Tech Science Press

Feature Selection Using Grey Wolf Optimization with Random Differential Grouping

R. S. Latha[1,*], B. Saravana Balaji[2], Nebojsa Bacanin[3], Ivana Strumberger[3], Miodrag Zivkovic[3] and Milos Kabiljo[3]

[1]Department of Computer Science and Engineering, Kongu Engineering College, Erode, 638060, Tamilnadu, India
[2]Department of Information Technology, Lebanese French University, Erbil, Iraq
[3]Singidunum University, Belgrade, 160622, Serbia
*Corresponding Author: R. S. Latha. Email: r.lslatha@yahoo.com

Abstract: Big data are regarded as a tremendous technology for processing a huge variety of data in a short time and with a large storage capacity. The user's access over the internet creates massive data processing over the internet. Big data require an intelligent feature selection model by addressing huge varieties of data. Traditional feature selection techniques are only applicable to simple data mining. Intelligent techniques are needed in big data processing and machine learning for an efficient classification. Major feature selection algorithms read the input features as they are. Then, the features are preprocessed and classified. Here, an algorithm does not consider the relatedness. During feature selection, all features are misread as outputs. Accordingly, a less optimal solution is achieved. In our proposed research, we focus on the feature selection by using supervised learning techniques called grey wolf optimization (GWO) with decomposed random differential grouping (DrnDG-GWO). First, decomposition of features into subsets based on relatedness in variables is performed. Random differential grouping is performed using a fitness value of two variables. Now, every subset is regarded as a population in GWO techniques. The combination of supervised machine learning with swarm intelligence techniques produces best feature optimization results in this research. Once the features are optimized, we classify using advanced kNN process for accurate data classification. The result of DrnDG-GWO is compared with those of the standard GWO and GWO with PSO for feature selection to compare the efficiency of the proposed algorithm. The accuracy and time complexity of the proposed algorithm are 98% and 5 s, which are better than the existing techniques.

1 Introduction

Big data processing in this internet world with advanced technologies, such as 4G, 5G, and 6G applications, is a challenging task. In the bulk of huge data repository, selecting the appropriate feature is still not optimal. Artificial intelligence, machine learning, and data mining algorithms have been used for optimal feature selection. A good data classifier must predict the data within a short time with high precision during feature selection. Effective preprocessing and crucial feature selection algorithms are required to achieve a fine quality of data dimensions during classification. If these two processes are highly optimal, then the classification process will be great. Thus, these aspects became the main motive for us to concentrate on effective feature selection algorithms.

Feature selection algorithms are necessary to achieve an optimal solution using many machine learning techniques. Machine learning algorithms still need some additional intelligent process for selecting highly relevant features in the database server. Demands and necessity makes us think beyond machine learning and bring a new meta-heuristic approach in feature selection. At present much optimization problems looks for swarm intelligence techniques for its optimal solution in nature. A group of animal behaviors are considered to develop a metaheuristic model, and an algorithm is developed. In this article, we use grey wolf optimization (GWO) for selecting optimal features after performing decomposition random differential grouping on the feature dataset.

The selection of appropriate feature is an interdisciplinary field of data mining. Big data cannot be easily handled with existing feature selection, classification, and optimization techniques. This aspect is necessary because big data are regarded as a collection of high dimensional data from various application sources. These new techniques help in computing the huge data with high force. Data are significant knowledge base for every business process. A core data transaction occurs every minute in the online business. This is a huge data repository were we have to implement mining techniques. In this research, we focus on selecting the appropriate features in the dataset without missing any data features and optimizing the feature for classification. The main contribution we introduce in this work is decomposition of data with random differential grouping. Decomposition divides the data and its subsets by fitness value calculation. In this work, all features are considered for decomposition. For effectiveness of feature selection, The features are divided in to subsets for an effective feature selection, and the relatedness is checked using random differential grouping. The group of population is regarded as an input for the GWO algorithm.

The rest of the papers have the following sections: Section 2 provides the literature review of this approach. Section 3 presents the proposed method description and working principal. Section 4 evaluates the results and comparison. Finally, Section 5 presents the conclusion with future scope.

2 Related Work

Sun et al. [1] defined the high-dimensional data process by using the learning-based local search algorithm. The main issue is nonlinear complicated problems are converted into linear local groups by using a local component learning strategy. The relevant features are mined along with marginal framework. This proposed new technique uses machine learning with mathematical analysis. Fong et al. [2] implemented the accurate selection of features from the huge high-dimensional data by using an intelligent classifier. Swarm search algorithm is a novel method used for feature selection and produces optimal solutions. This mechanism is a metaheuristic approach for selecting relevant data from the dataset. The aforementioned approach is an advantageous method with flexibility using a fitness value of the data during classification. Fong et al. [3] suggested a new coefficient-based feature selection model. This method has higher accuracy compared with the other clustering techniques. Clustering with a coefficient variation has a balanced process between the generalization and the problem of over fitting to

produce an optimal solution. Finally, the fastest discrimination approach is called hyper piping, which is used for checking data groups with high classification accuracy.

Peralta et al. [4] suggested the evolutionary algorithm for feature selection. Such a process uses a map reduce strategy for generating the subsets of features in a huge database. This work classifies the original dataset into various blocks for a mapping-based computation process. Classified blocks are integrated by using weights of features in reducing a phase as a finalized vector. Data assessment is performed by using SVM, logistic regression, and naïve Bayes classifier in the big data framework. The big data uses ant colony optimization (ACO) with swarm search (SS) for selecting the appropriate features with stream mining. Harde et al. [5] suggested a light weight-based novel ACO to achieve high accuracy in streaming infrastructure. This infrastructure will validate high-dimensional data during feature extraction. Selvi and Valarmathi et al. [6] suggested the firefly algorithm for feature selection for optimal local searching purposes. Firefly search is effective for local searching process. However, this work cannot provide global optimal solution in the search results. The selected features classified using various algorithms, such as kNN and multilayer perceptron utilizing a neural network classification. The outcome of this work is effective with firefly algorithm in the local optimal search process. Viegas et al. [7] demonstrated the selection of high-dimensional features by using genetic algorithm. This algorithm helps in targeting the different feature sets with unique metrics. According to Viegas, the problem of diversity in feature selection projections by hypothesis deviation are effectively handled by genetic algorithms. The classification result is more efficient than conventional classifiers.

Anusuya et al. [8] applied the map reduce-based graph mining process by using Apriori Particle Swarm Optimization (APSO). The behavior of the incoming streamed data with high dimensions is addressed in this article. The dynamic composition of the large data scale was computed using the PSO algorithm. Another category of big data classification uses different classification strategies by using machine learning, swarm intelligence, etc. Jun et al. [9] used the extreme learning method for huge data classification by fuzzy-based positive and negative categories. The central hidden layer is used to pull out the classified data from the hidden layer. The least square support vector machine and proximal-based support vector machine learning used by Huang et al. in [10,11] effectively classify the data within a short time. Here, extreme learning approaches are used to classify the disjoint vectors. Van der Char [12] proposed distributed computing-based online data classification. The learners in the distributed system used to learn the collected data sources from the distributed environment. After collecting the data which is heterogeneous as well as distributed nature is considered as distributed data context with joint classification. Grolinger et al. [13] suggested the map reduce-based data classification. The big data-based map reduce classification model is a difficult task. This process implements parallel computing and efficiently classifies the data. Rebentrost et al. [14] proposed the effective big data classification using quantum SVM. In this approach, the SVM is used in a quantum processing computer for analyzing vector logarithmic complexity with training data samples.

The various search methodology, such as greedy algorithm and evolutionary techniques [15], are used in effective feature selection. One of the effective feature selection algorithms is evolutionary algorithm [16], which is widely used in big data processing. Big data computed with cloud computing server decomposes the single complex problem into multiple subproblems. Each subproblem is individually addressed and finally combined to produce effective results [17]. The optimal decomposer and optimizer in the cloud computing infrastructure provide high performance in feature selection. Decomposition in different datasets are performed in papers [18–24] at cloud storage for an optimal solution.

3 Proposed DrnDG-GWO Methodology

Numerous optimal models are available for feature selection in the big data area. The complex nature of processing such big data is still a big challenge. Data mining techniques have proven to be the best in processing high dimensionality datasets. The preprocessing steps play a vital role in reducing the dataset for further processing the big data. Among the preprocessing phases, feature selection is an important step in selecting the optimal and relevant features for further classification or clustering. In this work, a new variant of random differential grouping with GWO is proposed as a hybrid feature selection approach to find the optimal solution. A modified kNN classifier is used for feature classification. The dataset for training and testing will be divided on the basis of the k-fold validation. An overview of the proposed architecture is shown in Fig. 1. The proposed feature selection is applied on different big datasets. The initial dataset features are decomposed into subcomponents based on the needed features for further classification. Fitness component is used for differential grouping the similar component in the dataset. Finally, the selected subcomponents are optimized by using the GWO, which is used to find the optimal solution based on the first, second, and third possible solutions.
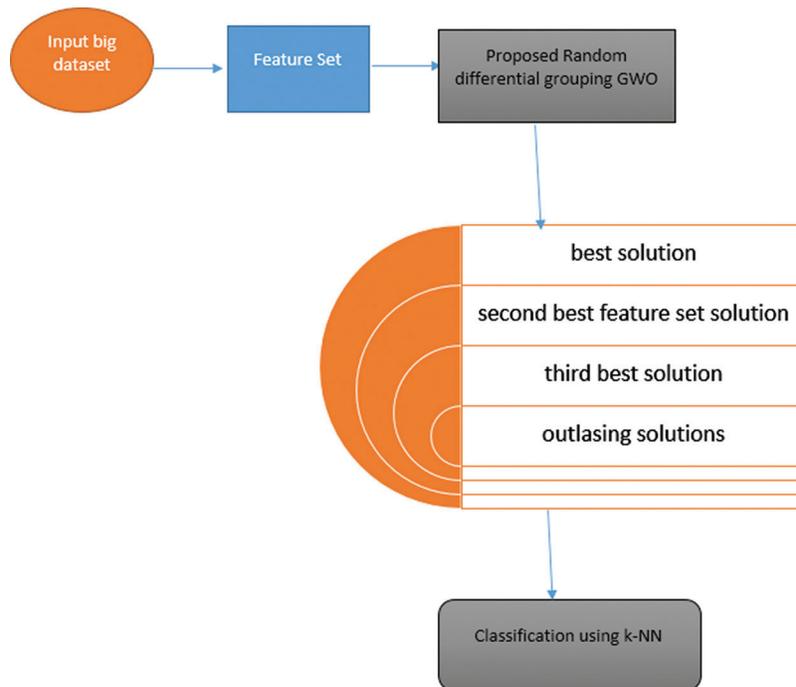
Figure 1: Architectural overview

3.1 Decomposition Using Random Differential Grouping

Interdependence searching has been proposed for the decomposition problem to address the high computational complexity of the supervised learning algorithms [25]. Our proposed decomposition system found the inter relationship between the variables using the two subsets of the decision variables rather than two whole set of decision variables. This approach will increase the speed of the decomposition process. The differential grouping [26] identifies the interaction between the variables by using the fitness value changes of the decision variables. The fitness changes cause the change of decision variable from x to y; then, we considered that x and y are in interaction and has a relationship. The differential grouping method is proven to be the best for decomposition. The existing solutions on optimizing the big data have poor performance because the feature interactions are not properly

addressed, and the dataset is statically decomposed. This process may consider only the even number of features. In this work, novel random differential grouping is proposed to dynamically decompose the big dataset and group the features that are interacting into the subset, thereby ensuring the probability of the proposed method.

The main objective of the proposed rnDG is to dynamically decompose the n feature vector into the subsets with a dimension of $m = r*c$. Dataset X consists of n feature vectors. The X is randomly decomposed into m subsets with s features in each subset where $s < n$, and it is measured as follows:

$$\{X_1 = f_1, f_2 \ldots f_s\}$$
$$\{X_2 = f_1, f_2 \ldots f_s\} \tag{1}$$
$$\{X_m = f_1, f_2 \ldots f_s\}.$$

Each subset is represented as a population in GWO for optimization. "s" is the number of features in each subset (ss). The size of the subset is "sz". Subset "$ss1$" consisting of individuals is represented as follows:

$$\{id_1 = 0, \ 1, \ 1..1\},$$
$$\{id_2 = 0, \ 1, \ 1..1\}, \tag{2}$$
$$\{id_{sz} = 0, \ 1, \ 1..1\}.$$

1 in the subset indicates that the feature is selected for feature subset, and 0 denotes that the feature is not selected for feature subset selection. Let dv be the decision variables of X, and U_x is the unit vector of the decision. X_1 is the subset of the decision variables, where $X_1 \subset X$, and UX_1 is the subset of U_X. Unit vector U = $\{u_1, u_2 \ldots u_n\}$, where $u_i = 0$.

Variable d is the differential function, and u is the vector of U_X. The directional derivative of d in u is expressed as follows:

$$X_u d(X) = \sum_{i=1}^{n} \frac{\partial d(x)}{\partial x_i} \times u_i. \tag{3}$$

The two subset interactions is declared by using the following:

All the decision variables are set as lower bounds (lb) on the search space of $X_{(l,l)}$.

The transition of decision variable subset X_1 of the search space $X_{(l,l)}$ to the upper bounds (ub) is denoted as $X_{(u,l)}$.

The fitness difference ∂_1 is calculated using $X_{(l,l)}$ and $X_{(u,l)}$ as follows:

$$\partial_1 = X_{(l,l)} - X_{(u,l)}. \tag{4}$$

The transition of decision variable subset X_2 of the search space between $X_{(l,l)}$ and $X_{(u,l)}$ is declared as the middle between the lower and the upper bounds, which are denoted as $X_{(l,m)}$ and $X_{(m,u)}$, respectively.

The fitness difference ∂_2 is calculated using $X_{(l,m)}$ and $X_{(m,u)}$ as follows:

$$\partial_2 = X_{(l,m)} - X_{(m,u)}. \tag{5}$$

If the difference between ∂_1 and ∂_2 is greater than the threshold value ϑ, then an interaction exists between the two subsets of features X_1 and X_2.

Threshold ϑ is calculated as the magnitude of the objective [27] as follows:

$$\vartheta = \alpha \times \min\{|d(x_1)|, \quad |d(x_2)| \ldots |d(x_m)|\}. \tag{6}$$

The nine features are decomposed into two subsets by using the proposed decomposition, which is represented as follows:

$$d(x) = \{f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9\}, \tag{7}$$

Such expression is decomposed using the proposed rnDG as $\{(f_1, f_7, f_9, f_2), (f_3, f_5, f_8, f_6, f_4)\}$, as shown in Fig. 2.
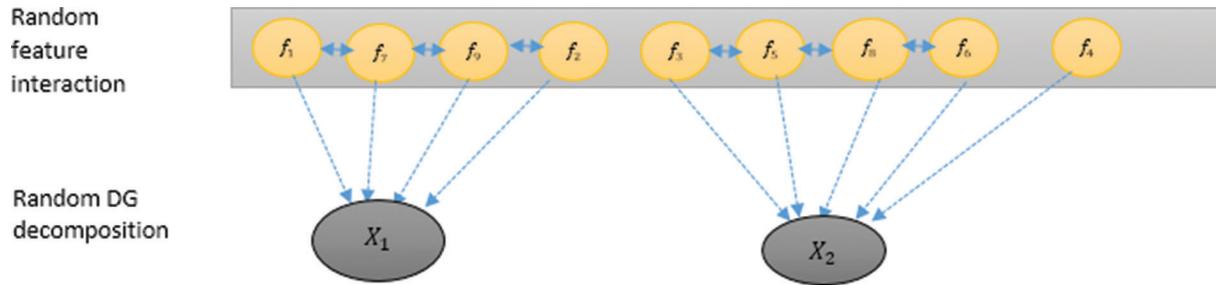
Figure 2: Proposed random differential grouping decomposition of features

---

Algorithm 1

Input: fitness function (ft), upper bounds (9ub), lower bounds (lb), and threshold ϑ

Output: subset of features

Initialize the subset size (sz), number of subsets m, and gen.

Step 1: Dynamically decompose f features into m subsets that have z features.

Step 2: gen=1, set all the decision variables to lower bound $X_{(l,l)}$=lb

Step 3: For i=1 to m

Step 4:   For j=t to sz

Step 5:   Initialize the individuals with 0 and 1

Step 6:   End for

Step 7: End for

Step 8: Calculate the fitness value $ft_{(l,l)}$=d$(X_{(l,l)})$

Step 9:   Randomly assign one feature to subset X_1. The remaining features are assigned to X_2.

Step 10: While X_2 is not empty do

Step 11:   Call algorithm 2 as $[X_1^* = interact(X_1, X_2, X_{(l,l)}, ft_{(l,l)})]$

Step 12:   If $(X_1^*==X_1)$ *then*

Step 13:   If X_1 contains a decision variable, then

Step 14:   add those features to the selected feature group 1

Step 15:   End if

Step 16:   Assign the first feature of X_2 to X_1 and delete that feature from X_2

Step 17:   Else

Step 18: $X_1=X_1^*$

(Continued)

Algorithm 1: (continued)

Step 19:   Delete the features of X_1 from X_2

Step 20:   End if

Step 21: End while

Step 22: Return the decomposed subsets.

Algorithm 2: $[X_1^* = interact(X_1, \; X_2, \; X_{(l,l)}, ft_{(l,l)})]$

Step 1: $X_{(u,l)} = X_{(l,l)}; \quad X_{(u,l)} = \text{ub}(X_1)$

Step 2: Calculate the fitness change ∂_1 by using (4)

Step 3: $X_{l,m}(X_2) = \frac{(\text{lb}(X_2) \; + \text{ub}(X_2))}{2}$ (8)

Step 4: $X_{u,m}(X_2) = \frac{(\text{lb}(X_2) \; + \text{ub}(X_2))}{2}$ (9)

Step 5:   Calculate the fitness change ∂_2 by using (5)

Step 6: If $(\partial_1 - \partial_2) > \vartheta$, then

Step 7:   If X_2 has a decision variable, then

Step 8: $X_1 = X_1 \cup_2^X$ (10)

Step 9:   End if

Step 10: End if

Step 11: Return X_1

The random differential grouping algorithm dynamically decomposes the features of the dataset into subsets. One feature is assigned to subset 1 at random. Then, the interaction between the features of the subset is calculated using algorithm 2. If no interaction occurs, then the features are assigned to one subset group. If an interaction is found using algorithm 1, then it will be divided to further two groups with equal size. The interaction between the first group X_1 with the two additional subgroups is identified and added to the respective decomposition group, such as X_1 and X_2. The next process of the proposed technique is GWO-based feature selection. The decomposed subsets are given as inputs to the GWO algorithm for feature selection.

3.2 DG-based GWO (RnDG-GWO)

GWO is considered the most advanced optimization algorithm due to the nature of hunters that are ready to catch their prey because of the crowd that is carefully organized. The GWO is a meta heuristic algorithm that is similar to GWO with the process of attacking and looking around their prey. The four levels of GWO leadership hierarchy are alpha (α), beta (β), delta (δ), and omega (ω). α is the male or female in charge of making decisions. β will help the α in making decisions. δ are the guides or elders or hunters. ω follows all wolves, and it will be the optimal solution, whereas the α, β, and ω are the first, second, and third best solutions.

In our proposed work, the decomposed subset from algorithm 1 is initialized as the population of n wolves in the first stage of algorithmic initialization. Every wolf is related to the desirable solutions, and n is the number of features in the original dataset. Fig. 3 shows the feature selection of the sample of nine features. According to the algorithm, the selected features that will improve the accuracy are represented as a binary value 1. Meanwhile, the features that will not improve the accuracy are represented as binary value 0.
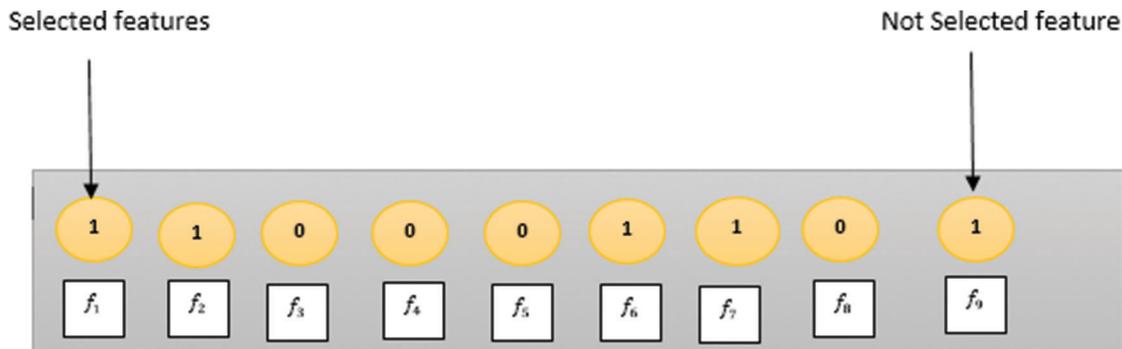
Figure 3: Feature selection (solution) representation sample

Algorithm 3:

Input: Grey wolf population X_1 and X_2, maximum no. of iteration tm, size of dataset N, control parameter c, and learning factors $c1$ and $c2$

Output: Optimal solution

Step 1: For i=1 to N

Step 2:   For j=1 to n

Step 3:   Call algorithm 1 to decompose the feature set

Step 4:   End for

Step 5:   End for

Step 6: // Use the decomposition to initialize the population

Step 7: For t=1 to tm

Step 8:   Update c using,

$$c(t) = c_s - (c_s - c_f) \times \left(\frac{1}{tm} \right)^2, \tag{11}$$

where c_s and c_f — initial and end values of the control parameter

Step 9: For i=1 to N

Step 10:   For j=1 to n

Step 11:   Calculate D_α, D_β, and D_δ using

$$D_\alpha = |E_1 \times X_\alpha - X| \tag{12}$$

$$D_\beta = |E_2 \times X_\beta - X| \tag{13}$$

$$D_\delta = |E_3 \times X_\delta - X| \tag{14}$$

where

$$X_1 = X_\alpha - C_1 \times D_\alpha, \tag{15}$$

$$X_2 = X_\alpha - C_2 \times D_\alpha, \tag{16}$$

(Continued)

Algorithm 3: (continued)

$$X_3 = X_\alpha - C_3 \times D_\alpha, \tag{17}$$

$$E = 2 * r2, \tag{18}$$

$$C = |2a * r1 - a|, \tag{19}$$

$$a = 2 - t* \frac{2}{tmax}, \tag{20}$$

r1, r2 ∈ [0, 1].

Step 12: Calculate the coefficient of weights $w1$, $w2$, and $w3$ [28] by using

$$w_1 = \frac{|X_1|}{|X_1 + X_2 + X_3|}, \tag{21}$$

$$w_2 = \frac{|X_2|}{|X_1 + X_2 + X_3|}, \tag{22}$$

$$w_3 = \frac{|X_3|}{|X_1 + X_2 + X_3|}, \tag{23}$$

Step 13: Update the position of the grey wolf by using 15, 16, 17, and

$$X_{(t+1)} = c_1 r_1 \left(\sum_{i=1}^{3} w_i * X_i(t) \right) + c_2 r_2 \left(\sum_{i=1}^{3} X_{ibest} * X_i(t) \right), \tag{24}$$

where c_1 is the social learning factor, and c_2 is the cognitive learning factor

Step 14:   End for

Step 15: End for

Step 16: Update the fitness value by using

$$fitness = \alpha \gamma R \ (D) + \beta \frac{|S|}{|X|}, \tag{25}$$

|S| — length of the selected subcomponent cardinality

|X| — cardinality of all dataset features

α and β — weight parameters for accurate classification, where α ∈ [0, 1], and β = 1 − α

Step 17: End for.

In this proposed work, kNN is used as a classifier for the classification problem. The distance used to classify the data by using k-NN classifier is the Euclidean distance [28,29] used to find the k nearest neighbors by using:

$$ED_n(P, \ Q) = \sqrt{\sum_{I=1}^{n} (Q_i - P_i)^2}, \tag{26}$$

where n is the number of features in the dataset, and Q_i and P_i are the selected features in the given subset component. The dataset is divided into two parts for testing and training by using the k-fold validation. In this work, $k = 5$. Meanwhile, $k-1$ partitions are considered for training and the remaining ones for testing. The values generated from the search agent of GWO are continuous values. Hence, the resulted format is converted into the binary format by using the transform function called sigmoid function to improve the classification accuracy. The sigmoid function is as follows:

$$X_i = \frac{x_i - min}{max - min}, \tag{27}$$

$$X_{si} = \frac{1}{1 + e^{-X_i}}, \tag{28}$$

$$X_b = \begin{cases} 1 & if \ R > X_{si} \\ 0 & if \ R < X_{si} \end{cases}, \tag{29}$$

where X_{si} is the search agent; $i - 1, \ldots, n$ denotes the number of features; max and min are the maximum and minimum values of the continuous feature vector; and $X_b \in [0, 1]$ is the classification result. The overall workflow of our proposed feature selection scheme is represented in Fig. 4.

Fig. 4 illustrates the proposed feature selection algorithm with classification. The input dataset features are first divided into subcomponents by using the random differential grouping. Then, the decomposed feature subsets are optimized using GWO to provide the optimal feature set for classification by using kNN. The prediction accuracy using the proposed feature selection is analyzed with a big dataset. Hence, our proposed feature selection algorithm using random differential grouping with GWO minimizes the dataset features by selecting the most relevant features and maximizes the classification accuracy based on the performance of the kNN classifier on various big datasets.

4 Results and Discussions

This section discusses the experimented result of the proposed feature selection algorithm using MATLAB (2018a). It is examined with the benchmark datasets from the UCI machine learning repository to check the efficiency and strength of the proposed algorithm. The datasets with original and reduced features using the proposed algorithm is shown in Tab. 1.

Existing standard GWO and GWO with PSO are used to compare the efficiency of the proposed algorithm. The algorithm parameters are taken from the literature and are equally compared. In this proposed work, the kNN classifier is used with the hybrid feature selection algorithm and considered the supervised learning for a fast implementation. The dataset is divided in the ratio of 4:1 as training and testing because this work uses the fivefold cross-validation technique. The comparison of selecting the number of features using the existing algorithms with proposed are shown in Tab. 2, and the comparison results are illustrated in Fig. 5.

The experimented results shows that our proposed algorithm efficiently reduces the irrelevant features than the original and other standard feature selection algorithms. Less number of relevant features leads to improved classification accuracy. The accuracy is calculated using the following equation:

$$Accuracy = \frac{Correctly \ classified \ instances}{total \ number \ of \ instances} \times 100. \tag{30}$$

The accuracy comparison of the original dataset and feature selection algorithm accuracy are shown in Tab. 3 and illustrated in Fig. 6. The evaluated results show that the proposed rnDG-GWO feature selection-based kNN classification obtains higher accuracy of 92%–98% on various benchmark datasets compared with the other existing feature selection-based classification accuracy.
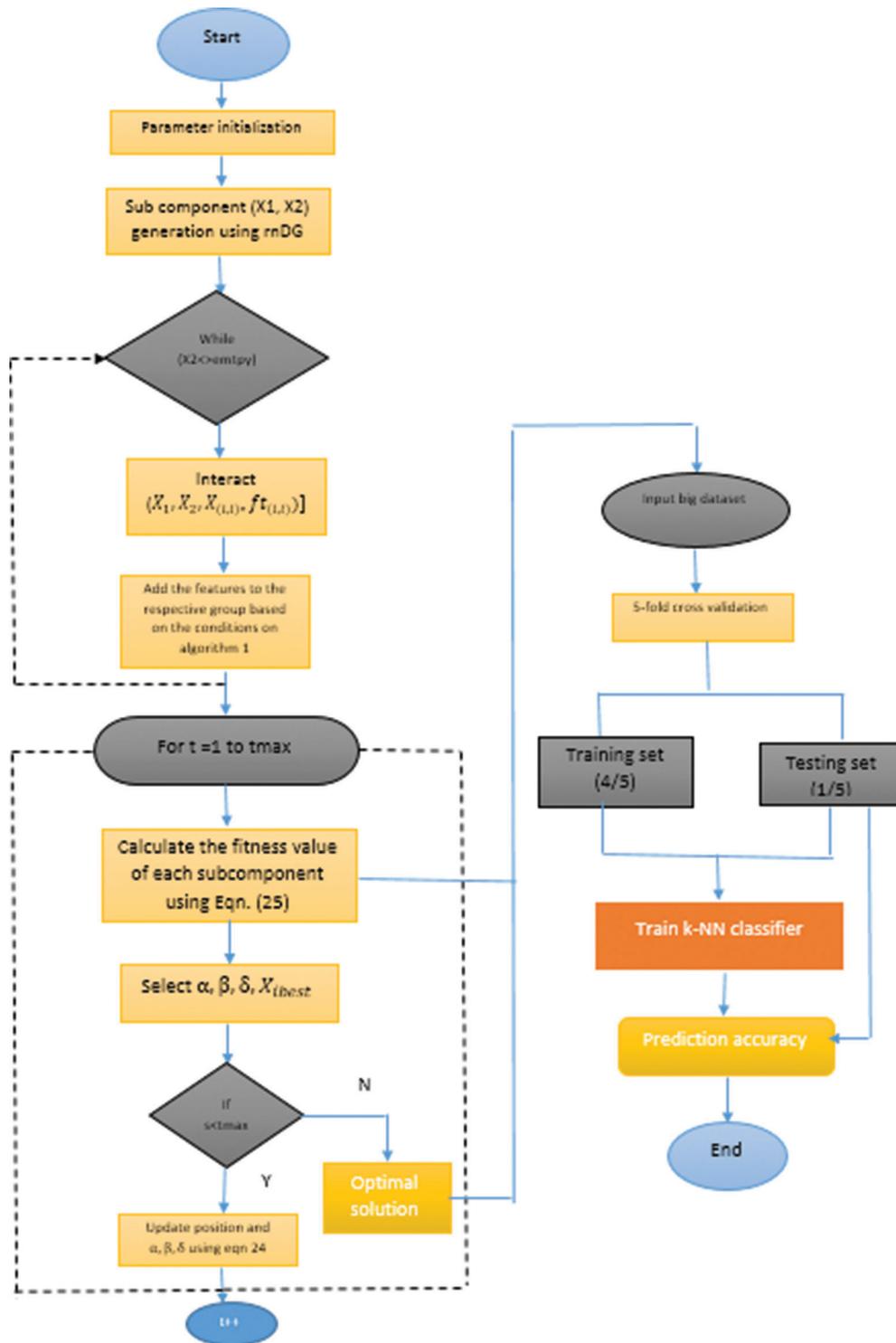
Figure 4: Work flow diagram of proposed feature selection with classification

Table 1: Feature selection using proposed model

Datasets	Original features	Reduced features using rnDG-GWO
Breast cancer	9	3
WineEW	13	4
Parkinson's	22	5
Diabetes	19	4
vehicle	18	5

Table 2: Feature selection comparison of the FS algorithms

Datasets	Original features	Standard PSO	Standard GWO	GWO +PSO	Reduced features using rnDG-GWO
Breast cancer	9	5	5	4	3
WineEW	13	8	7	6	4
Parkinson's	22	10	12	6	5
Diabetes	19	8	10	5	4
Vehicle	18	12	11	6	5
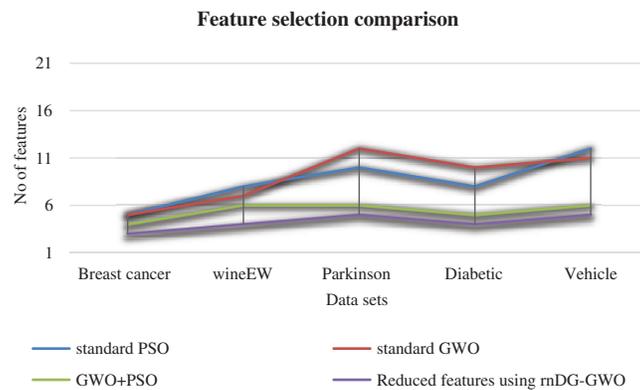
Figure 5: Feature selection comparison of the various algorithms

Table 3: Accuracy comparison of various algorithms

Datasets	Original accuracy without FS	Standard PSO	Standard GWO	GWO +PSO	rnDG-GWO
Breast cancer	96	96.5	96	96.7	98.43
WineEW	92	91	91.8	92.345	97.54
Parkinson's	78	81	82	92	97.34
Diabetes	60	62	63.78	65.1	92.43
Vehicle	72	78	82	96	98.23
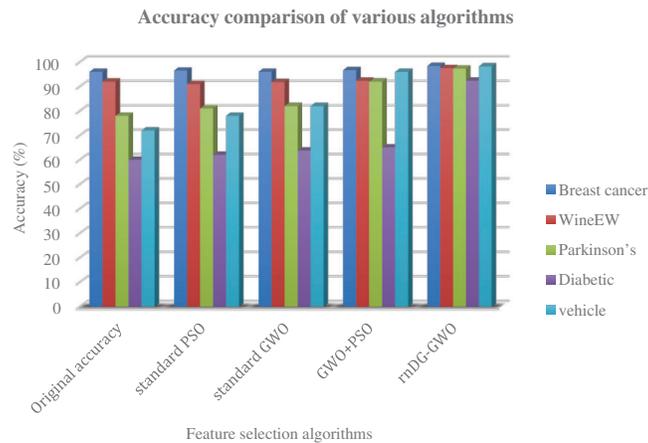
Figure 6: Accuracy of the classification with FS and without FS of various algorithms

Hence, our proposed decomposed random differential grouping with GWO (DrnDG-GWO) for selecting optimal features provides high accuracy of prediction on breast cancer, wineEW, Parkinson, diabetes, and vehicle datasets. The computational time comparison of these algorithms are shown in Tab. 4 and illustrated in Fig. 7.

Table 4: Computational time (seconds) comparison of the proposed and existing algorithms

Datasets	standard PSO	standard GWO	GWO+PSO	rnDG-GWO
Breast cancer	8.8	8.48	7	6
WineEW	8.4	7.23	6	5.2
Parkinson's	8.5	7.34	6	5
Diabetes	9.6	9.5	7.4	5.3
Vehicle	8.9	9.25	6.6	5.12
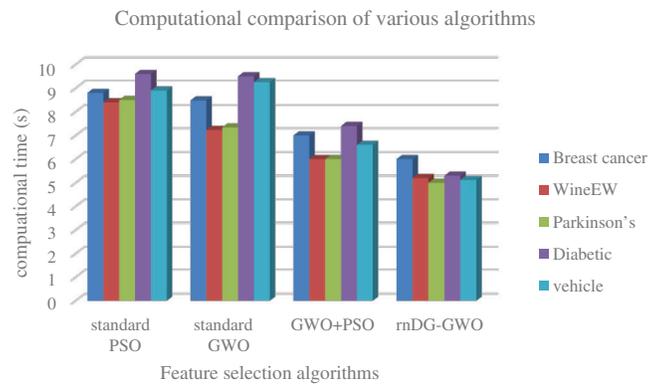
Figure 7: Computational time comparison in seconds

Tab. 4 illustrates the comparison of various feature selection methods based on classification and terms of computational time. The observed result shows that our proposed algorithm predicts in the shortest amount of time compared with the other traditional algorithms.

The observed results confirmed that the accuracy is comparatively low when no feature selection is utilized, and the classification accuracy considerably higher than the original when the proposed feature selection is used. The work aims to reduce the number of features and increase the classification accuracy with the experimentation. Hence, our proposed DrnDG-GWO feature selection-based classification is best in terms of reducing the number of features, increasing the accuracy, and reducing the computational time.

5 Conclusions

Feature selection is considered an important preprocessing step in selecting the appropriate feature from a huge dataset. Research on various selection methodology is highly improving. In our research, we focus on accurate selection of features from the dataset and increasing the accuracy of the classifier. Swarm intelligence is considered an interesting metaheuristic method for various optimization processes. Here, we use GWO, which is a swarm optimization technique for selecting the feature with global optimal solution. Decomposition of data with random differential grouping process optimally reduces the feature sets from the huge original dataset. This reduced feature set is optimized on the basis of the three optimal solutions of GWO. The accuracy, number of features retrieved, and computational cost are compared to contrast the efficiency of the proposed algorithm with existing standard GWO and GWO with PSO for feature selection. In this comparative analysis, our proposed feature optimization technique shows improved result. The numbers of features and classification are proportionate. The optimal solution would be to reduce features that are highly relevant. The classification accuracy will be relatively high when we perform classification with reduced features. In our research work, we focus on the above-mentioned two strategies for improving the feature selection and optimization. The feature reduction process with optimal feature selection is performed on the basis of decomposition and random differential grouping. Finally, reduced features are further optimized for optimal solution. The KNN classifier is used for feature classification. The result is evaluated with existing models, and our research work indicates that it outperforms them all. Decision tree-based decomposition can be suggested with swarm intelligence techniques in the future. Machine learning strategies can be used for feature selection to obtain an optimal solution.

Conflicts of Interest: The authors declare that no conflict of interest regarding the publication of the paper.

References

[1] S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software,* vol. 69, pp. 46–61, 2014.

[2] Y. Sun, S. Todorovic and S. Goodison, "Local-learning-based feature selection for high-dimensional data analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1610–1626, 2009.

[3] S. Fong, X. S. Yang and S. Deb, "Swarm search for feature selection in classification," in *Proc. ICCSE, Sydney*, Australia, pp. 902–909, 2013.

[4] S. Fong, J. Liang, R. Wong and M. Ghanavati, "A novel feature selection by clustering coefficients of variations," in *Proc. ICDIM, Phitsanulok*, Thailand, pp. 205–213, 2014.

[5] D. Peralta, S. Del Río, S. Ramírez Gallego, I. Triguero and J. M. Benitez *et al.*, "Evolutionary feature selection for big data classification: A mapreduce approach," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[6] S. Harde and V. Sahare, "ACO swarm search feature selection for data stream mining in big data," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, no.12, pp. 12087–12089, 2015.

[7] R. Senthamil Selvi and M. Valarmathi, "An improved firefly heuristics for efficient feature selection and its application in big data," *Biomedical Research*, vol. 28, 2017.

[8] F. Viegas, L. Rocha, M. Goncalves, F. Mourao, G. Thiago *et al.*, "A genetic programming approach for feature selection in highly dimensional skewed data," *Neurocomputing*, vol. 273, pp. 554–569, 2018.

[9] D. Anusuya, R. Senthilkumar and T. S. Prakash, "Evolutionary feature selection for big data processing using Map reduce and APSO," *International Journal of Computational Research and Development*, vol. 1, no. 2, pp. 30–35, 2017.

[10] W. Jun, W. Shitong and F. Chung, "Positive and negative fuzzy rule system, extreme learning machine and image classification," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 4, pp. 261–271, 2011.

[11] G. B. Huang, H. Zhou, X. Ding and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2011.

[12] J. Chen, G. Zheng and H. Chen, "ELM mapreduce: map reduce accelerated extreme learning machine for big spatial data analysis," in *Proc. ICCA, Hangzhou*, China, pp. 400–405, 2013.

[13] C. Tekin and M. Vander Schaar, "Distributed online big data classification using context information," in *Proc. Allerton*, Monticello, IL, USA, pp. 1435–1442, 2013.

[14] K. Grolinger, M. Hayes, W. A. Higashino, A. L'Heureux, D. S. Allison *et al.*, "Challenges for mapreduce in big data," in *Proc. IEEE World Congress on Services, IEEE, Anchorage*, AK, USA, pp. 182–189, 2014.

[15] P. Rebentrost, M. Mohseni and S. Lloyd, "Quantum support vector machine for big data classification," *Physical Review Letters*, vol. 113, no. 13, 2014.

[16] Y. Liu, F. Tang and Z. Zeng, "Feature selection based on dependency margin," *IEEE Transactions on Cybernetics*, vol. 45, no. 6, pp. 1209–1221, 2014.

[17] V. Stanovov, C. Brester, M. Kolehmainen and O. Semenkina, "Why don't you use evolutionary algorithms in Big data?," in *IOP Conf. Series: Materials Science and Engineering*, vol. 173, no. 1, 2017.

[18] R. Wang, F. Zhang, T. Zhang and P. J. Fleming, "Cooperative co-evolution with improved differential grouping method for large-scale global optimisation," *International Journal of Bio-Inspired Computation*, vol. 12, no. 4, pp. 214–225, 2018.

[19] Y. Sun, M. Kirley and S. K. Halgamuge, "On the selection of decomposition methods for large scale fully non-separable problems," in Proc. ICGA,Yangon, Myanmar, pp. 1213–1216, 2015.

[20] Y. Sun, M. Kirley and S. K. Halgamuge, "A recursive decomposition method for large scale continuous optimization," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 5, pp. 647–661, 2017.

[21] X. M. Hu, F. L. He, W. N. Chen and J. Zhang, "Cooperation coevolution with fast interdependency identification for large scale optimization," *Information Sciences*, vol. 381, pp. 142–160, 2017.

[22] A. Chen, Z. Ren, Y. Yang, Y. Liang and B. Pang, "A historical interdependency based differential grouping algorithm for large scale global optimization," in *Proc. GECCO*, Kyoto, Japan, pp. 1711–1715, 2018.

[23] M. Shanmugam, A. Ramasamy, S. Paramasivam and P. Prabhakaran, "Monitoring the turmeric finger disease and growth characteristics using sensor based embedded system—a novel method," *Circuits and Systems*, vol. 7, no. 08, pp. 1280–1289, 2016.

[24] S. Maheswaran, M. Ramya, P. Priyadharshini and P. Sivaranjani, "A real time image processing based system to scaring the birds from the agricultural field," *Indian Journal of Science and Technology*, vol. 9, no. 30, pp. 1–5, 2016.

[25] H. Ge, L. Sun, X. Yang, S. Yoshida and Y. Liang, "Cooperative differential evolution with fast variable interdependence learning and cross-cluster mutation," *Applied Soft Computing*, vol. 36, pp. 300–314, 2015.

[26] M. N. Omidvar, X. Li, Y. Mei and X. Yao, "Cooperative co-evolution with differential grouping for large scale optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 378–393, 2013.

[27] Y. Mei, M. N. Omidvar, X. Li and X. Yao, "A competitive divide-and-conquer algorithm for unconstrained large-scale black-box optimization," *ACM Transactions on Mathematical Software (TOMS)*, vol. 42, no. 2, pp. 1–24, 2016.

[28] M. El-Hasnony, S. I. Barakat, M. Elhoseny and R. R. Mostafa, "Improved feature selection model for big data analytics," *IEEE Access*, vol. 8, pp. 66989–67004, 2020.

[29] M. AbdelBasset, D. Shahat, I. E. Henawy, V. H. C. Albuquerque and S. Mirjalili, "A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection," *Expert Systems with Applications*, vol. 139, pp. 122824–122837, 2020.