

Research on Feature Optimization Algorithm for Rice Crops Under Mountainous Conditions

Yuyao Wang^{1,*}

¹ School of Surveying & Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

* Corresponding author: Yuyao Wang (Email: 212204010001@home.hpu.edu.cn)

Abstract: To improve the accuracy and efficiency of rice recognition in mountainous areas, this paper proposes a feature selection method combining the Gray Wolf Optimization (GWO) algorithm and the Random Forest classifier. First, the Relief algorithm is applied to further filter the initial feature set by calculating feature weights and setting appropriate thresholds, optimizing the feature subset's dimensionality and classification accuracy. Then, the GWO algorithm is used to further optimize the feature subset, reducing feature redundancy and correlation, and improving the quality of the final feature set. The GWO algorithm simulates the hunting behavior of gray wolf packs, dynamically adjusting the positions of the wolves to gradually approach the optimal solution. During the optimization process, the Random Forest classifier serves as the fitness function to evaluate the classification accuracy of each feature subset. Through an iterative update process, the GWO_RF algorithm successfully selects the optimal feature subset, achieving high recognition accuracy with reduced computational complexity. Experimental results show that the proposed method significantly improves recognition accuracy and effectively reduces computation time, providing a new approach for the efficient analysis of large-scale remote sensing images.

Keywords: Gray Wolf Optimization, Feature Selection, Random Forest, Relief Algorithm, Rice Recognition.

1. Introduction

Feature selection is a critical step in data mining, aiming to identify the most relevant features from a large set of raw features while eliminating redundant and irrelevant ones. This process reduces computational costs, enhances model performance, and improves data interpretability [1]. Current research on feature selection primarily falls into two categories: one evaluates the importance of each feature to the target task, selecting essential features and discarding unimportant ones; the other integrates multidimensional features through algorithms to achieve dimensionality reduction while preserving as much feature information as possible. However, the features obtained through the latter approach often lack interpretability [2-5]. This issue is particularly pronounced in land cover classification tasks, where reducing high-dimensional features via algorithmic methods hinders the understanding of their real-world significance and the characteristics they represent across different categories. Therefore, this study adopts a feature selection approach based on quantifying feature importance, identifying the optimal feature subset from the original feature set while preserving the interpretability of each feature.

In mountainous rice feature extraction, this study extracted a total of 55 spectral, textural, polarimetric, and topographic features [6-7]. Due to potential coupling among these features, a high level of redundancy may exist, making it challenging for subsequent classification models to effectively capture key features. Traditional feature selection methods often focus on a specific type of feature and fail to consider the complementarity between multiple features. This limitation is particularly problematic in complex mountainous environments, where a single feature selection strategy may overlook critical features or introduce irrelevant ones, ultimately reducing the accuracy of rice identification [8]. Therefore, a feature selection method that can fully exploit multi-source feature information and enhance the robustness

of mountainous rice identification is urgently needed.

To address this issue, this study proposes a feature selection model that integrates the Mutual Information (MI) algorithm, Relief algorithm, Grey Wolf Optimization (GWO) algorithm, and Random Forest (RF) algorithm, forming the MI-Relief-GWO_RF model, abbreviated as MRG_RF. This model is used to determine the optimal feature subset for rice identification under mountainous conditions.

2. Methodology

The MRG_RF model first employs the Mutual Information (MI) algorithm to identify a subset of features with weak relevance to the target task. Then, the Relief algorithm ranks the initially selected features based on their importance weights and filters out features exceeding a predefined threshold. The remaining features are retained for the next stage of feature selection. Subsequently, the Grey Wolf Optimization (GWO) algorithm is combined with the Random Forest (RF) algorithm to form a wrapper-based feature selection method (GWO_RF). In this process, the GWO algorithm searches for the optimal feature subset within the feature space, while the RF algorithm serves as the fitness function to evaluate the performance of the feature subsets. Finally, the GWO_RF method is applied to refine the preselected feature set, resulting in the optimal feature subset.

2.1. Random Forest

2.1.1. Decision Tree

A decision tree is a recursive partitioning model used for classification and regression tasks. Its core idea is to iteratively select the most informative features to split the dataset into increasingly homogeneous subsets, ultimately improving prediction accuracy. At each split, the algorithm determines the optimal feature and threshold using selection criteria such as information gain, gain ratio, or the Gini index. The decision tree construction consists of two main phases: tree growth and pruning. The tree growth phase recursively

partitions the data until all samples are classified or further division is not possible. The pruning phase prevents overfitting by simplifying the tree structure to enhance generalization performance [9].

(1) Tree Growth

The tree growth process recursively partitions the dataset until either all samples belong to a single category or a predefined stopping criterion is met. The steps include:

1. Selecting the optimal feature for splitting, based on information gain, gain ratio, or Gini index, to maximize subset purity.

2. Partitioning the dataset into subsets according to the chosen feature values and generating child nodes.

3. Recursively repeating the feature selection and splitting process for each child node until a stopping condition is reached, such as a homogeneous class distribution or a minimum number of samples per node.

(2) Tree Pruning

To prevent overfitting, pruning simplifies the tree while maintaining predictive accuracy. The steps include:

1. Setting a threshold to determine pruning, such as a change in loss function or validation performance.

2. Starting from leaf nodes and evaluating whether pruning a node improves overall model performance. If merging the node and its children reduces loss or improves performance, pruning is applied.

3. Repeating the process until no further pruning improves performance, resulting in a more generalizable tree.

2.1.2. Random Forest

Random Forest is an ensemble learning algorithm composed of multiple decision trees. By introducing randomness in both data sampling and feature selection, it reduces overfitting and variance, enhancing model robustness. The key idea is to train multiple trees on different data subsets and aggregate their predictions through majority voting (classification) or averaging (regression) [10-11].

2.1 Construction of Random Forest

The random forest algorithm consists of the following steps:

1. Generating multiple bootstrap samples from the original dataset using sampling with replacement. Each sample is used to train an individual decision tree.

2. For each decision tree, randomly selecting a subset of features instead of considering all features, reducing feature dominance and ensuring diversity.

3. Constructing decision trees independently using the selected samples and features.

4. Aggregating predictions from all trees via majority voting for classification tasks or averaging for regression tasks.

Random Forest is effective in handling high-dimensional data, mitigating noise sensitivity, and dealing with missing values. However, it has drawbacks such as increased computational complexity and reduced interpretability. The number of trees is a crucial hyperparameter affecting model performance. Too few trees lead to instability, while too many trees increase computational cost.

To determine the optimal number of trees, experiments were conducted by testing tree counts from 100 to 1000 (step size = 100) and evaluating classification accuracy. As shown

in Figure 1, the feature groups tested include spectral (Spectral), texture (Texture), terrain (Terrain), polarization (Polarization), and their combinations (ST, STt, STtP). Results indicate that increasing tree count reduces error, but computational efficiency decreases. Accuracy stabilizes at 600 trees, making it the optimal choice balancing performance and efficiency.

This study employs Random Forest for feature selection due to its ability to evaluate all features while mitigating overfitting risks and enhancing robustness. Its capability to handle high-dimensional data ensures reliable feature selection outcomes.

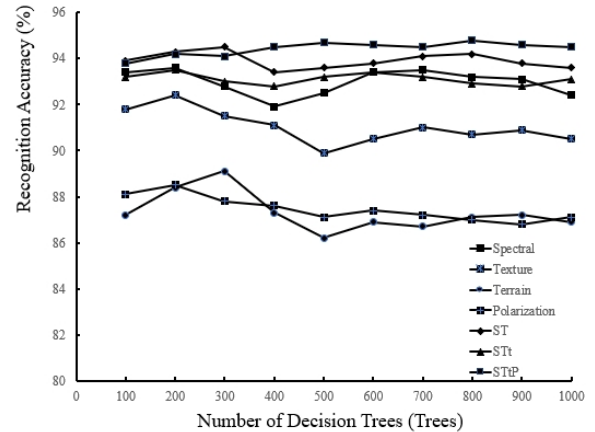


Figure 1. Impact of Number of Decision Trees on Recognition Accuracy

In the feature selection process, we applied both single-feature algorithms and fused-feature algorithms to extract features from mountainous rice samples, using random forests as the classifier. To reduce errors introduced by random sampling of samples, each algorithm's computation was performed 100 times, and the average value was taken to achieve more stable rice recognition results. The final computational results are shown in Table 2.

From the analysis in Table 2, it can be seen that the recognition rate of feature combinations is significantly better than that of single-feature algorithms, and its recognition accuracy remains stable above 90%. This result suggests that, in the task of rice extraction in mountainous areas, feature combinations can enhance recognition performance and stability compared to single features. Therefore, to further optimize the feature selection process, this paper will use the feature set obtained from the STtP feature combination as the initial feature set for subsequent feature selection.

Table 1. Result Statistical Analysis

Feature Set	Number of Features	Recognition Accuracy (%)	Time (min)
Spectral	26	92.5	1286.3
Texture	21	91.1	1109.9
Terrain	4	84.3	201.5
Polarization	4	85.5	231.3
ST	47	93.2	1327.6
STt	51	92.8	1335.1
STtP	55	93.1	1341.9

2.1.3. Mutual Information Algorithm

The Mutual Information (MI) algorithm is a feature

selection method based on information theory, used to measure the degree of correlation between two variables. It is commonly applied in fields such as feature selection and data analysis [12]. The principle of MI is to calculate the deviation between the joint distribution and the individual marginal distributions of two random variables, quantifying their correlation. For two variables X and Y, the mutual information value can be calculated using the following formula (1):

$$MI(X;Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \quad (1)$$

Where $p(x,y)$ represents the joint probability distribution of X and Y; $p(x)$ and $p(y)$ represent the marginal probability distributions of X and Y, respectively; and $MI(X;Y)$ represents the mutual information value between variables X and Y.

The larger the MI value, the stronger the correlation between the two variables. When $MI(X;Y)=0$, it indicates that the two variables are completely independent.

The initial feature set in this study includes 55 features. Since combining multiple features may result in highly redundant features, affecting the efficiency of subsequent feature selection and model accuracy, the mutual information algorithm is used for preliminary feature screening. The MI values are divided into three correlation ranges:

When the MI value is greater than 0.5, it indicates a high correlation (S) between the feature and the target variable; when the MI value is between 0.2 and 0.5, it represents a moderate correlation (W); and when the MI value is less than 0.2, it indicates low or no correlation (N) with the target variable.

The calculation results are shown in the table below.

Table 2. Correlation Analysis of Feature Set

Feature Set	Number of Features	Recognition Accuracy (%)	Time (min)
Original Feature	55	93.1	1341.9
MI (W+N)	39	93.5	608.1
MI (N)	6	68.7	92.4

From Table 2, it can be seen that after applying the mutual information algorithm, highly correlated features were removed, reducing the number of features from 55 to 36. Compared to the original feature set, the recognition accuracy improved by 0.4%, and the training time was reduced to 45.3% of the original time. When only low-correlation features were selected, the feature dimension was reduced to 6, resulting in a recognition accuracy that was too low to meet practical requirements. The subsequent feature selection should use a feature set that includes both moderate and low correlation features as the initial feature set. The MI(W+N) feature set effectively improves recognition accuracy and efficiency. Therefore, this feature set is selected as the initial feature set for subsequent feature selection.

2.1.4. Relief Algorithm

From the above initial feature set, it can be observed that although the feature dimension has been reduced, the feature set still possesses high dimensionality and large computational complexity. If directly applied to subsequent

optimization algorithms, it may lead to excessively long computation times and affect the recognition performance of the final optimal feature subset. Therefore, secondary feature selection on the initially screened feature set is crucial. This not only reduces the computation and time complexity but also further enhances the accuracy and stability of the algorithm [13-15].

The Relief algorithm is a feature selection method designed for binary classification problems. By introducing a weight evaluation mechanism based on sample distances, it not only considers the relationship between each sample and other samples within the same class but also incorporates the differences with samples from other classes, thus providing a more comprehensive measure of feature importance. This makes Relief a widely used and classic algorithm in the field of feature selection and is regarded as one of the most successful algorithms in the preprocessing phase of feature selection.

In this paper, the Relief algorithm is applied to the secondary screening of the mountainous rice identification feature set. The MI (W+N) feature set is used as input to calculate the weight value of each feature. A weight threshold is set to complete the secondary feature selection. The specific steps are as follows: randomly select a sample S, find k nearest neighbor samples NH within the same class of S, and simultaneously find k nearest neighbor samples NM from different classes. Then, using an iterative update method, the feature weight $\omega(X)$ is calculated, as shown in equations (2).

$$\text{diff}(X,S,S')=\begin{cases} \frac{|S[X]-S'[X]|}{\max(X)-\min(X)} & X \text{ is continuous} \\ 0 & X \text{ is discrete, and } S[X]=S'[X] \\ 1 & X \text{ is discrete, and } S[X]\neq S'[X] \end{cases} \quad (2)$$

In the equations, X represents a feature in the feature set; S and S' represent two randomly selected samples; k represents the number of nearest neighbors; m represents the number of iterations; C represents the class label; P(C) represents the probability that the feature belongs to class C; S[X] represents the feature value of sample S; NH_j represents the j-th nearest neighbor sample in class C; NM(C)_j represents the j-th nearest neighbor sample in the non-C class; max(X) and min(X) represent the maximum and minimum values of feature X, respectively; Class(S) represents the class of sample S; and $\text{diff}(X, S, S')$ represents the difference in feature X between sample S and S'.

The Relief algorithm is essentially a method for calculating feature weights and does not directly perform feature selection[16]. Therefore, a threshold for feature weights can be set manually to select high-weight features, forming the optimal feature subset. By setting an appropriate threshold, not only can the feature set's dimensionality be optimized, but it can also ensure that the selected features contribute significantly to the recognition of typical regions.

In this paper, the feature weights calculated by the Relief algorithm are first normalized into percentages to facilitate a more intuitive analysis of the weight distribution range. Different threshold ranges are set based on the feature weight distribution, and features with lower weights are discarded. The performance of the feature subsets corresponding to different weight thresholds is then evaluated using the classifier's recognition accuracy. Specifically, the feature weights are divided into five ranges: 0.005, 0.010, 0.015, and 0.020. Based on the random forest classifier, the recognition

accuracy and number of features are calculated for each weight range to select the optimal feature threshold range.

From the recognition rate results corresponding to the different weight ranges in Table(3), it can be observed that when the weight threshold is 0.010, the recognition accuracy reaches the optimal value, and the computational efficiency is relatively high. This indicates that within this weight threshold range, the feature set can effectively represent the features. Further analysis reveals that when the weight threshold increases to 0.015 or 0.020, although the number of features decreases significantly, the recognition accuracy declines to varying degrees. Therefore, after considering both recognition accuracy and computational efficiency, the feature set with a weight threshold of 0.010 is selected as the initial feature set for the subsequent optimal feature subset, with a feature dimension of 28. This choice not only ensures a high recognition rate but also significantly reduces computational complexity, helping to improve the efficiency and accuracy of subsequent optimization algorithms.

Table 3. The Impact of Different Weight Thresholds on Recognition Accuracy

Weight	Number of Features	Recognition Accuracy (%)	Time (minutes)
0.005	33	88.1	591.8
0.010	28	91.6	473.5
0.015	21	86.3	681.1
0.020	12	82.1	737.9

2.1.5. Grey Wolf Optimization Algorithm

After the feature subset is initially filtered through the Relief algorithm, some features still exhibit high redundancy and correlation. Therefore, the Grey Wolf Optimization (GWO) algorithm is used for further selection and optimization of the feature subset [17].

The Grey Wolf Optimization (GWO) algorithm is a metaheuristic optimization algorithm based on the hunting behavior of grey wolves in nature. The algorithm simulates the social hierarchy and hunting strategy of grey wolves, dividing them into four levels: α wolves, β wolves, δ wolves, and ω wolves, each representing different roles in the population. The α wolf is the leader of the group, responsible for providing decision-making direction; the β wolf is the secondary leader, assisting the α wolf in decision-making; the δ wolf maintains the group's order; the ω wolf is a common member, responsible for executing tasks. In the GWO algorithm, the population's search process is mainly divided into three stages: surrounding, hunting, and attacking, dynamically adjusting the search range and direction to gradually approach the global optimal solution [18].

The core idea of the GWO algorithm is to simulate the hunting behavior of grey wolf groups to find the optimal solution. In the solution space, each grey wolf represents a candidate solution, and its position represents a feature subset. By dynamically adjusting the grey wolves' positions, the algorithm gradually approaches the optimal feature subset. The algorithm updates the positions of the α , β , and δ wolves as reference points, guiding other grey wolves to update their positions, causing the population to converge toward the optimal solution. GWO has the advantages of being simple to implement, fast convergence speed, and strong global search capability. It has been widely applied in feature selection, image processing, path planning, and machine learning.

In the GWO algorithm, the search behavior of the population is achieved by surrounding and attacking prey. The position update formula is based on the prey's position and the average distance of other wolves in the group, iteratively adjusting positions to achieve efficient search in the solution space. Although GWO has strong global search ability, its performance depends on the specific problem and requires parameter tuning based on the actual scenario. For example, controlling the dynamic weight of the search factors can effectively balance global and local search capabilities, improving the algorithm's convergence speed and optimization effect.

In this paper, a combination of the Random Forest algorithm and GWO (GWO_RF) is proposed, using the Random Forest classifier as the fitness function for the GWO algorithm to evaluate the feature subsets. Specifically, the GWO algorithm selects a feature subset during each iteration, inputs it into the Random Forest classifier for classification, and uses classification accuracy as the fitness function's objective value. Based on the fitness values, the GWO algorithm adjusts the position of each grey wolf, optimizing the feature subset selection process. The GWO_RF algorithm not only improves feature selection efficiency but also reduces the classifier's computational complexity, ultimately obtaining the optimal feature subset.

The flowchart of the GWO_RF algorithm is shown in Figure 2:

(1) Initialize the population: Randomly generate a certain number of grey wolves, where each grey wolf's position represents a feature subset. Each element in the position vector corresponds to a feature, with values of 0 or 1 indicating whether the feature is selected. The population size, maximum iteration count, and relevant parameters (such as a , A , C , etc.) are also set.

(2) Calculate fitness: Use the Random Forest algorithm to evaluate the classification accuracy of each grey wolf's corresponding feature subset, treating the classification accuracy as the grey wolf's fitness value. A higher fitness value indicates stronger classification capability of the feature subset.

(3) Select α , β , and δ wolves: Sort the population based on the fitness values of the grey wolves, selecting the three wolves with the highest fitness values as the α (optimal solution), β (second-best solution), and δ (third-best solution) wolves. The remaining wolves are considered ordinary members.

(4) Update the grey wolf positions: The positions of ordinary grey wolves are updated based on the positions of the α , β , and δ wolves, as well as the dynamically adjusted coefficients a and A , using the following formula:

$$D_\alpha = |C_1 \cdot \vec{X}_\alpha - \vec{X}| \quad (3)$$

$$D_\beta = |C_2 \cdot \vec{X}_\beta - \vec{X}| \quad (4)$$

$$D_\delta = |C_3 \cdot \vec{X}_\delta - \vec{X}| \quad (5)$$

$$\vec{X}(t+1) = \frac{\vec{X}_\alpha + \vec{X}_\beta + \vec{X}_\delta}{3} \quad (6)$$

The position vectors \vec{X}_α and \vec{X}_β and \vec{X}_δ represent the positions of the α , β , and δ wolves, respectively; \vec{X} denotes the current position of the grey wolf. The coefficients

C_1, C_2 and C_3 are random weight factors, while D_α, D_β and D_δ represent the distances between the grey wolf and the $\alpha, \beta,$ and δ wolves.

(5) Iterative update: Steps (2) to (4) are repeated. In each iteration, the positions of the $\alpha, \beta,$ and δ wolves, along with their corresponding fitness values, are updated. The positions of the grey wolves in the population are gradually optimized, and the search for the global optimal feature subset continues.

(6) Termination condition: The algorithm terminates when the maximum number of iterations is reached, or when the fitness value of the α wolf no longer shows significant changes.

(7) Output result: The feature subset corresponding to the α wolf is output as the optimal feature subset. The GWO_RF algorithm effectively reduces the feature dimensionality while improving classification accuracy, ultimately yielding the optimal feature subset.

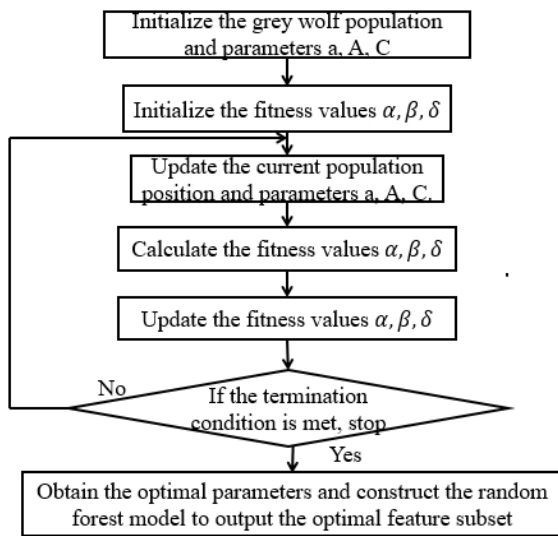


Figure 2. The process of the GWO_RF

2.2. Comparison Experiment

To verify the effectiveness of the proposed MRG_RF feature selection model, this paper compares it with classical feature selection methods, including the commonly used Information Gain (IG) and Genetic Algorithm (GA), to comprehensively evaluate the performance of the three methods in terms of accuracy improvement and computational efficiency.

Table 4. Results of Different Feature Selection Methods

Method	Number of Features	Recognition Accuracy (%)	Time (Minutes)
Proposed Method	19	92.3	426.0
GA	25	87.7	496.9
IG	31	89.1	523.7

The MRG_RF model proposed in this paper outperforms traditional feature selection methods. In terms of feature count, the proposed method selects only 19 efficient features, ensuring low redundancy while simplifying the feature set. Regarding accuracy, the multi-optimization strategy used in this method better captures the key nonlinear relationships and potential patterns in the data, resulting in the most precise

feature combination. In terms of efficiency, the proposed method achieves higher accuracy with a lower time cost.

3. Conclusion

This study presents a novel feature selection algorithm for rice crop recognition in mountainous regions, which integrates the Relief algorithm and the Grey Wolf Optimizer (GWO) to effectively reduce feature redundancy and enhance the classification accuracy. The proposed method first utilizes the Relief algorithm to identify important features from an initial feature set, followed by further optimization using the GWO to refine the feature subset. The results show that the combined use of Relief and GWO significantly improves both feature selection efficiency and classification performance, as demonstrated by the increased recognition accuracy and reduced computation time.

By leveraging these algorithms, the study successfully identifies an optimal feature subset that balances high classification accuracy with low computational complexity. This research demonstrates the potential of machine learning-based feature selection techniques in precision agriculture, particularly in the challenging environment of mountainous regions. Future work will explore the application of this method to larger-scale agricultural monitoring tasks and further refinement of the feature selection process for even more complex agricultural scenarios.

References

- [1] L. Zhang, H. Li, and X. Xu, "A novel feature selection method for remote sensing image classification based on mutual information and relief algorithm," *Remote Sensing*, vol. 12, no. 5, pp. 1012–1025, May 2020.
- [2] H. Zhang, Z. Chen, and M. Zhang, "A hybrid feature selection method for hyperspectral image classification using relief and genetic algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2790–2801, Apr. 2020.
- [3] Y. Wang, L. Zhang, and X. Li, "Gray wolf optimization algorithm based feature selection for remote sensing image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 177–189, Feb. 2020.
- [4] Z. Li, L. Zhang, and X. Chen, "Feature selection for high-dimensional remote sensing image classification using adaptive genetic algorithm and random forests," *Remote Sensing*, vol. 13, no. 3, pp. 459–471, Mar. 2021.
- [5] R. Liu, Z. Huang, and X. Chen, "A hybrid feature selection method based on relief and particle swarm optimization for remote sensing applications," *IEEE Access*, vol. 8, pp. 134285–134294, 2020.
- [6] M. Zhang, X. Wang, and J. Yang, "Efficient feature selection method based on gray wolf optimizer for land cover classification in complex terrain," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 90, pp. 102115, Dec. 2020.
- [7] J. Liu, Z. Han, and Y. Ma, "A novel feature selection method based on improved relief and gray wolf optimization algorithm for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 7, pp. 4319–4329, Jul. 2022.
- [8] Y. Xu, S. Zhang, and Q. Lu, "Feature selection and classification of remote sensing images based on hybrid GWO and SVM," *Remote Sensing*, vol. 13, no. 18, pp. 3548–3560, Sept. 2021.

- [9] X. Wang, J. Yang, and Z. Zhang, "Remote sensing image classification using a feature selection method based on gray wolf optimization algorithm and decision tree," *Comput. Geosci.*, vol. 149, pp. 104670, Oct. 2021.
- [10] Y. Tang, W. Xu, and J. Chen, "A new hybrid feature selection approach for remote sensing image classification: relief and mutual information-based grey wolf optimization," *J. Appl. Remote Sens.*, vol. 14, no. 2, pp. 026501, Apr. 2020.
- [11] L. Liu, S. Xie, and W. Zheng, "Remote sensing image classification using hybrid feature selection and machine learning methods," *Sensors*, vol. 21, no. 3, pp. 798–812, Mar. 2021.
- [12] Z. Wang, H. Zhang, and P. Wang, "Feature selection for remote sensing image classification based on hybrid algorithms of gray wolf optimization and random forests," *J. Sensors*, vol. 2021, pp. 183458, Feb. 2021.
- [13] J. Yang, W. Xie, and Y. Chen, "A novel feature selection method for hyperspectral remote sensing image classification based on hybrid genetic algorithm and gray wolf optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9060–9074, Nov. 2021.
- [14] H. Chen, Y. Wu, and M. Jiang, "Feature selection for high-dimensional remote sensing data using a hybrid method of the genetic algorithm and simulated annealing," *Comput. Geosci.*, vol. 130, pp. 38–47, Feb. 2020.
- [15] W. Liu, X. Zhang, and H. Yang, "A novel hybrid feature selection method based on reliefF and whale optimization algorithm for remote sensing image classification," *Remote Sensing*, vol. 12, no. 22, pp. 3745–3758, Nov. 2020.
- [16] X. Xie, W. Li, and J. Li, "Hybrid feature selection algorithm based on gray wolf optimization and principal component analysis for hyperspectral remote sensing image classification," *Computers, Environment and Urban Systems*, vol. 87, pp. 101610, Mar. 2021.
- [17] Q. Zhang, X. Zhao, and L. Zhang, "Improving remote sensing image classification through feature selection and dimensionality reduction methods," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. VI-3, pp. 37–46, Sept. 2021.
- [18] M. Liu, Z. Li, and Y. Yu, "A new approach for feature selection in remote sensing image classification using hybrid feature selection algorithms," *J. Earth Sci., Vol. 2020*, pp. 320–329, Jun. 2020.