bCSO	bGWO	bCSA	bBCS	bSSA
Breastcancer	95.89	95.59	95.72	95.00	96.51	96.85
BreastEW	94.48	92.72	94.04	93.99	96.62	97.12
CongressEW	95.59	93.47	94.17	94.57	96.88	96.43
HeartEW	80.45	75.88	76.06	75.25	76.68	80.45
IonosphereEW	91.67	83.50	85.75	86.19	88.26	91.91
SonarEW	90.46	80.65	81.79	84.12	87.26	92.19
WineEW	92.33	90.51	91.90	92.37	97.47	97.90
Vote	93.33	90.37	92.51	92.28	92.73	93.13
Tic-tac-toe	75.41	74.43	73.83	75.33	77.22	79.00
Zoo	96.89	96.59	97.70	96.00	96.71	97.16
SpectEW	82.26	82.42	82.76	83.14	83.83	84.28
PenglungEW	90.42	89.98	90.23	90.90	91.15	91.42
Prostate Tumor	91.54	91.23	91.42	92.00	92.46	92.36
11 Tumors	92.02	91.88	92.26	92.66	93.08	93.30
Lung Cancer	93.24	93.24	93.32	94.80	94.47	94.45
Overall Mean	90.40	88.16	88.90	89.24	90.75	91.86

TABLE 4. The mean accuracy of bSSA on original datasets (without feature selection).

Dataset	Mean Accuracy					
	bGA	bCSO	bGWO	bCSA	bBCS	bSSA
Breastcancer	94.94	94.74	94.91	94.11	95.56	**95.89**
BreastEW	93.54	91.89	93.25	93.11	95.67	**96.16**
CongressEW	94.64	92.64	93.38	93.68	**95.93**	95.48
HeartEW	79.65	75.20	75.42	74.54	75.93	**79.65**
IonosphereEW	90.76	82.76	85.03	85.38	87.39	**91.00**
SonarEW	89.56	79.93	81.10	83.33	86.40	**91.28**
WineEW	91.42	89.70	91.13	91.50	96.51	**96.93**
Vote	**92.41**	89.56	91.73	91.41	91.82	92.21
Tic-tac-toe	74.66	73.77	73.21	74.62	76.46	**78.22**
Zoo	95.93	95.73	96.88	95.10	95.76	**96.20**
SpectEW	81.45	81.68	82.06	82.36	83.01	**83.45**
PenglungEW	89.52	89.18	89.47	90.04	90.26	**90.51**
Prostate Tumor	90.63	90.42	90.65	91.13	**91.55**	91.45
11 Tumor	91.11	91.06	91.48	91.79	92.17	**92.38**
Lung Cancer	92.32	92.41	92.53	**93.91**	93.54	93.51
Overall Mean	89.13	88.55	88.97	89.49	90.05	**90.43**

the error rate and number of features selected. The proposed bSSA uses the objective function discussed in Eq. (6) for selecting the best feature set [26]. The fitness function defined in Eq. (6) is maximized while selecting relevant features.

$$Fitness = \alpha_1 \gamma_x(D) + \beta_1 \frac{|S - F_s|}{|S|} \tag{6}$$

where $\gamma_x(D)$ is accuracy related to decision D. F_s represents the selected features, S is the total count of features, $\alpha_1 \in [0, 1]$ and β_1 are constants which represent the importance of the classification accuracy and the subset length respectively. The sum values of α_1 and β_1 must be 1 ($\alpha_1 + \beta_1 = 1$).

Eq. (6) can be changed into a minimization problem using error values and ratio of elected feature. The error minimizing fitness function described in Eq. (7).

$$Fitness = \alpha_1 E_R(D) + \beta_1 \frac{|F_s|}{|S|} \tag{7}$$

where, $E_R(D)$ is error rate.

V. RESULTS AND DISCUSSION

The efficacy of the anticipated bSSA has been evaluated on 15 benchmark FS datasets [88]. Table 1 presents a complete delineation of all the data sets considered. In this paper,

TABLE 5. The mean accuracy and mean number of selected features on original test datasets (20%) (without feature transformation).

Dataset	Mean Accuracy						Mean number of selected features					
	bGA	bCSO	bGWO	bCSA	bBCS	bSSA	bGA	bCSO	bGWO	bCSA	bBCS	bSSA
Breastcancer	96.90	96.86	96.80	96.41	97.22	**97.69**	6.11	6.41	6.53	6.35	5.60	**5.12**
BreastEW	95.47	93.95	95.10	95.38	**97.13**	97.07	19.27	18.34	20.05	14.47	15.62	**11.51**
CongressEW	96.59	94.71	95.23	95.97	96.92	**97.03**	6.11	7.24	8.33	7.28	6.15	**4.35**
HeartEW	81.28	76.87	76.92	76.35	77.08	**84.83**	8.26	7.34	6.85	9.32	6.62	**6.04**
IonosphereEW	92.63	84.61	86.72	87.46	88.72	**96.31**	16.16	18.22	20.79	21.47	14.55	**11.06**
SonarEW	91.40	81.72	82.70	85.36	87.71	**95.70**	36.28	46.21	39.75	40.78	45.99	**29.81**
WineEW	93.30	91.70	92.93	93.72	97.97	**98.29**	9.26	8.77	9.49	8.86	8.96	**6.81**
Vote	94.31	91.55	93.54	93.63	93.21	**95.69**	7.12	8.41	7.53	7.35	6.61	**4.12**
Tic-tac-toe	76.19	75.42	74.66	76.43	77.62	**78.98**	7.33	7.41	8.54	7.35	7.61	**5.12**
Zoo	97.90	97.86	97.22	97.41	98.80	**98.13**	6.33	6.74	6.36	6.66	5.45	**5.35**
SpectEW	83.49	83.72	84.11	84.42	85.09	**85.54**	9.43	9.61	9.46	9.37	9.53	**9.18**
PenglungEW	91.13	90.79	91.08	91.66	**91.88**	92.14	188.41	138.74	128.81	138.45	121.21	**120.46**
Prostate Tumor	92.71	92.50	92.73	93.23	93.66	**93.55**	5143.23	2353.17	3825.17	2148.47	2219.22	**2216.95**
11 Tumors	93.57	93.52	93.95	94.27	94.66	**94.87**	6138.33	5623.87	4671.07	4293.63	4251.36	**4201.85**
Lung Cancer	94.54	94.63	94.75	96.16	**95.78**	95.75	6144.03	4792.83	4641.17	3426.43	3416.95	**3395.11**
Overall Mean	92.13	89.36	89.90	90.53	91.56	**93.44**	1183.04	870.22	893.99	676.41	675.50	**668.86**

TABLE 6. Results for PCA-based transformed (20%) of test datasets.

Dataset	Mean Accuracy						Mean number of selected features					
	bGA	bCSO	bGWO	bCSA	bBCS	bSSA	bGA	bCSO	bGWO	bCSA	bBCS	bSSA
Breastcancer	97.00	96.94	96.73	96.53	97.32	**97.76**	6.31	6.16	6.46	6.60	5.41	**5.11**
BreastEW	95.73	94.31	95.24	95.61	96.72	**97.09**	18.35	19.23	19.39	13.87	14.57	**11.79**
CongressEW	96.59	94.73	95.19	95.93	96.90	**97.32**	6.11	7.27	8.29	7.24	6.21	**4.29**
HeartEW	81.43	77.00	76.81	76.48	77.20	**84.59**	8.44	7.30	6.81	9.24	6.58	**6.01**
IonosphereEW	92.74	87.42	86.85	88.03	92.59	**96.86**	16.09	17.22	19.50	21.28	14.32	**11.00**
SonarEW	91.50	84.52	83.13	85.82	88.02	**96.00**	34.22	43.22	36.12	41.34	39.12	**26.79**
WineEW	94.33	92.52	93.64	94.14	98.11	**99.09**	9.14	8.64	9.44	8.75	8.73	**6.65**
Vote	94.56	91.58	93.56	93.70	93.44	**97.73**	7.06	8.34	7.45	7.28	6.53	**4.21**
Tic-tac-toe	76.38	75.53	74.71	76.56	77.69	**80.08**	7.25	7.36	8.47	7.28	7.47	**5.09**
Zoo	97.97	97.91	98.87	97.50	97.53	**99.10**	6.29	6.67	6.30	6.58	5.39	**5.38**
SpectEW	84.74	84.98	85.37	85.69	86.36	**86.82**	9.29	9.47	9.32	9.23	9.39	**9.04**
PenglungEW	92.32	91.97	92.26	92.85	93.08	**93.34**	186.53	137.35	127.52	137.07	120.00	**119.26**
Prostate Tumor	93.64	93.42	93.66	94.16	94.59	**94.61**	5086.65	2327.29	3783.09	2124.84	2185.91	**2182.56**
11 Tumors	94.88	94.83	95.27	95.59	95.98	**96.20**	6064.67	5556.38	4615.02	4242.11	4200.34	**4151.43**
Lung Cancer	95.67	95.76	95.89	97.32	**96.93**	96.90	6058.01	4725.73	4576.19	3378.46	3369.11	**3347.58**
Overall Mean	91.97	90.23	90.48	91.06	92.16	**94.22**	1168.29	859.18	882.62	668.07	666.61	**660.41**

the KNN classifier has been used by the proposed bSSA to select the optimal feature subset. The k value of KNN is empirically set to 4 and the data set is divided into training, validation, and test sets based on $K-fold$ cross-validation. To validate the efficiency of the proposed bSSA, $K-1$ fold has been used for training and validation and the rest for testing. This process is repeated N times. So, each method is evaluated $K \times N$ times for each data set. The proposed model uses 60% data for training, 20% for validation, and remaining 20% for testing. The data in training, validation, and test set will be completely different which reduces the chances of over-fitting.

The efficiency of the proposed bSSA has been evaluated in terms of average accuracy, average count of selected features, mean computational time, mean error rate. Further, to experimentally validate its performance, it is compared with bGA [24], bCSO [25], bGWO [26], bCSA [28], and bBCS [29]. The experiments are performed using a computer with 2.35 GHz with Core i3 processor, 8 GB of RAM and Matlab 2017b. The parameter settings of considered

TABLE 7. Results for fastICA-based transformed (20%) of test datasets.

Dataset	Mean Accuracy						Mean number of selected features					
	bGA	bCSO	bGWO	bCSA	bBCS	bSSA	bGA	bCSO	bGWO	bCSA	bBCS	bSSA
Breastcancer	96.98	96.97	96.95	96.60	97.19	**97.93**	6.02	6.26	6.32	6.10	5.47	**5.03**
BreastEW	95.55	94.04	95.24	95.57	97.30	**97.47**	18.98	17.91	19.40	13.91	15.24	**11.16**
CongressEW	96.67	94.81	95.37	96.16	97.08	**97.56**	6.02	7.07	8.06	6.99	6.00	**4.08**
HeartEW	81.35	76.96	77.04	76.50	77.21	**85.06**	8.13	7.17	6.63	8.96	6.45	**5.88**
IonosphereEW	92.70	84.70	86.85	87.64	88.87	**96.66**	15.93	17.79	20.12	20.64	14.20	**10.90**
SonarEW	91.47	81.80	82.82	85.53	87.86	**96.10**	35.75	45.15	38.47	39.20	44.85	**29.25**
WineEW	93.37	91.80	93.07	93.91	98.14	**99.39**	9.12	8.57	9.18	8.51	8.73	**6.47**
Vote	94.38	91.65	93.68	93.82	93.37	**97.95**	7.01	8.22	7.29	7.06	6.44	**3.97**
Tic-tac-toe	76.26	75.51	74.77	76.59	77.75	**80.20**	7.21	7.24	8.26	7.06	7.42	**4.93**
Zoo	97.85	97.90	98.95	97.62	97.39	**99.02**	6.23	6.58	6.15	6.39	5.32	**5.11**
SpectEW	85.29	85.53	85.93	86.24	86.92	**87.98**	9.21	9.39	9.24	9.15	9.31	**8.97**
PenglungEW	92.92	92.56	92.86	93.46	93.68	**93.94**	184.75	136.05	126.31	135.76	118.86	**118.12**
Prostate Tumor	94.25	94.03	94.27	94.77	**95.21**	95.10	5048.50	2309.83	3754.72	2108.90	2169.51	**2166.12**
11 Tumors	95.50	95.44	95.88	96.21	96.61	**96.83**	6013.12	5509.15	4575.79	4206.05	4164.64	**4116.14**
Lung Cancer	96.29	96.39	96.51	97.95	97.56	**97.63**	6000.46	4680.84	4532.72	3346.36	3337.11	**3315.78**
Overall Mean	92.05	90.01	90.67	91.24	91.24	**94.54**	1158.43	851.81	875.24	662.07	661.30	**654.79**

TABLE 8. Performance analysis of the proposed and existing methods on PCA-fastICA based transformed (20%) of test datasets.

Dataset	Mean Accuracy						Mean number of selected features					
	bGA	bCSO	bGWO	bCSA	bBCS	bSSA	bGA	bCSO	bGWO	bCSA	bBCS	bSSA
Breastcancer	97.16	97.06	97.10	96.86	97.43	**98.31**	5.89	5.97	5.83	5.94	5.07	**4.67**
BreastEW	96.00	94.73	95.97	96.10	96.96	**98.03**	17.24	18.06	18.11	13.06	13.20	**10.76**
CongressEW	97.00	95.80	95.86	96.19	97.29	**98.16**	6.02	6.53	8.01	6.94	5.95	**3.85**
HeartEW	82.18	77.13	77.07	77.00	77.93	**85.43**	7.84	7.10	5.99	8.09	5.21	**5.03**
IonosphereEW	93.17	87.99	87.42	88.43	93.11	**97.62**	15.64	16.77	18.73	20.60	13.46	**10.69**
SonarEW	92.17	85.09	83.92	86.39	88.52	**96.83**	33.17	41.20	35.21	40.62	42.29	**23.21**
WineEW	94.83	93.17	94.29	94.86	98.91	**99.43**	9.01	8.11	9.09	8.02	7.94	**5.98**
Vote	95.11	92.18	93.96	94.20	93.86	**98.12**	6.64	7.11	7.01	6.87	6.19	**3.97**
Tic-tac-toe	77.21	76.60	75.10	77.18	78.20	**81.27**	6.95	7.04	7.97	6.77	7.10	**4.72**
Zoo	98.03	98.25	99.01	98.12	97.75	**99.40**	6.01	5.78	5.98	6.12	5.23	**4.98**
SpectEW	85.84	86.09	86.49	86.80	89.49	**93.95**	9.13	9.31	9.16	9.07	9.23	**8.89**
PenglungEW	93.52	93.16	93.47	94.06	94.29	**94.85**	183.00	134.76	125.11	134.47	117.73	**116.01**
Prostate Tumor	94.86	94.64	94.88	95.39	95.83	**96.72**	5010.64	2292.51	3726.56	2093.08	2153.24	**2139.80**
11 Tumors	96.12	96.06	96.51	96.83	97.24	**98.46**	5962.01	5462.33	4536.90	4170.30	4129.24	**4001.15**
Lung Cancer	96.92	97.01	97.14	98.59	98.20	**98.77**	5943.46	4636.37	4489.66	3314.57	3305.40	**3204.28**
Overall Mean	92.68	91.00	91.21	91.80	92.80	**95.99**	1148.18	843.93	867.29	656.30	655.09	**648.60**

algorithms have been depicted in Table 2. As metaheuristic algorithms return different values in different runs, to reduce the randomization effect each algorithm has been executed 30 times and their mean values are used for comparison. For better comparison and show the importance of feature representation, the performance of bSSA has been evaluated on original test datasets (with and without optimal features) and transformed test datasets using PCA, fastICA and PCA-fastICA and depicted in Tables 4 - 8. The mean accuracy of all the considered methods over original datasets (with all features) have been tabulated in Table 4. It is pertinent from the table that the proposed bSSA achieves the optimal accuracy of 91.30%. The average number of features selected and accuracy of the bSSA over original datasets with optimal features are shown in Table 5. It is seen in the table that bSSA returns the best accuracy using the lowest number of features for 90% of the datasets.

Furthermore, to evince the importance of feature transformation, the proposed bSSA has been tested on PCA, fastICA, and PCA-fastICA based transformed datasets. Table 6 - Table 8 depicts the mean number of features selected and accuracy of PCA, fastICA, and PCA-fastICA

TABLE 9. The mean computational time (in seconds) of original and PCA-fastICA-based transformed datasets.

Dataset	Standard dataset						Transformed dataset					
	bGA	bCSO	bGWO	bCSA	bBCS	bSSA	bGA	bCSO	bGWO	bCSA	bBCS	bSSA
Breastcancer	46.53	45.68	43.56	46.75	48.08	**41.83**	45.01	46.53	42.73	46.03	48.60	**41.16**
BreastEW	39.19	40.67	35.71	41.11	38.15	**37.73**	41.06	40.14	36.76	41.27	38.36	**36.16**
CongressEW	39.31	39.40	37.88	41.71	37.58	**36.98**	36.83	38.00	39.10	41.08	36.87	**35.04**
HeartEW	30.84	30.41	27.39	29.85	30.62	**24.75**	28.82	27.90	27.39	29.90	27.14	**24.59**
IonosphereEW	28.13	29.11	27.17	28.96	31.26	**24.71**	26.67	28.79	28.34	29.02	31.15	**24.55**
SonarEW	26.67	30.55	25.93	27.90	30.47	**24.53**	26.89	27.27	30.37	28.21	31.92	**25.28**
WineEW	17.44	26.11	31.43	24.12	29.02	**19.57**	29.84	29.36	25.13	23.51	20.32	**17.16**
Vote	21.88	21.06	23.93	20.48	20.81	**21.44**	17.45	18.79	18.01	20.25	19.04	**14.40**
Tic-tac-toe	36.87	25.87	25.08	25.01	29.82	**26.66**	27.93	30.23	28.79	24.52	26.12	**22.25**
Zoo	18.01	18.82	19.83	18.39	17.84	**14.29**	16.34	17.70	16.71	19.05	15.78	**13.38**
SpectEW	19.14	20.01	21.08	19.55	**14.96**	15.19	17.37	18.82	17.76	20.25	16.77	**14.22**
PenglungEW	47.04	46.18	44.04	47.26	48.61	**42.29**	45.51	47.04	43.20	46.54	49.13	**41.61**
Prostate Tumor	9607.87	5991.01	8070.56	5026.22	5012.31	**4911.03**	9545.42	5952.07	8018.10	4993.55	4979.73	**4879.11**
11 Tumors	27128.44	21700.83	21135.63	18603.61	18406.21	**17985.32**	26965.67	21570.63	21008.82	18491.99	18295.77	**17877.41**
Lung Cancer	33490.25	26100.69	26788.26	21815.44	21459.25	**20963.25**	33205.58	25878.83	26560.56	21630.01	21276.85	**20785.06**
Overall Mean	4706.51	3611.10	3757.17	3054.42	3017.27	**2945.97**	4671.76	3584.81	3739.45	3032.35	2994.24	**293.43**

TABLE 10. The mean error values returned by the proposed bSSA over original and PCA-fastICA-based transformed datasets.

Dataset	Standard dataset						Transformed dataset					
	bGA	bCSO	bGWO	bCSA	bBCS	bSSA	bGA	bCSO	bGWO	bCSA	bBCS	bSSA
Breastcancer	0.0415	0.0404	0.0404	0.0394	0.0404	**0.0384**	0.0404	0.0404	0.0415	0.0404	0.0415	**0.0344**
BreastEW	0.0364	0.0374	0.0293	0.0303	0.0283	**0.0364**	0.0404	0.0394	0.0415	0.0384	0.0374	**0.0283**
CongressEW	0.0324	0.0364	0.0354	0.0313	0.0303	**0.0283**	0.0344	0.0334	0.0324	0.0344	0.0293	**0.0283**
HeartEW	0.1284	0.1324	0.1365	0.1304	0.1304	**0.1274**	0.1385	0.1324	0.1314	0.1335	0.1304	**0.1254**
IonosphereEW	0.1021	0.1223	0.0920	0.1021	0.0900	**0.0698**	0.1112	0.1223	0.1001	0.1062	0.1072	**0.0708**
SonarEW	0.1031	0.1021	0.0981	0.1132	0.0971	**0.0728**	0.1031	0.1041	0.1031	0.1001	0.0849	**0.0718**
WineEW	0.0293	0.0313	0.0354	0.0394	0.0425	**0.0303**	0.0354	0.0334	0.0354	0.0415	0.0344	**0.0273**
Vote	0.0415	0.0293	0.0394	0.0455	0.0394	**0.0303**	0.0435	0.0465	0.0374	0.0384	0.0324	**0.0283**
Tic-tac-toe	0.2143	0.2224	0.2538	0.2234	0.2133	**0.2113**	0.2032	0.2133	0.2012	0.2153	0.2052	**0.1860**
Zoo	0.0415	0.0313	0.0364	0.0384	0.0394	**0.0263**	0.0384	0.0415	0.0384	0.0364	0.0303	**0.0273**
SpectEW	0.1477	0.1523	0.1570	0.1500	0.1500	**0.1465**	0.1593	0.1523	0.1511	0.1535	0.1500	**0.1442**
PenglungEW	0.0830	0.0586	0.0788	0.0910	0.0788	**0.0606**	0.0870	0.0930	0.0748	0.0768	0.0648	**0.0566**
Prostate Tumor	0.0623	0.0440	0.0591	0.0683	0.0591	**0.0415**	0.0653	0.0698	0.0561	0.0576	0.0486	**0.0425**
11 Tumors	0.0463	0.0451	0.0451	0.0440	0.0451	**0.0429**	0.0451	0.0451	0.0463	0.0451	0.0463	**0.0384**
Lung Cancer	0.0406	0.0417	0.0327	0.0338	**0.0316**	0.0406	0.0451	0.0440	0.0463	0.0429	0.0417	**0.0316**
Overall Mean	0.0766	0.0751	0.0786	0.0743	0.0671	**0.0793**	0.0807	0.0758	0.0773	0.0758	0.0723	**0.0627**

transformed datasets. It can be visualized from the tables that the proposed bSSA with PCA-fastICA outperform others. The overall mean accuracy and number of features selected, returned by bSSA on PCA-fastICA transformed dataset is 95.26% and 7.78 respectively. If the average accuracy and average number of features selected in all tables (refer Table 4, 5, 6, 7, and 8) are compared, it can be seen that bSSA shows very excellent results in the transformed data sets, especially in the PCA-fastICA transformed datasets. Moreover, bSSA also perform better than other if no feature representation method such as PCA, fastICA, PCA-fastICA is used. Thus, from the above analysis, it can be pertinent that the feature representation generally enhances the performance.

Furthermore, the mean computational time of all the considered methods is also compared to assess the computational efficiency of the proposed bSSA. The computational time of all the considered methods are depicted in Table 9. Reported results indicate that bSSA is computationally efficient than other considered methods. Further, the error rates are also computed and shown in Table 10. Table 10 shows that the proposed bSSA returns the least error values for five

TABLE 11. Results of the Wilcoxon test for statistically significance level at $\alpha = 0.05$.

Datasets	bSSA-bGA		bSSA-CSO		bSSA-bGWO		bSSA-bCSA		bSSA-bBCS	
	$p - value$	SIG	$p - value$	SIG	$p - value$	SIG	$p - value$	SIG	$p - value$	SIG
Breastcancer	3.60E-10	+	3.21E-10	+	8.26E-11	+	1.97E-09	+	4.02E-08	+
BreastEW	1.53E-08	+	3.94E-09	+	2.32E-09	+	1.11E-10	+	3.54E-09	+
CongressEW	3.34E-08	+	1.31E-08	+	1.39E-10	+	1.22E-09	+	4.47E-02	+
HeartEW	3.63E-09	+	6.63E-09	+	2.56E-09	+	2.86E-02	-	1.52E-11	+
IonosphereEW	3.59E-13	+	2.76E-07	+	3.90E-09	+	2.68E-13	+	2.39E-06	+
SonarEW	4.54E-07	+	2.80E-08	+	1.32E+00	+	1.63E+00	+	5.45E-06	+
WineEW	1.12E-10	+	7.98E-12	+	5.54E-07	+	3.17E-06	+	4.42E-08	+
Vote	8.81E-14	+	1.07E-14	+	4.85E-11	+	1.97E-10	+	8.50E-13	+
Tic-tac-toe	1.88E-07	+	2.37E-05	+	0.00E+00	+	3.61E-08	+	1.92E-09	+
Zoo	7.26E-10	+	3.35E-10	+	2.77E-08	+	4.44E-07	+	1.39E-09	+
SpectEW	9.49E-11	+	9.49E-08	+	3.45E-07	+	3.85E-08	+	5.75E-09	+
PenglungEW	4.66E-09	+	5.33E-07	+	1.27E-07	+	3.80E-09	+	2.68E-07	+
Prostate Tumor	6.77E-07	+	1.02E-06	=	4.07E-06	=	2.86E-10	+	4.09E-10	+
11 Tumors	7.75E-11	+	1.33E-10	+	6.45E-01	+	6.83E-06	+	2.40E-09	+
Lung Cancer	1.67E-06	+	3.34E-10	+	1.43E-09	+	4.92E-06	+	2.08E-08	+

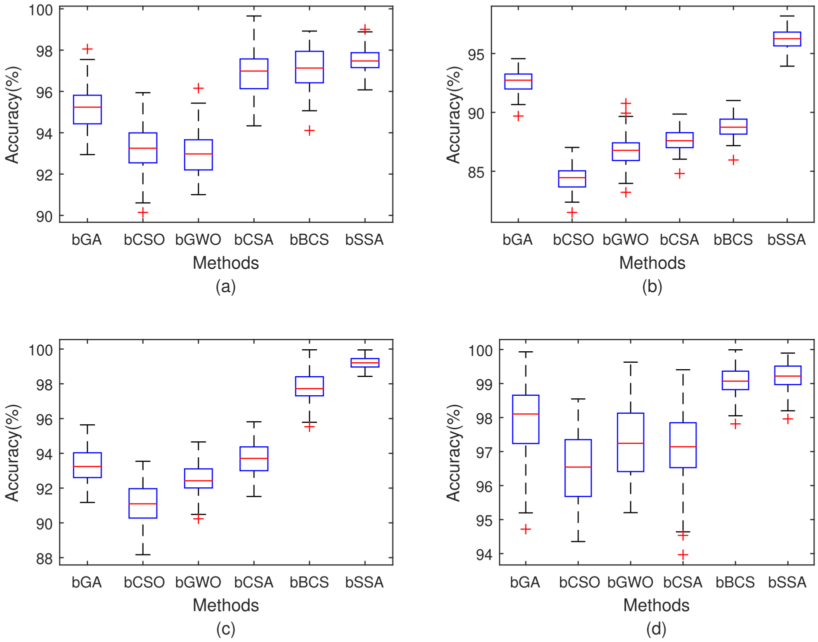


FIGURE 2. Box plot graphs for the accuracy of original dataset (a) BreastEW, (b) IonosphereEW, (c) WineEW, and (d) Zoo dataset.

original and nine transformed datasets. bBCS, bCSO, and bGA show better efficiency over BreastEW, WineEW, and Vote datasets.

Moreover, wilcoxon rank sum is also conducted in Table 11 to statistically validate the efficacy of the proposed bSSA with feature transformation. Table 11 shows the corresponding $p - value$ along with SIG (significance). Significance level is decided by comparing $p - value$ and level of significance α. The null hypothesis is accepted if $p - value > 0.05$ and symbolized by '$=$' else, it is denied and represented by '$+$' or '$-$' symbol based on $p - value$. In table, '$+$' sign means that the proposed algorithm is significantly different and better, '$=$' means there is no significant (major) difference and '$-$' indicates that the compared algorithm is better than the proposed algorithm. It can be observed from the table that there are '$+$' for maximum number of datasets which means the proposed bSSA with feature transformation is significantly different and better than other considered algorithms. There are a few datasets such as HeartEW for which bCSA shows better performance than the proposed

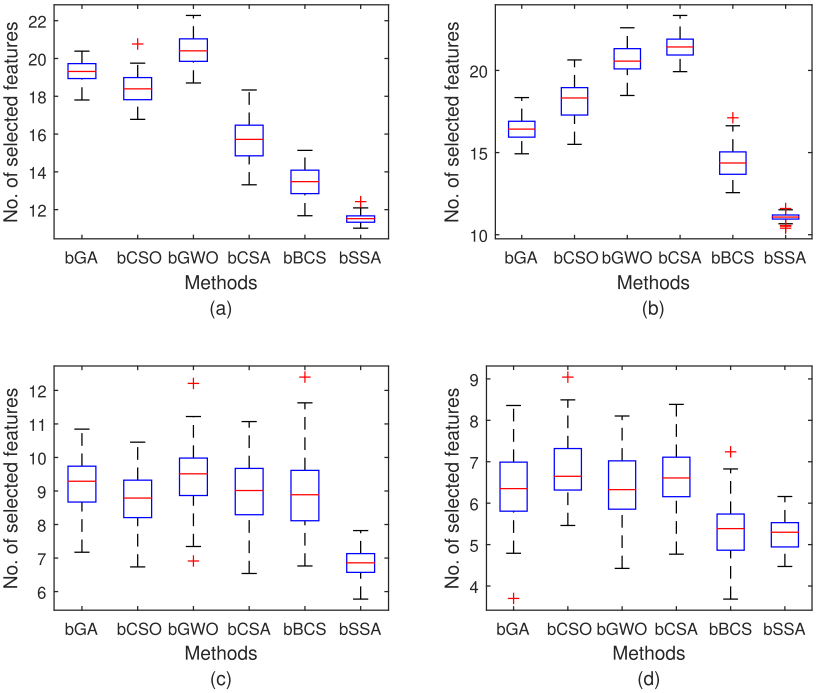


FIGURE 3. Box plot graphs for the number of selected features of original dataset (a) BreastEW, (b) IonosphereEW, (c) WineEW, and (d) Zoo dataset.

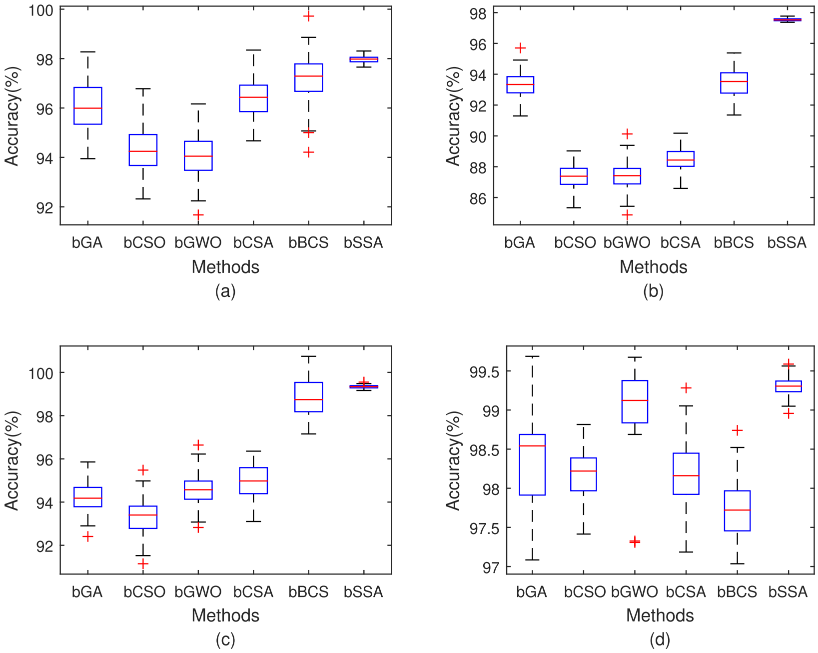
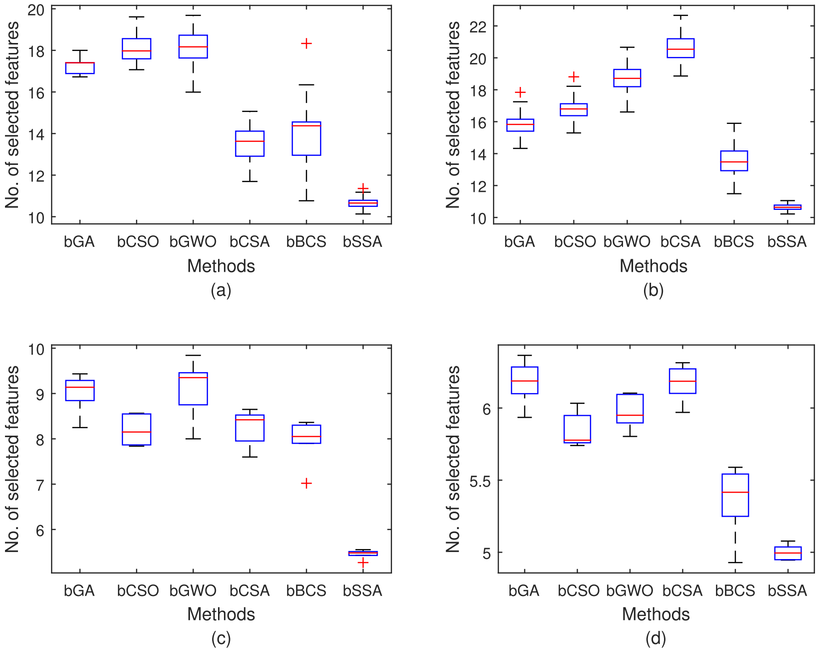


FIGURE 4. Box plot graph to measure the consistency of accuracy for PCA-fastICA-based transformed datasets (a) BreastEW, (b) IonosphereEW, (c) WineEW, and (d) Zoo dataset.

algorithm and Prostate Tumor for which bGWO and bCSA show similar performance.

Moreover, box plot graphs have also been drawn to appraise the consistency and stability of all the considered methods. The box plots show the distribution of data graphically and are generally used to analyze any method's consistency. One may quickly discover the symmetry, median, and skewness of box plots. To analyze the consistency, box plot graphs for original and PCA-fastICA based transformed datasets are plotted in Figs. 2 - 5. In the box plot, x-axis represents the method's name and the y-axis depicts the respective performance criterion. It is perceived



(a) (b)

(c) (d)

FIGURE 5. Box plot graphs to measure the consistency in the selected features of PCA-fastICA-based transformed datasets.

from the box plots that bSSA shows a minimal variation in the transformed data set based on PCA-fastICA. If the box plot graphs of original datasets, datasets with optimal features, and transformed datasets are examined, then the proposed bSSA feature selection method's consistency on PCA-fastICA-based transformed datasets can be easily observed. Hence, from the experimental analysis, the superiority of the proposed bSSA is observed.

VI. CONCLUSION

This paper introduced a new binary salp swarm algorithm (bSSA) to select the relevant features of a data set. To select the optimal feature set, the proposed bSSA first transforms the original data using PCA-fastICA based transformed datasets followed by feature selection steps. As the feature selection methods only reduce the unnecessary and unimportant features, not the correlated and higher-order dependencies among features. Therefore, to eradicate redundancy, correlation, and dependency among features simultaneously, the proposed bSSA first transforms the datasets using PCA-fastICA and then selects the optimal features. The proposed bSSA attains the mean accuracy of 93.97%, 94.56%, and 94.73% over original, PCA, ICA-based transformed datasets respectively. The proposed feature selection method achieves 95.26% mean accuracy using 7.78 features on PCA-fastICA transformed datasets. From the statistical and empirical analysis, it is found that the proposed bSSA feature selection method outperforms the existing methods for the majority of the performance measures. The proposed IWOA employs random choices to find the optimal solution, which means that the computing time and the solution quality are actually random variables. Due to this stochastic nature, its rigorous analysis would be very difficult. Moreover, the efficiency of the proposed bSSA with feature transformation may reduce on high dimensional imbalanced dataset since training, validation, and testing set could contain a different number of instances.

Future work may enhance the accuracy using a multi-label feature selection approach. New and more efficient variant of SSA and other meta-heuristics may perform better with fine tuning of control parameters. Moreover, multi-objective fitness function could also be used for improving the efficacy.

REFERENCES

[1] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017.

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

[3] M. Saraswat and K. V. Arya, "Feature selection and classification of leukocytes using random forest," *Med. Biol. Eng. Comput.*, vol. 52, no. 12, pp. 1041–1052, Dec. 2014.

[4] J. Wei, R. Zhang, Z. Yu, R. Hu, J. Tang, C. Gui, and Y. Yuan, "A BPSO-SVM algorithm based on memory renewal and enhanced mutation mechanisms for feature selection," *Appl. Soft Comput.*, vol. 58, pp. 176–192, Sep. 2017.

[5] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, Dec. 1997.

[6] H. Li, C.-J. Li, X.-J. Wu, and J. Sun, "Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine," *Appl. Soft Comput.*, vol. 19, pp. 57–67, Jun. 2014.

[7] M. E. ElAlami, "A filter model for feature subset selection based on genetic algorithm," *Knowl.-Based Syst.*, vol. 22, no. 5, pp. 356–362, Jul. 2009.

[8] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowe, "A survey on filter techniques for feature selection in gene expression microarray analysis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 4, pp. 1106–1119, Jul. 2012.

[9] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.

[10] C. Freeman, D. Kulić, and O. Basir, "An evaluation of classifier-specific filter measure performance for feature selection," *Pattern Recognit.*, vol. 48, no. 5, pp. 1812–1826, May 2015.

[11] B. Tran, B. Xue, and M. Zhang, "Genetic programming for feature construction and selection in classification on high-dimensional data," *Memetic Comput.*, vol. 8, pp. 3–15, Dec. 2016.

[12] I. T. Jolliffe, "Principal component analysis and factor analysis," in *Principal Component Analysis*. New York, NY, USA: Springer, 1986, pp. 115–128.

[13] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowl.-Based Syst.*, vol. 24, no. 7, pp. 1024–1032, Oct. 2011.

[14] L.-C. Lee, C.-Y. Liong, K. Osman, and A. A. Jemain, "Comparison of several variants of principal component analysis (PCA) on forensic analysis of paper based on IR spectrum," in *Proc. AIP Conf.*, vol. 1750, 2016, Art. no. 060012.

[15] V. Rokhlin, A. Szlam, and M. Tygert, "A randomized algorithm for principal component analysis," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1100–1124, Jan. 2010.

[16] D. Lopez-Paz, S. Sra, A. Smola, Z. Ghahramani, and B. Schölkopf, "Randomized nonlinear component analysis," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1359–1367.

[17] A. K. Chattopadhyay, S. Mondal, and A. Biswas, "Independent component analysis and clustering for pollution data," *Environ. Ecol. Statist.*, vol. 22, no. 1, pp. 33–43, Mar. 2015.

[18] C. Bugli and P. Lambert, "Comparison between principal component analysis and independent component analysis in electroencephalograms modelling," *Biometrical J.*, vol. 49, no. 2, pp. 312–327, Apr. 2007.

[19] F. Yao, J. Coquery, and K.-A. Lê Cao, "Independent principal component analysis for biologically meaningful dimension reduction of large biological data sets," *BMC Bioinf.*, vol. 13, no. 1, p. 24, 2012.

[20] A. ChandraPandey, D. SinghRajpoot, and M. Saraswat, "Data clustering based on data transformation and hybrid step size-based cuckoo search," in *Proc. 11th Int. Conf. Contemp. Comput. (IC3)*, Aug. 2018, pp. 1–6.

[21] Y. Kaya and H. Pehlivan, "Comparison of classification algorithms in classification of ECG beats by time series," in *Proc. 23nd Signal Process. Commun. Appl. Conf. (SIU)*, May 2015, pp. 407–410.

[22] Y. Kaya, H. Pehlivan, and M. E. Tenekeci, "Effective ECG beat classification using higher order statistic features and genetic feature selection," *Biomed. Res.*, vol. 28, no. 17, pp. 7594–7603, 2017.

[23] E. Oja and Z. Yuan, "The FastICA algorithm revisited: Convergence analysis," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1370–1381, Nov. 2006.

[24] R. Leardi, R. Boggia, and M. Terrile, "Genetic algorithms as a strategy for feature selection," *J. Chemometrics*, vol. 6, no. 5, pp. 267–281, Sep. 1992.

[25] S. Gu, R. Cheng, and Y. Jin, "Feature selection for high-dimensional classification using a competitive swarm optimizer," *Soft Comput.*, vol. 22, no. 3, pp. 811–822, Feb. 2018.

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

[27] A. C. Pandey and D. S. Rajpoot, "Feature selection method based on grey wolf optimization and simulated annealing," *Recent Patents Comput. Sci.*, vol. 12, pp. 1–18, Apr. 2019.

[28] G. I. Sayed, A. E. Hassanien, and A. T. Azar, "Feature selection via a novel chaotic crow search algorithm," *Neural Comput. Appl.*, vol. 31, no. 1, pp. 171–188, Jan. 2019.

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

[30] Z. Zhao and H. Liu, "Spectral feature selection for supervised and unsupervised learning," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, 2007, pp. 1151–1157.

[31] J. M. Peña and R. Nilsson, "On the complexity of discrete feature selection for optimal classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1517–1522, Aug. 2010.

[32] L. Du and Y.-D. Shen, "Unsupervised feature selection with adaptive structure learning," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 209–218.

[33] Y. Han, Y. Yang, Y. Yan, Z. Ma, N. Sebe, and X. Zhou, "Semisupervised feature selection via spline regression for video semantic recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 2, pp. 252–264, Feb. 2015.

[34] Z. Zhao and H. Liu, "Semi-supervised feature selection via spectral analysis," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2007, pp. 641–646.

[35] Y. Han, Y. Yang, and X. Zhou, "Co-regularized ensemble for feature selection," in *Proc. IJCAI*, vol. 13, 2013, pp. 1380–1386.

[36] R. He, T. Tan, L. Wang, and W.-S. Zheng, "L_2, 1 Regularized correntropy for robust feature selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2504–2511.

[37] A. Kulhari, A. Pandey, R. Pal, and H. Mittal, "Unsupervised data classification using modified cuckoo search method," in *Proc. 9th Int. Conf. Contemp. Comput. (IC3)*, Aug. 2016, pp. 1–5.

[38] L. Wolf and A. Shashua, "Feature selection for unsupervised and supervised inference: The emergence of sparsity in a weight-based approach," *J. Mach. Learn. Res.*, vol. 6, pp. 1855–1887, Jan. 2005.

[39] A. Chandra Pandey, D. Singh Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method," *Inf. Process. Manage.*, vol. 53, no. 4, pp. 764–779, Jul. 2017.

[40] A. C. Pandey and A. Kulhari, "Semi-supervised spatiotemporal classification and trend analysis of satellite images," in *Advances in Computer and Computational Sciences*. Singapore: Springer, 2018, pp. 353–363.

[41] X. Chang, F. Nie, Y. Yang, and H. Huang, "A convex formulation for semi-supervised multi-label feature selection," in *Proc. AAAI*, 2014, pp. 1171–1177.

[42] S. F. da Silva, M. X. Ribeiro, J. D. E. S. Batista Neto, C. Traina-Jr., and A. J. M. Traina, "Improving the ranking quality of medical image retrieval using a genetic feature selection method," *Decis. Support Syst.*, vol. 51, no. 4, pp. 810–820, Nov. 2011.

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

[44] Y. Liu, F. Tang, and Z. Zeng, "Feature selection based on dependency margin," *IEEE Trans. Cybern.*, vol. 45, no. 6, pp. 1209–1221, Jun. 2015.

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

[46] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Hybrid step size based cuckoo search," in *Proc. 10th Int. Conf. Contemp. Comput. (IC3)*, Aug. 2017, pp. 1–6.

[47] S. Kohli, M. Kaushik, K. Chugh, and A. C. Pandey, "Levy inspired enhanced grey wolf optimizer," in *Proc. 5th Int. Conf. Image Inf. Process. (ICIIP)*, Nov. 2019, pp. 338–342.

[48] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Netw.*, vol. 4, pp, pp. 1942–1948, 1995.

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

[50] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Data clustering using hybrid improved cuckoo search method," in *Proc. 9th Int. Conf. Contemp. Comput. (IC3)*, Aug. 2016, pp. 1–6.

[51] R. N. Khushaba, A. Al-Ani, A. AlSukker, and A. Al-Jumaily, "A combined ant colony and differential evolution feature selection algorithm," in *Proc. Int. Conf. Ant Colony Optim. Swarm Intell.* Berlin, Germany: Springer, 2008, pp. 1–12.

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

[53] L.-H. Chen and H.-D. Hsiao, "Feature selection to diagnose a business crisis by using a real GA-based support vector machine: An empirical study," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 1145–1155, Oct. 2008.

[54] J. Derrac, S. García, and F. Herrera, "A first study on the use of coevolutionary algorithms for instance and feature selection," in *Hybrid Artificial Intelligence Systems*. Berlin, Germany: Springer, 2009, pp. 557–564.

[55] D. P. Muni, N. R. Pal, and J. Das, "Genetic programming for simultaneous feature selection and classifier design," *IEEE Trans. Syst., Man Cybern., B (Cybern.)*, vol. 36, no. 1, pp. 106–117, Feb. 2006.

[56] K. Neshatian and M. Zhang, "Dimensionality reduction in face detection: A genetic programming approach," in *Proc. 24th Int. Conf. Image Vis. Comput. New Zealand*, Nov. 2009, pp. 391–396.

[57] M. C. Lane, B. Xue, I. Liu, and M. Zhang, "Particle swarm optimisation and statistical clustering for feature selection," in *Proc. Australas. Joint Conf. Artif. Intell.* Springer, 2013, pp. 214–220.

[58] B. Tran, B. Xue, and M. Zhang, "Improved PSO for feature selection on high-dimensional datasets," in *Proc. Asia–Pacific Conf. Simulated Evol. Learn.* Cham, Switzerland: Springer, 2014, pp. 503–515.

[59] L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, "Improved binary PSO for feature selection using gene expression data," *Comput. Biol. Chem.*, vol. 32, no. 1, pp. 29–38, Feb. 2008.

[60] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," *Appl. Soft Comput.*, vol. 18, pp. 261–276, May 2014.

[61] L. Ke, Z. Feng, and Z. Ren, "An efficient ant colony optimization approach to attribute reduction in rough set theory," *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1351–1357, Jul. 2008.

[62] N. M. O'Boyle, D. S. Palmer, F. Nigsch, and J. B. Mitchell, "Simultaneous feature selection and parameter optimisation using an artificial ant colony: Case study of melting point prediction," *Chem. Central J.*, vol. 2, no. 1, p. 21, Dec. 2008.

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

[64] L. Ke, Z. Feng, Z. Xu, K. Shang, and Y. Wang, "A multiobjective ACO algorithm for rough feature selection," in *Proc. 2nd Pacific–Asia Conf. Circuits, Commun. Syst.*, vol. 1, Aug. 2010, pp. 207–210.

[65] C.-S. Yang, L.-Y. Chuang, Y.-J. Chen, and C.-H. Yang, "Feature selection using memetic algorithms," in *Proc. 3rd Int. Conf. Converg. Hybrid Inf. Technol.*, vol. 1, Nov. 2008, pp. 416–423.

[66] M. Kang, M. R. Islam, J. Kim, J.-M. Kim, and M. Pecht, "A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3299–3310, May 2016.

[67] E. Pashaei and N. Aydin, "Binary black hole algorithm for feature selection and classification on biological data," *Appl. Soft Comput.*, vol. 56, pp. 94–106, Jul. 2017.

[68] Y. Kaya, "Feature selection using binary cuckoo search algorithm," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, May 2018, pp. 1–4.

[69] G. I. Sayed, A. Tharwat, and A. E. Hassanien, "Chaotic dragonfly algorithm: An improved Metaheuristic algorithm for feature selection," *Int. J. Speech Technol.*, vol. 49, no. 1, pp. 188–205, Jan. 2019.

[70] R. Guha, M. Ghosh, S. Kapri, S. Shaw, S. Mutsuddi, V. Bhateja, and R. Sarkar, "Deluge based genetic algorithm for feature selection," *Evol. Intell.*, Mar. 2019, doi: 10.1007/s12065-019-00218-5.

[71] M. Taradeh, M. Mafarja, A. A. Heidari, H. Faris, I. Aljarah, S. Mirjalili, and H. Fujita, "An evolutionary gravitational search-based feature selection," *Inf. Sci.*, vol. 497, pp. 219–239, Sep. 2019.

[72] R. K. Agrawal, B. Kaur, and S. Sharma, "Quantum based whale optimization algorithm for wrapper feature selection," *Appl. Soft Comput.*, vol. 89, Apr. 2020, Art. no. 106092.

[73] S. Ouadfel and M. Abd Elaziz, "Enhanced crow search algorithm for feature selection," *Expert Syst. Appl.*, vol. 159, Nov. 2020, Art. no. 113572.

[74] H. Jia, J. Li, W. Song, X. Peng, C. Lang, and Y. Li, "Spotted hyena optimization algorithm with simulated annealing for feature selection," *IEEE Access*, vol. 7, pp. 71943–71962, 2019.

[75] K. K. Ghosh, S. Ahmed, P. K. Singh, Z. W. Geem, and R. Sarkar, "Improved binary sailfish optimizer based on adaptive β-hill climbing for feature selection," *IEEE Access*, vol. 8, pp. 83548–83560, 2020.

[76] M. Alweshah, S. Alkhalaileh, D. Albashish, M. Mafarja, Q. Bsoul, and O. Dorgham, "A hybrid mine blast algorithm for feature selection problems," *Soft Comput.*, Jul. 2020, doi: 10.1007/s00500-020-05164-4.

[77] M. Alweshah, S. A. Khalaileh, B. B. Gupta, A. Almomani, A. I. Hammouri, and M. A. Al-Betar, "The monarch butterfly optimization algorithm for solving feature selection problems," *Neural Comput. Appl.*, Jul. 2020, doi: 10.1007/s00521-020-05210-0.

[78] J. Li, S. Fong, L.-S. Liu, N. Dey, A. S. Ashour, and L. Moraru, "Dual feature selection and rebalancing strategy using Metaheuristic optimization algorithms in X-ray image datasets," *Multimedia Tools Appl.*, vol. 78, no. 15, pp. 20913–20933, Aug. 2019.

[79] Y. Zhang, D.-W. Gong, and J. Cheng, "Multi-objective particle swarm optimization approach for cost-based feature selection in classification," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 1, pp. 64–75, Jan. 2017.

[80] Y. Zhang, S. Cheng, Y. Shi, D.-W. Gong, and X. Zhao, "Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm," *Expert Syst. Appl.*, vol. 137, pp. 46–58, Dec. 2019.

[81] Y. Hu, Y. Zhang, and D. Gong, "Multiobjective particle swarm optimization for feature selection with fuzzy cost," *IEEE Trans. Cybern.*, early access, Sep. 14, 2020, doi: 10.1109/TCYB.2020.3015756.

[82] E. Elhariri, N. El-Bendary, and S. A. Taie, "Using hybrid filter-wrapper feature selection with multi-objective improved-salp optimization for crack severity recognition," *IEEE Access*, vol. 8, pp. 84290–84315, 2020.

[83] H. Chen, F. Liu, P. Chang, S. Yao, F. Huang, and J. Hu, "An improved salp swarm algorithm based on spark for feature selection," in *Proc. 15th Int. Conf. Comput. Sci. Edu. (ICCSE)*, Aug. 2020, pp. 424–429.

[84] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, vol. 46. Hoboken, NJ, USA: Wiley, 2004.

[85] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, "Non-Gaussianity from inflation: Theory and observations," *Phys. Rep.*, vol. 402, nos. 3–4, pp. 103–266, Nov. 2004.

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

[87] A. C. Pandey, A. K. Tripathi, R. Pal, H. Mittal, and M. Saraswat, "Spiral salp swarm optimization algorithm," in *Proc. 4th Int. Conf. Inf. Syst. Comput. Netw. (ISCON)*, Nov. 2019, pp. 722–727.

[88] D. Dheeru and E. Karra Taniskidou. *UCI Machine Learning Repository: Data Sets*. Accessed: May 10, 2020. [Online]. Available: http://archive.ics.uci.edu/ml/datasets.php

SAYAR SINGH SHEKHAWAT received the B.E. degree from the University of Rajasthan, Jaipur, and the M.Tech. degree from Rajasthan Technical University, Kota, India. He is currently a Research Scholar with Rajasthan Technical University, Kota. He has published two research articles in well-known SCI journals and presented his research in various conferences, and attended several national and international conferences and workshops. His research interests include algorithm design, nature inspired algorithms, and computational intelligence.

HARISH SHARMA is currently an Associate Professor with the Department of Computer Science and Engineering, Rajasthan Technical University, Kota. He has published more than 45 papers in various international journals and conferences. His research interest includes nature-inspired optimization techniques. He is the Secretary and a Founder Member of the Soft Computing Research Society of India. He is an Associate Editor of the *International Journal of Swarm Intelligence (IJSI)*, published by Inderscience. He has also edited special issues of the journals *Memetic Computing* and the *Journal of Experimental and Theoretical Artificial Intelligence*.

SANDEEP KUMAR received the B.E., M.Tech., and Ph.D. degrees in computer science and engineering. He is currently an Associate Professor of computer science and engineering with CHRIST (Deemed to be University), Bengaluru, India. He edited special issues for many journals, including *JDMSC*, *JIM*, *JSMS*, *JIOS*, *IJGUC*, *IJIIDS*, *RPCS*, and *RACSC*. He has published more than 50 papers in well-known SCI/SCOPUS indexed international journals and conferences, and attended several national and international conferences and workshops. He has authored/edited four books in the area of computer science. His research interests include nature inspired algorithms, swarm intelligence, soft computing, and computational intelligence.

ANAND NAYYAR (Senior Member, IEEE) received the Ph.D. degree in computer science (wireless sensor networks and swarm intelligence) from Desh Bhagat University, in 2017. He is currently working with the Graduate School, Duy Tan University, Da Nang, Vietnam. He is a Certified Professional with 75 Professional certificates from CISCO, Microsoft, Oracle, Google, Beingcert, EXIN, GAQM, Cyberoam, and many more. He has published more than 350 research papers in various national and international conferences, international journals (Scopus/SCI/SCIE/SSCI Indexed) with high impact factor. He has authored/coauthored cum Edited 25 Books of Computer Science. He has associated with more than 500 International Conferences as a Programme Committee/Chair/Advisory Board/Review Board Member. He has five Australian Patents to his credit in the area of wireless communications and artificial intelligence. He is also working in the area of wireless sensor networks, MANETS, swarm intelligence, cloud computing, the Internet of Things, blockchain, machine learning, deep learning, cyber security, network simulation, and wireless communications. He is a member of more than 50 Associations as a Senior and a Life Member and also acting as ACM Distinguished Speaker. He was awarded 27 Awards for Teaching and Research. Young Scientist, Best Scientist, Young Researcher Award, Outstanding Researcher Award, Excellence in Teaching and many more. He is acting as the Editor-in-Chief of IGI-Global, USA Journal titled *International Journal of Smart Vehicles and Smart Transportation (IJSVST)*.

BASIT QURESHI (Member, IEEE) received the bachelor's degree in computer science from Ohio University, Athens, OH, USA, in 2000, the master's degree in computer science from Florida Atlantic University, in 2002, and the Ph.D. degree in computer science from the University of Bradford, U.K., in 2011. He is currently an Associate Professor and the Chair of the Department of Computer Science, Prince Sultan University, Saudi Arabia. He has published over 60 research works in the area of robotics, the Internet of Things, machine learning, and computational intelligence. He is a member of the IEEE Computer Society, the IEEE Communications Society, and ACM.

• • •