STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, March 2026, pp 2155–2163.
Published online in International Academic Press (www.IAPress.org)

Enhancing Gene Selection in Microarray Dataset Using Binary Gray Wolf Optimization Algorithm and Statistical Dependence Technique

Hanadi Dawood Saleem, Talal Fathal Hussein, Fatima Mahmood Hasan, Omar Saber Qasim*

Department of Mathematics, Mosul University, Iraq

Abstract The task of choosing the most pertinent subset of features is crucial for boosting classification accuracy while reducing computational demands in machine learning. To address this issue, the paper explores two strategies: Binary Gray Wolf Optimization (BGWO) and the Statistical Dependence (SD) method. The procedure starts with the SD technique to identify the features that have the greatest impact on classification results. Subsequently, the BGWO algorithm is utilized alongside the K-Nearest Neighbors (KNN) classifier to further refine the selection to the most vital features. The proposed SD-BGWO method surpasses traditional techniques by either enhancing classification accuracy or by lowering the number of features needed, thus optimizing the feature selection process in both efficiency and effectiveness.

Keywords Statistical dependence, Binary Gray wolf optimization, Classification, Gene selection

AMS 2010 subject classifications 62-XX, 92B05, 92D20, 92C40,92D50

DOI: 10.19139/soic-2310-5070-2430

1. Introduction

This algorithm was developed by mathematically modeling the social hierarchy and hunting behavior of gray wolves, translating these natural processes into a computational framework. The performance evaluation of the GWO algorithm demonstrated its superior exploration and exploitation capabilities compared to other swarm intelligence methods. The fundamental GWO algorithm was initially designed for continuous search spaces; however, researchers adapted it to operate in discrete search spaces by modifying the search mechanism. The Binary Gray Wolf Optimization (BGWO) algorithm was created as a consequence of this adaption, and it utilizes the binary values 0 and 1 for operation within binary search domains [2, 3, 4].

Entropy between two distinct random variables is used to calculate statistical dependence (SD), one method of filter calculation. The entropy of a variant is calculated as the average data required to describe the variant. This particular filter is used. to rearrange the characteristics of datasets in line with their importance in influencing the classification accuracy [5, 6].

The hybrid algorithm (SD-BGWO) to select the best features is suggested in this studyThe goal of the suggested SD-BGWO algorithm is to speed up the process of locating and choosing the most crucial features while combining the benefits of the SD and BGWO algorithms. The great efficiency of this technique has been proved by experimental results, particularly when the sample size is small and the number of features is big.

In Sections 3 and 4, a synopsis of the SD approach and feature selection is given before the GWO algorithm is introduced. We discuss the suggested SD-BGWO algorithm in Section 5. While Section 7 gives the most significant overall conclusions, Section 6 discusses the study's complete results.

*Correspondence to: Omar Saber Qasim (Email: omar.saber@uomosul.edu.iq). Department of Mathematics, Mosul University, Mosul, Nineveh, Iraq (41002).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

2. Grey wolf optimization (GWO)

Swarm Intelligence (SI) techniques represent a class of population-based metaheuristic that simulate the collective behavior of animal groups, such as swarms, flocks, herds, and schools. Among these, the Grey Wolf Optimizer (GWO) emerges as an innovative algorithm within the SI category, uniquely modeled after the hierarchical structure and hunting patterns of grey wolves. Unlike other SI techniques, GWO specifically emulates the social order observed in grey wolves during their hunting processes [1, 7].

Grey wolves are apex predators and members of the Canidae family. known for hunting in highly organized packs, which typically consist of 5 to 12 individuals. These packs are characterized by a well-defined social hierarchy. As depicted in Figure 1,The social structure is divided into several layers. The Alpha wolf, which can be either male or female, stands at the top of this hierarchy, making crucial decisions regarding the pack's movements, hunting times, and resting locations. The Alpha is not necessarily the strongest, but rather the most adept at leading and organizing the pack [8]. The Beta wolf is directly beneath the Alpha and serves as a supporting leader by helping to keep the pack in order and oversee its activities. The Beta acts as a mentor and advisor to the Alpha and is the most likely candidate to take over leadership if the Alpha is incapacitated or steps down.

The third rank is δ (delta (Dl).), which is mainly responsible for monitoring the boundaries of the territory, warning the wolf pack in case of any danger and caring for the weak and wounded grey wolves. A wolf is referred to as delta if it does not fit into any of the aforementioned classifications. Alpha and beta wolves are subordinate to and controlled by the dominant delta wolves [9]. The fourth rank (lowest level) is the lowest level grey wolf in the society, called ω (Omega (Om)), which has to submit to all the other dominant grey wolves and is allowed to eat in the end. It may seem that the ω wolves do not play an important role in the wolf pack, but it is necessary for adjusting the internal relations between the populations [1, 10]. In conclusion, grey wolves hunt in groups. Hunting consists of the following major stages:

- The first stage is the near aim of the hunting during continuous pursuit.
- The second stage: start chasing the aim by encircling, and harassing it until its movement is stopped.
- The third stage: raid the aim.

3. Mathematical modelling

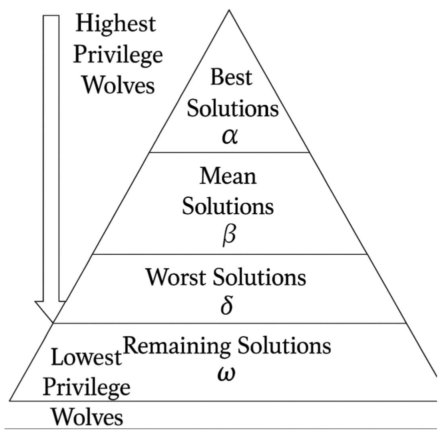

Figure 1. Typical wolf family hierarchies and behaviors encompass: (a) a structured leadership hierarchy; (b) two-dimensional position vectors that identify grey wolf locations; and (c) three-dimensional position vectors detailing the spatial distribution of grey wolves. [8].

The Wolves, according to the GWO, are the best option out of all the alternatives, whereas the second and third optimal solutions are denoted by β and δ, respectively. The alternative options are mentioned by ω. The chief wolves are in charge of the hunting procedure.. As the first step of catching, the catching goal happens by surrounding behavior of α, β and δ [11].

Encircling prey: The following mathematical model can be used to explain how grey wolves encircle each other:

$$\overrightarrow{D} = \left|\overrightarrow{C} \cdot \overrightarrow{X}_p(t) - \overrightarrow{X}(t)\right|, \tag{1}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X}_p(t) - \overrightarrow{A} \cdot \overrightarrow{D}, \tag{2}$$

In this particular instance, t is the iteration. The location of the prey (target) is denoted by .,$\overrightarrow{X}_p$, the gray wolf's role by $\overrightarrow{X}$, and the distance vector by , $\overrightarrow{D}$. Vector entry-wise multiplication is indicated by the operator.

where $\overrightarrow{A}$ and $\overrightarrow{C}$ are control coefficient vector and convergence factor set respectively and are calculated as follows:

$$\overrightarrow{A} = 2\overrightarrow{a} \cdot \overrightarrow{r} - a, \tag{3}$$

$$\overrightarrow{A} = 2\overrightarrow{r_2}, \tag{4}$$

where components of are progressively reduced from 2 to 0 throughout the course of iterations, and r_1, r_2 are random vectors in $[0, 1]$ [1].

3.0.1. Hunting: When the prey is closest to the gray wolves α, β and δ their positions can be used to determine the prey's position The three optimum solutions at the iteration t are represented by $\overrightarrow{X_\alpha}$, $\overrightarrow{X_\beta}$ and $\overrightarrow{X_\delta}$ which represent the position of α, β and δ wolves, respectively. The following equations represent the update for the wolf positions as a result [8, 9]:

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{C_1} \cdot \overrightarrow{X_\alpha} - \overrightarrow{X}\right|, \quad \overrightarrow{D_\beta} = \left|\overrightarrow{C_2} \cdot \overrightarrow{X_\beta} - \overrightarrow{X}\right|, \quad \overrightarrow{D_\delta} = \left|\overrightarrow{C_3} \cdot \overrightarrow{X_\delta} - \overrightarrow{X}\right| \tag{5}$$

$$\overrightarrow{X_1} = \left|\overrightarrow{X_\alpha} - \overrightarrow{A_1} \cdot \overrightarrow{D_\alpha}\right|, \quad \overrightarrow{X_2} = \left|\overrightarrow{X_\beta} - \overrightarrow{A_2} \cdot \overrightarrow{D_\beta}\right|, \quad \overrightarrow{X_3} = \left|\overrightarrow{X_\delta} - \overrightarrow{A_3} \cdot \overrightarrow{D_\delta}\right| \tag{6}$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3}. \tag{7}$$

The parameters $\overrightarrow{a}$ are updated using the equation that follow:

$$\overrightarrow{a} = 2 - t \cdot \frac{2}{Max_{iter}}, \tag{8}$$

where Max_{iter} is the maximum number of t iterations allowed for optimization [3, 4].
and t is the number of iterations. Figure 2 displays the Grey Wolf Optimizer (GWO) algorithm 1 pseudo-code.

Start
To select a wolf population to start with X_i $(i = 1, \ldots, n)$
Select value at random. A, C, and a
Calculate each wolf's fitness function.
$x_\alpha =$ the first-best position vector
$x_\beta =$ the position that is second-best vector
$x_\delta =$ a position vector that ranks third best.
while ($t < Max_{iter}$ of iterations)
for each position vector
The current search agent's position is updated using equation (7)
end for
Update A, a, and c
To compute each search agent's fitness function
x_α, x_β, and $x_\delta$should be updated.
set $t = t + 1$
end while
return x_α
END

Algorithm 1. GWO algorithm's pseudocode.

4. Binary gray wolf optimization (BGWO)

In the Grey Wolf Optimizer (GWO), wolf positions can be defined anywhere within a continuous search space. However, when it comes to specific tasks like feature selection, these positions are constrained to binary values of 0 or 1. This restriction converts the search space in the Binary Grey Wolf Optimizer (BGWO) into a hypercube. Wolves adjust their positions by altering certain parameters, enabling them to move towards or away from the hypercube. This process is guided by specific mathematical equations [12, 13].

This work proposes a binary version of the Grey Wolf Optimization (GWO) technique for selecting features. Here, three vectors are integrated into the equations used to update wolf positions. These vectors represent the top three candidate solutions, with each wolf's position being adjusted accordingly. All candidate solutions are represented in binary form, ensuring that the wolves' positions align within the hypercube. The GWO algorithm is applied to update these positions, with a focus on maintaining binary constraints as outlined in Equation (9). The core equation for updating The GWO algorithm is explained in depth below [13, 14]:

$$X_i^{t+1} = Crossover\left(X_1, X_2, X_3\right), \tag{9}$$

The wolves in BGWO are represented by the binary vectors (x_1, x_2, x_3), where (a, b, c) is a desirable intersection of (x_1, x_2, x_3) and (crossover(a, b, c)). The wolves employ equations (10), (13) and (16) to determine α, β, and δ, respectively.

$$X_1^d = \begin{cases} 1 & \text{if} \quad X_\alpha^d + bstep_\alpha^d \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

whereX_α^d is a vector which demonstrates wherever the α wolf is positioned in dth,dimension, and $bstep_\alpha^d$is a binary step in the dth dimension. This formula is used to determine $bstep_\alpha^d$:

$$bstep_\alpha^d = 1 \text{ if } cstep_\alpha^d \geq \text{rand}, \tag{11}$$

Rand is a random number in the closed range $[0, 1]$ that is generated by a uniform distribution, where $cstep_\alpha^d$ is the step size in the dimension d continuous-valued.

$cstep_\alpha^d$ It is computed using the sigmoid function and the following equation [14]:

$$cstep_{\alpha}^{d} = \frac{1}{1 + e^{-10\left(A_1^d D_{\alpha}^d - 0.5\right)}}, \tag{12}$$

In the dimension, A_1^d and $D_{\alpha}^d$are computed using equations (3) and (5).X_2^d and X_3^d can be estimated by the subsequent equations.

$$X_2^d = 1 \text{ if } X_{\beta}^d + bstep_{\beta}^d \geq 1, \tag{13}$$

$$bstep_{\beta}^d = 1 \text{ if } cstep_{\beta}^d \geq \text{rand}, \tag{14}$$

$$cstep_{\beta}^{d} = \frac{1}{1 + e^{-10\left(A_1^d D_{\beta}^d - 0.5\right)}}, \tag{15}$$

$$X_3^d = \begin{cases} 1 & \text{if} \quad X_{\delta}^d + bstep_{\delta}^d \geq 1 \\ 0 & \quad\text{otherwise} \end{cases} \tag{16}$$

$$bstep_{\delta}^d = \begin{cases} 1 & \text{if} \quad cstep_{\delta}^d \geq \text{rand} \\ 0 & \quad\text{otherwise} \end{cases} \tag{17}$$

$$cstep_{\delta}^{d} = \frac{1}{1 + e^{-10\left(A_1^d D_{\delta}^d - 0.5\right)}}, \tag{18}$$

The intersection will be applied to the solutions of equations a, b and c as follows:

$$X_d = \begin{cases} a_d & \text{if} \quad \text{rand} < \frac{1}{3} \\ b_d & \text{if} \quad \frac{1}{3} \leq \text{rand} < \frac{2}{3} \\ c_d & \quad \text{otherwise} \end{cases} \tag{19}$$

is the result in dimension d. of the transaction. This random number constrains, within a closed range, the binary values of the first three parameters of the normal distribution.[15, 16]. The pseudo code for the BGWO algorithm 2 is displayed .

```
Start
    Select X_i (i = 1, ..., n)
    Select value A, a, and C
    Calculate each wolf's fitness function
        x_α = the first-best position vector
        x_β = the position that is second-best vector
        x_δ = a position vector that ranks third best.
  while (t < Max_iter of iterations)
    for each position vector
        Compute x_1; x_2; x_3 using Eqs. (10), (13) and (16)
        Apply crossover among x_1; x_2; x_3 using Eq. (9)
    end for
        Update A, a, and C
        Compute the fitness function
        Update x_α, x_β and x_δ
        set t = t + 1
        end while
  return x_α
END
```

Algorithm 2. The pseudo code of the BGWO

5. Feature Selection

Selecting the most relevant features is essential for optimizing model performance and reducing dataset complexity. Various strategies are employed for this task. The Statistical Dependence (SD) approach focuses on the interrelationships between dataset features to gauge their significance, helping to filter out less useful data. Alternatively, metaheuristic algorithms, such as the Binary Gray Wolf Optimizer (BGWO), are used when datasets are transformed into discrete formats. BGWO utilizes binary vectors to identify and select key features. Both methods aim to streamline datasets by focusing on the most impactful features, thereby enhancing the efficiency and accuracy of predictive models [17, 18].

6. Statistical Dependence (SD)

Calculating feature dependence with associated class label values is more precise when done using the statistical dependency (SD) feature ranking approach. The first step is to calculate each feature value dataset feature at a certain stage of the quantization scale (QS). Flexible and effective measurement of the characteristic quantization scale enables each bin to hold roughly identical samples over the whole dataset. Instead of using the typical uniform QS in this method, bins are chosen to provide some statistical validity to the occurrence of various quantization levels , and Eg. (20) is used to calculate the SD between the y class labels and the discrete value of the x feature. [19].

$$\sum_{x\in X}\sum_{y\in Y} p\left(x,y\right)\frac{p\left(x,y\right)}{p\left(x\right)p\left(y\right)}. \tag{20}$$

If Eq. (20) generates a greater value.The feature values and the category labels have a significant degree of dependence.The minimum value is produced, This implies that the features are totally independent of labels [6, 20].

7. The Proposed Hybrid Algorithm SD-BGWO

The Statistical Dependence (SD) method is used as a first and crucial step in the SD-BGWO hybrid strategy to assess and rank the characteristics according to their impact on classification accuracy. The following feature selection procedure is guided by this ranking. During the Binary Grey Wolf Optimization (BGWO) stage, features are chosen using a binary vector, which is randomly started with a length equal to the entire number of features. A value of "1" denotes selection, while a value of "0" denotes exclusion. Each dataset is split into five folds (K=5) using the K-Fold Cross-Validation technique to ensure robustness and reduce bias. Four folds are utilized for training and one for testing in each iteration. As depicted in Figure 4, this approach ensures that only the most relevant features, identified in the SD phase, are selected for further analysis, thereby enhancing the effectiveness of the feature selection process:

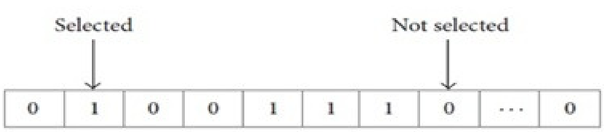

Figure 2. A description of the features in BGWO

A binary solution vector utilized in the BGWO stage is shown in the above figure. Features with a value of "1" are picked for categorization, while those with a value of "0" are excluded from the chosen feature set. In order to estimate wolf positions, the BGWO algorithm first determines classification accuracy using a KNN classifier, which is then employed within a relevance function.

$$FF = 0.1 * C + 0.9 \left(\frac{q}{p - q} \right), \tag{21}$$

where (p) is the total number of features,(c) is the classification accuracy, and (q) is the size of the chosen gene collection.in Algorithm 3 displays the pseudoplot of the suggested SD-BGWO framework.

```
Start
    A wolf population X_i (i=1.....n) is chosen.
    Choose an initial value. A, a, and C
    Determine the SD approach for feature selection
    Compute the fitness function for every wolf.
        x_α = the first-best position vector
        x_β = the position that is second-best vector
        x_δ = a position vector that ranks third best.
  while (t < Max_iter of iterations)
    for each position vector
        Compute x_1; x_2; x_3 using Eqs. (10), (13) and (16)
        Apply crossover among x_1; x_2; x_3 using Eq. (9)
    end for
        Update A, a, and C
        Compute the fitness function
        Update x_α, x_β and x_δ
        set t = t + 1
        end while
  return x_α
END
```

Algorithm 3. The SD-BGWO algorithm's proposed pseudocode.

The suggested SD-BGWO method has many drawbacks despite its encouraging outcomes. It mainly depends on the KNN classifier's performance, which can impact the total accuracy if the parameters are not adjusted correctly. Furthermore, overfitting and high computational cost are potential problems for the approach, particularly when working with small or high-dimensional data.

8. Experimental Results and Discussion

The hybrid method was validated using SD-BGWO on three distinct classification datasets (DLBCL, Colon, and Prostate).80% of each dataset was used for training and 20% for testing each time, using the K-Fold Cross-Validation technique to divide each dataset into five equal halves (K=5). Using this approach, the performance of the proposed model was fairly and properly assessed on three classification datasets.

Table 1. An explanation of the datasets used

Data	#Samples	#Features	Target class
Dataset1 (Prostate)	102	12601	2
Dataset2 (DLBCL)	77	7130	2
Dataset3 (Colon)	62	2001	2

Table 2. A comparison of the average feature selection for the SD-BGWO and the BGWO

Dataset	# Feature selection SD-BGWO	# Feature selection BGWO	# All features
Dataset1 (Prostate)	16.4	30	12601
Dataset2 (DLBCL)	20.8	30	7130
Dataset3 (Colon)	19.8	30	2001

Table 3. Shows the average experimental results for the SD-BGWO and BGWO

Datasets	Algorithms	Training dataset			Testing dataset	
		CA	SE	SP	MCC	CA
Dataset1	SD-BGWO	0.9265	0.9599	0.8904	0.8533	0.9412
		(0.0294)	(0.0258)	(0.0349)	(0.0594)	(0.0360)
	BGWO	0.6735	0.6269	0.8122	0.3818	0.7941
		(0.0408)	(0.0630)	(0.0472)	(0.0431)	(0.01120)
Dataset2	SD-BGWO	0.9885	0.9952	0.9714	0.9689	0.9200
		(0.0105)	(0.0106)	(0.0391)	(0.0286)	(0.0632)
	BGWO	0.9000	0.9009	0.9058	0.7231	0.9640
		(0.0408)	(0.1060)	(0.0590)	(0.1396)	(0.0515)
Dataset3	SD-BGWO	0.8762	0.7726	0.9572	0.7534	0.8800
		(0.0353)	(0.0386)	(0.0295)	(0.0717)	(0.0671)
	BGWO	0.7381	0.595	1	0.585	0.8600
		(0.0558)	(0.0672)	(0)	(0.0723)	(0.1084)

9. Conclusion

This study divided the feature selection procedure into two phases utilizing an innovative algorithm. The first step was to generate a preliminary collection of characteristics using the Statistical Dependence (SD) technique. The Binary Gray Wolf Optimization (BGWO) technique followed in order to decrease and refine the subset from the first phase. To evaluate the effectiveness of these features, the K-Nearest Neighbors (KNN) classifier was employed to assess the fitness of the feature subsets. The performance of the proposed SD-BGWO algorithm was validated on three distinct datasets, showing significant improvements in classification accuracy. Comparative results with the original methodology are detailed in Tables 2 and 3.

REFERENCES

1. S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," Advances in Engineering Software, vol. 69, pp. 46–61, 2014.
2. N. Singh and S. Singh, "A modified mean gray wolf optimization approach for benchmark and biomedical problems," Evolutionary Bioinformatics, vol. 13, p. 1176934317729413, 2017.
3. M. Shahjalal et al., "A Binary Gray Wolf Optimization algorithm for deployment of Virtual Network Functions in 5G hybrid cloud," Computer Communications, vol. 193, pp. 63–74, 2022.
4. O. S. Qasim and N. M. Noori, "A new hybrid algorithm based on binary gray wolf optimization and firefly algorithm for features selection," Asian-European Journal of Mathematics, vol. 14, no. 10, p. 2150172, 2021.
5. K. Tieu, G. Dalley, and W. E. L. Grimson, "Inference of non-overlapping camera network topology by measuring statistical dependence," in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, IEEE, 2005.
6. O. S. Qasim, M. S. Mahmoud, and F. M. Hasan, "Hybrid binary dragonfly optimization algorithm with statistical dependence for feature selection," International Journal of Mathematical, Engineering and Management Sciences, vol. 5, no. 6, pp. 1420–1431, 2020.
7. Y. Li, X. Lin, and J. Liu, "An improved gray wolf optimization algorithm to solve engineering problems," Sustainability, vol. 13, no. 6, p. 3208, 2021.
8. A. Entesar, H. S. Mahmood, and O. S. Qasim, "Solving nonlinear differential equations using a hybrid method between the variational iteration method and grey wolf optimization algorithm," Journal of Interdisciplinary Mathematics, vol. 25, no. 2, pp. 573–585, 2022.
9. E. Emary et al., "Feature subset selection approach by gray-wolf optimization," in Afro-European Conference for Industrial Advancement, Springer, 2015.
10. E.-S. El-Kenawy and M. Eid, "Hybrid gray wolf and particle swarm optimization for feature selection," International Journal of Innovative Computing, Information and Control, vol. 16, no. 3, pp. 831–844, 2020.
11. O. S. Qasim and Z. Y. Algamal, "A gray wolf algorithm for feature and parameter selection of support vector classification," International Journal of Computing Science and Mathematics, vol. 13, no. 1, pp. 93–102, 2021.

12. N. Paharia, R. S. Jadon, and S. K. Gupta, "Feature selection using improved multiobjective and opposition-based competitive binary gray wolf optimizer for facial expression recognition," Journal of Electronic Imaging, vol. 31, no. 3, p. 033039, 2022.
13. K. Issac, N. Bharanidharan, and H. Rajaguru, "Feature Selection Using Elephant Herding Optimization Hybridized with Gray Wolf Optimization for Anomaly Detection in Wafer Manufacturing," in Advanced Computational Paradigms and Hybrid Intelligent Computing, Springer, 2022, pp. 515–522.
14. P. Hu, J.-S. Pan, and S.-C. Chu, "Improved binary grey wolf optimizer and its application for feature selection," Knowledge-Based Systems, vol. 195, p. 105746, 2020.
15. E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," Neurocomputing, vol. 172, pp. 371–381, 2016.
16. N. A. Al-Thanoon, O. S. Qasim, and Z. Y. Algamal, "Improving nature-inspired algorithms for feature selection," Journal of Ambient Intelligence and Humanized Computing, vol. 13, no. 6, pp. 3025–3035, 2022.
17. O. A. Alomari et al., "Gene selection for microarray data classification based on Gray Wolf Optimizer enhanced with TRIZ-inspired operators," Knowledge-Based Systems, vol. 223, p. 107034, 2021.
18. N. A. Al-Thanoon, O. S. Qasim, and Z. Y. Algamal, "Variable selection in Gamma regression model using binary gray Wolf optimization algorithm," in Journal of Physics: Conference Series, IOP Publishing, 2020.
19. M. Balistrocchi and B. Bacchi, "Modelling the statistical dependence of rainfall event variables through copula functions," Hydrology and Earth System Sciences, vol. 15, no. 6, pp. 1959–1977, 2011.
20. F. Cánovas-García et al., "Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery," Computers & Geosciences, vol. 103, pp. 1–11, 2017.