# Hybridization of Modified Grey Wolf Optimizer and Dragonfly for Feature Selection

Said Al Afghani Edsa and Khamron Sunat(✉)

College of Computing, Khon Kaen University, Khon Kaen 40000, Thailand
saidal.a@kkumail.com, skhamron@kku.ac.th

**Abstract.** There are numerous techniques designed to enhance the performance of machine learning models, with feature selection being one of the key strategies. Although many feature selection methods exist, our study presents a novel hybrid approach that merges two metaheuristic techniques: the Modified Grey Wolf Optimizer (MGWO) and the Dragonfly Algorithm (DA). This innovative method not only boosts the model's performance but also emphasizes the most pertinent features. Our experimental results showcase robust model performance, achieving an F1-score of 90% on our experimental dataset, surpassing other approaches. Further results and discussions are provided in this paper, .

**Keywords:** Modified Grey Wolf Optimizer · Dragonfly Algorithm · Support Vector Classifier · Feature · Selection · Model Performance

## 1 Introduction

There are several fundamental strategies for improving the performance of machine learning models. These strategies encompass adding more predictor features, enlarging the training dataset, adjusting or updating model parameters, enhancing feature engineering, and data preprocessing, among other techniques.

Numerous researchers have developed methods to enhance model performance [3, 8], with feature selection emerging as a leading strategy. While various feature selection techniques are available, the adoption of metaheuristic approaches has been on the rise. In light of this, our study aims to formulate an effective feature selection method that optimizes model outcomes. Our primary goal is to identify and prioritize crucial features that substantially contribute to optimal model results.

It's worth noting that previous research [11] has advocated for feature selection using the Modified Grey Wolf Optimization, especially for high-dimensional data. Yet, they encountered challenges, particularly in the evaluation of the fitness function. From their findings, it became evident that while the objective was to achieve both outstanding model performance and a concise set of selected features, the enhancement of the fitness function was lacking. To address this gap, we incorporated a rapid fitness function, designed to boost accuracy during the training phase and consequently cut down on training time. Simultaneously, we adopted a sampling technique tailored for large datasets and maintained model consistency using cross-validation throughout both the


C. Anutariya and M. M. Bonsangue (Eds.): DSAI 2023, CCIS 1942, pp. 35–42, 2023.
https://doi.org/10.1007/978-981-99-7969-1_3

feature selection and training stages. Our analysis further delves into a performance comparison, emphasizing the role of the Support Vector Machine in classification tasks.

We present a nature-inspired feature selection approach that combines the Modified Grey Wolf Optimization (MGWO) with the Dragonfly Algorithm. Our ambition with this hybrid technique is to pinpoint the most relevant features and achieve unparalleled model performance.

## 2 Research Method

This research uses a literature study and experimental approach. Other researchers such as Seyedali Mirjalili in his research [8] developed GWO hybridization for the case of feature selection with a binary approach, inspired by that we try to develop a combination of GWO and DA which is used as an indicator of feature selection, which aims to enrich the feature selection method, moreover this principle can also be used for other optimization cases because this method is derived from the meta heuristic method as well.

The process flow carried out in this experiment is given below (Fig. 1):

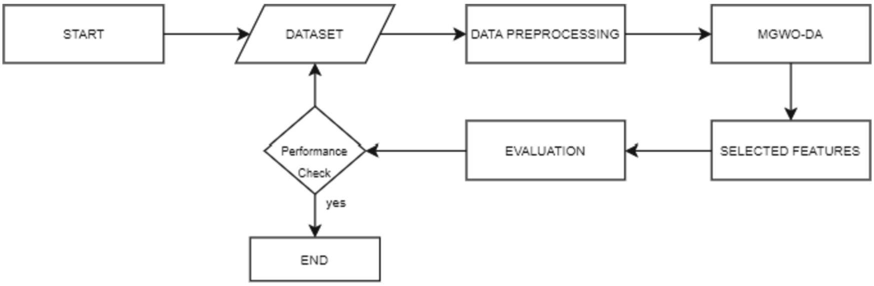


**Fig. 1.** Process flow.

**A. Grey Wolf Optimizer (GWO)**

Grey wolf Optimization (GWO) is the swarm intelligence optimization technique which was first introduced in [3]. It is inspired by the leadership hierarchy and hunting process of the grey wolf in nature. The simple mechanism of GWO makes it easy to implement over other NIAs. Also, it has fewer decision variables, less storage required, and does not possess any rigorous mathematical equations of the optimization problem. Muro [5] explained the hunting behavior of wolf into three stages as:

1. Social hierarchy: The social hierarchy of grey wolf has four levels: alpha $\alpha$, beta $\beta$, delta $\delta$ and omega $\omega$. The leaders are responsible for decision making and is denoted as the alpha wolf. The second level, called beta wolf works as a helping hand to the alpha for any activity. In the third level, the delta wolf is placed, which plays the role of scapegoat in grey-wolf packing. The rest of the wolves are categorized as omega wolf and is dominated by all other wolves.

2. Encircling the Prey: The encircling process is given by the following mathematical equation:

$$\boldsymbol{D} = \boldsymbol{C}.\boldsymbol{X}_{\boldsymbol{p},\boldsymbol{t}} - \boldsymbol{X}_{\boldsymbol{t}}, \tag{1}$$

$$\boldsymbol{X}_{\boldsymbol{t}+1} = \boldsymbol{X}_{\boldsymbol{p},\boldsymbol{t}} - \boldsymbol{A}.\boldsymbol{D} \tag{2}$$

where **A** and **C** are coefficient vectors, $\boldsymbol{X}_{\boldsymbol{p,t}}$ denotes the position vector of the prey at current iteration $\boldsymbol{t}$, and $\boldsymbol{X}_{\boldsymbol{t+1}}$ denotes the position vector of a grey wolf at next iteration. The vectors are determined as:

$$\boldsymbol{A} = 2\boldsymbol{a}.\boldsymbol{r}_1 - \boldsymbol{a}, \tag{3}$$

$$\boldsymbol{C} = 2\boldsymbol{r}_2 \tag{4}$$

with condition vector **a** is a linearly decreasing parameter from 2 to 0 and $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are random vectors in [0,1].

3. Hunting To encircle the position of prey, and the wolf position is approximated by the average of the position guided by alpha ($\boldsymbol{X}_1$), beta ($\boldsymbol{X}_2$), and gamma ($\boldsymbol{X}_3$) wolves [3]. The position of prey is estimated as:

$$\boldsymbol{X}_{\boldsymbol{t}+1} = \frac{\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3}{3}. \tag{5}$$

Many researchers have attempted modifications of the GWO, ranging from aspects of movement mutation to hunting strategy alterations by changing existing formulas, as seen in [4, 8, 11]. In this paper, we draw inspiration from [11] due to its thorough modifications, which span from the Initialization strategy, competitive strategy, multi-convergence factor, to the incorporation of differential evolution. Given the extensive modifications made by these researchers, we've endeavored to further modify it in terms of position updates by integrating the Dragonfly Algorithm.

**B. Dragonfly Algorithm (DA)**

The Dragonfly algorithm is inspired by the behavior of those who form a flock that has static and dynamic properties. Static and dynamic properties are then raised into a meta heuristic method because the meta heuristic has the main properties of exploration (static in terms of flocks) and exploitation (when it has found an optimal). In the simulation for the mathematical model, the dragonfly's behavior is divided into several parts, namely separation, alignment, cohesion, attraction for food, and distraction from enemies. In summary, the position of all dragonflies in a swarm is determined by the following equation, which $\boldsymbol{X}_{\boldsymbol{t}}$ describes the position of the dragonfly by the follwong formula:

$$\boldsymbol{X}_{\boldsymbol{t}+1} = \boldsymbol{X}_{\boldsymbol{t}} + \Delta\boldsymbol{X}_{\boldsymbol{t}+1}, \tag{6}$$

where

$$\Delta\boldsymbol{X}_{\boldsymbol{t}+1} = (sS_i + aA_i + cC_i + fF_i + w\Delta\boldsymbol{X}_{\boldsymbol{t}}). \tag{7}$$

The position vector (6) gives the position of the dragonfly. However, when there are no neighboring solutions, the dragonflies are required to fly in random search space, and their position is updated using the modified equation for the position vector (see [6]):

$$\boldsymbol{X_{t+1}} = \boldsymbol{X_t} + \boldsymbol{X_t}.levy(d). \tag{8}$$

**C. Hybridization: MGWO – DA (Proposed Method).**

Similarly to GWO - PSO, a vector a is used which has a value from 2 and 0 and can be calculated by:

$$a_{da} = 2 - l(\frac{2}{Max\_iter}), \tag{9}$$

where $l$ is the iteration factor. To combine MGWO [11] and DA we try to use the generalized position by using the levy multiplier in the case of DA with GWO which is formed in the following equation:

$$\boldsymbol{X_i^d} = (\boldsymbol{X_j^d} - \boldsymbol{A_k D_l}) * Levy + \boldsymbol{X_j^d} \tag{10}$$

with $\boldsymbol{i, k} = 1, 2, 3; \boldsymbol{j, l}$ represents for α, β and δ. The updated position of the wolf is given by:

$$\boldsymbol{X_{t+1}} = \boldsymbol{X_t} + r.(\boldsymbol{X_\alpha} - \boldsymbol{X_t}) + levy(d).\boldsymbol{X_t}, \tag{11}$$

where r comes from random number from 0 to 1.

This concept is rooted in the idea that, in the absence of prey or during the search for prey, wolves exhibit random movement, directing themselves either toward potential prey or other destinations. We aim to integrate this behavior with that of dragonflies, which also move randomly in the absence of neighboring individuals in their vicinity [10]. This combination can provide wolves with a potential advantage in exploration (Fig. 2).

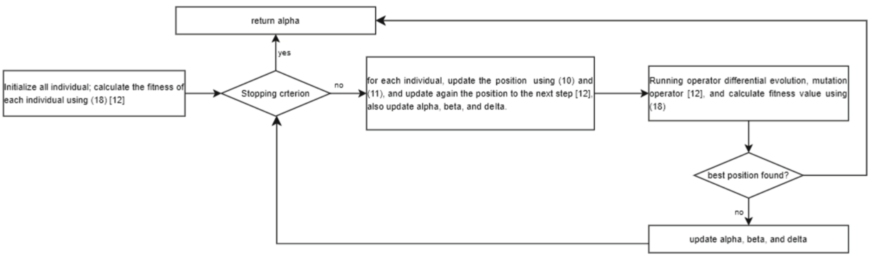


**Fig. 2.** Overview of the Procedure.

**D. Support Vector Machine: Classifier**

For details related to the support vector machine (SVM) model see [7]. The following is the general procedure related to this SVM:

Variables and parameters:

$X = \{x_1, x_2, \ldots, x_n\}$: training sample
$Y = \{y_1, y_2, \ldots, y_n\} \subset \{\pm 1\}$ : label from sample (binary case)

Kernel : type of kernel used

Par : parameter of the kernel
C : cost slak (penalty)

$\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_n]$ : lagrange multiplier
b : bias

Procedure:

1. Calculate the kernel K matrix
2. Determine constraints for quadratic programming (QP)
3. Determine objective function of (2) using:

$$\underset{\alpha}{argmax} \sum \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{12}$$

Subject to:

$$\sum_{i=1}^{l} \alpha_i y_i = 0; C_i \geq \alpha_i \geq 0,\ for\ i = 1, 2, \ldots, l \tag{13}$$

4. Find the solution of QP, and $\alpha_i$, b
5. Use the output from (4) to perform predictions:

$$y_{prediction} = sign\left(w^T + b\right) \tag{14}$$

$$y_{prediction} = sign(\sum_{i=1}^{l} \alpha_i y_i (x_i x) + b) \tag{15}$$

The hybridization of MGWO-DA is employed for feature selection in the model. To evaluate the model's performance, accuracy and F1-score metrics are utilized. Accuracy can be calculated as follows:

$$Accuracy = \frac{True\ Negatives + True\ Positive}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \tag{16}$$

For F1-score, it can be calculated as follows:

$$F1 - score = \frac{2 \times True\ Positive}{2 \times True\ Positive + False\ Positive + False\ Negative} \tag{17}$$

**E. MGWO-DA Feature Selection Process (Proposed Method).**
In the domain of feature selection, two primary principles are often emphasized: enhancing the performance of a model and minimizing the number of features used, ideally

fewer than the original number of features. Several studies in the literature, such as [10, 11], have employed these principles. In our paper, we propose a novel principle that places an emphasis on assessing the robustness and efficacy of the selected features. After this selection phase, we conduct a further evaluation of our model's performance using various metrics like accuracy, AUC, and the F1 score. For increased efficiency during evaluation, we utilize a subset of the data, say 90%. This approach substantially trims down computational time. To ensure consistent and unbiased results, we incorporate stratified cross-validation. This ensures both rapid and reliable performance estimations. Furthermode, we use the following formula [11] as the objective function of our proposed algorithm for feature selection:

$$fitness = \beta \times \frac{d}{D} + \alpha \times error\ rate \tag{18}$$

with

$$errorrate = 1 - correct\ rate, \tag{19}$$

where d as number selected feature, D as total features, α as control for performance and number of selected features, and β defined as 1 – α [11].

In our methodology, accuracy assessment is anchored on the outcomes of cross-validation. In the evaluation phase of the features we've selected, we apply stratified cross-validation to a subset of the entire dataset. This approach not only ensures a balanced representation of each class but also promotes computational efficiency (Fig. 3).

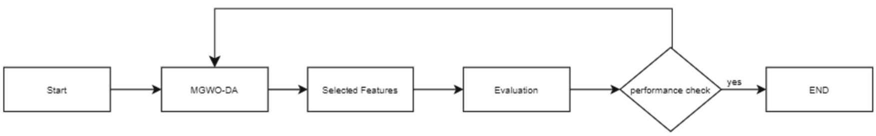


**Fig. 3.** Feature selection process.

## 3 Result and Discussion

In this experiment, we made use of the Sonar dataset, previously referenced in other research as experimental data [9]. The dataset contains 208 rows, 60 columns, and has a binary target class distribution: class 1 at 53% and class 0 at 47%. During the modeling phase, we adhered to standard procedures, partitioning the data into training and testing subsets. We assessed the model's effectiveness using the accuracy, and F1-score metrics and also analyzed its susceptibility to overfitting (Table 1).

To guarantee consistent model performance, cross-validation was employed. Although the Support Vector Machine (SVM) was our primary choice for this experiment, our selection was not exclusively tied to SVM. The main aim was to evaluate our novel feature selection algorithm across various methodologies utilizing the same

**Table 1.** Experimental Result

| Evaluation metric | Metric | MGWO-DA | GWO | MGWO | PSO | WOA | GA |
|---|---|---|---|---|---|---|---|
| Accuracy | AVG | **0.8339** | 0.7704 | 0.8148 | 0.7605 | 0.5929 | 0.7915 |
| | STD | **0.0282** | 0.0430 | 0.0050 | 0.0514 | 0.0775 | 0.0391 |
| | BEST | **0.9038** | 0.8309 | 0.8123 | 0.8451 | 0.7887 | 0.8450 |
| F1 Score | AVG | **0.8326** | 0.7915 | 0.8109 | 0.7854 | 0.7121 | 0.8082 |
| | STD | **0.0287** | 0.0402 | 0.0107 | 0.0464 | 0.0495 | 0.0448 |
| | BEST | **0.9031** | 0.8461 | 0.8420 | 0.8607 | 0.8235 | 0.8607 |

**Table 2.** Index of Features by MGWO-DA

| Metrics | Accuracy | F1-score | Index of Selected Features |
|---|---|---|---|
| AVG | **0.8339** | **0.8326** | [0, 2, 5, 8, 27, 33, 35, 39, 40, 45, 48, 53, 54, 59] |
| STD | **0.0282** | **0.0287** | |
| BEST | **0.9038** | **0.9031** | |

model. Moreover, our algorithm is compatible with other machine learning models, like the k-Nearest Neighbors (KNN) and Naive Bayes (Table 2).

Experimental results demonstrate consistent performance, indicating that this hybrid method for feature selection holds significant promise for application in modeling and other optimization problems. Its value lies not only in enhancing accuracy but also in selecting features that significantly influence predictions.

For comparative purposes, we utilized the same dataset and model (Naive Bayes) but incorporated different metaheuristics, as discussed in [12]. In that reference, the proposed metaheuristic method, Leopard Seal Optimization (LSO), achieved a peak accuracy of 97.62%. However, using the method we introduced, namely GWO-DA, the accuracy, AUC, and F1-score reached 100% (with a minimum of 100 iterations). This result is particularly significant given the unbalanced class distribution in the dataset, where accuracy alone might not sufficiently capture model performance. Our findings suggest that the GWO-DA hybridization technique offers compelling advantages in terms of accuracy enhancement.

Regarding the algorithm's complexity, it's worth noting that the required iterations (k) combined with the number of agents (in this case, wolves) (n) lead to a big-O notation of O(kn) for the method we've developed.

## 4 Conclusion

Based on the study's findings, the hybridization of MGWO-DA demonstrates competitive outcomes. Notably, configuring particle count and iteration numbers during training significantly impacts model performance. Hence, parameter selection warrants careful consideration for optimal results. Future research exploring binary optimization or BGWO-DA principles on high-dimensional datasets holds promise, enriching our understanding of feature selection and metaheuristic methods.

**Acknowledgments.** The author would like to thank the College of Computing and Khon Kaen University through the KKU Scholarship for ASEAN and GMS Countries' Personal Academic Year 2023 program.

## References

1. Yang, Q., Zhang, C., Zhang, S.: Data preparation for data mining. Appl. Artif. Intell. **17**(5–6), 375–381 (2003)
2. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 3rd edn. Morgan Kaufmann Publishers, USA (2000)
3. Mirjalili, S., Mirjalili, S.M., Lewis, A.: Grey wolf optimizer. Adv. Eng. Softw. **69**, 46–61 (2014)
4. Preeti, Deep, K.: A modified Lévy flight grey wolf optimizer feature selection approach to breast cancer dataset. In: Thakur, M., Agnihotri, S., Rajpurohit, B.S., Pant, M., Deep, K., Nagar, A.K. (eds.) Soft Computing for Problem Solving. LNNS, vol. 547. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-6525-8_31
5. Muro, C., Escobedo, R., Spector, L., Coppinger, R.P.: Wolf-pack (Canis lupus) hunting strategies emerge from simple rules in computational simulations. Behav. Proc. **88**(3), 192–197 (2011)
6. Mirjalili, S.: Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput. Appl. **27**(4), 1053–1073 (2015). https://doi.org/10.1007/s00521-015-1920-1
7. Cristianini, N., Ricci, E.: Support vector machines. In: Kao, MY. (eds.) Encyclopedia of Algorithms. Springer, Boston, MA (2008)
8. Al-Tashi, Q., Abdul Kadir, S., Rais, H., Mirjalili, S., Alhussian, H.: Binary optimization using hybrid grey wolf optimization for feature selection. IEEE Access **7**, 39496–39508 (2019)
9. Altay, O., Varol Altay, E.: A novel hybrid multilayer perceptron neural network with improved grey wolf optimizer. Neural Comput. Applic. **35**, 529–556 (2023)
10. Al-Tashi, Q., Rais, H., Jadid, S.: Feature selection method based on grey wolf optimization for coronary artery disease classification. In: Saeed, F., Gazem, N., Mohammed, F., Busalim, A. (eds.) Recent Trends in Data Science and Soft Computing. IRICT 2018. AISC, vol. 843, pp. 257–266. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-99007-1_25
11. Pan, H., Chen, S., Xiong, H.: A high-dimensional feature selection method based on modified Gray Wolf Optimization. Appl. Soft Comput. **135**, 110031 (2023)
12. H. Rabie, A., Nehal, A.M.I., Saleh, M.: Leopard seal optimization (LSO): a natural inspired meta-heuristic algorithm. Commun. Nonlinear Sci. Num. Simul. **125**(1), 107338 (2023)