IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 July 2023, accepted 8 August 2023, date of publication 21 August 2023, date of current version 25 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3307196

RESEARCH ARTICLE

Hyper-Spectral Image Pixel Classification Based on Golden Sine and Chaotic Spotted Hyena Optimization Algorithm

XIPING YANG AND LIFANG CHENG
School of Mathematics, Zhengzhou University of Aeronautics, Zhengzhou 450046, China

Corresponding author: Xiping Yang (yangmath@163.com)

This work was supported in part by the National Natural Scientific Foundation of China "Study on the Predictability of the Large Curve Path of Kuroshio" under Grant 41906003; in part by the Key Scientific and Technological Project of the Higher Education Institutions of Henan Province, China, "Research on Real-Time Early Warning Mechanism of Intelligent Video Surveillance Under Big Data," under Grant 22A520012; and in part by the Key Scientific and Technological Project of the Higher Education Institutions of Henan Province, China, "Research on Intelligent Synthesis of Multimedia Graphics and Texts Based on Aesthetic Perception," under Grant 23A120004.

ABSTRACT Hyperspectral remote sensing technology is a breakthrough technology that integrates imaging and spectral methods, but it also faces issues such as classification accuracy being controlled by dimensionality and spectral variability. Based on this, this study proposes a Golden Sine Spotted Hyena Optimization (GSSHO) algorithm that integrates the Golden Sine method, chaos strategy, and tournament strategy to help hyperspectral remote sensing technology achieve higher work efficiency. The implementation of the algorithm can be roughly divided into three stages. This includes initializing using chaotic strategies, calculating the individual fitness value and selecting the best cluster, and after updating the fitness value, continuing to select the optimal solution until the stop condition of the algorithm is met. Among them, the golden sine algorithm provides a new data update method for the spotted hyena optimization algorithm, avoiding situations where individual data cannot be searched for. The Salinas and Pavian Centre datasets were used in the experiment to validate the effectiveness of the improved spotted hyena optimization algorithm. At the same time, four other common algorithms were selected for performance comparison experiments with the algorithm. The experimental results show that the GSSHO algorithm has excellent dimensionality reduction and convergence capabilities. Compared with the total number of original bands, in both datasets, the two values decreased to the total of 32 and 11, respectively, equivalent to 1/6 and 1/9 of the original data; The fitness function values are 0.0836 and 0.0315 respectively. And compared with the other four algorithms, this algorithm also significantly outperforms other algorithms in all aspects of indicators. Therefore, the spotted hyena optimization algorithm based on the golden sine algorithm and chaotic strategy can make band selection more efficient.

INDEX TERMS Spotted hyena algorithm, chaos strategy, golden sine, optimization problem, pixel classification.

I. INTRODUCTION

Remote sensing is a means of using sensors to extract environmental features, and realizing the absorption of electromagnetic wave length based on spectroscopy, which is called spectral remote sensing technology. This technology can help humans analyze and recognize observed objects, and obtain the desired information. Compared with traditional imaging techniques, hyperspectral remote sensing technology has many advantages, such as higher resolution. Because some subtle differences in the spectrum cannot be recognized by traditional remote sensing, resulting in problems such as classification errors. And hyperspectral technology can effectively avoid this situation and lay a solid foundation for subsequent data detection. At the same time, this technology can also combine multiple types of features of the data,

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

improving the accuracy and stability of detection. However, these advantages also bring corresponding defects. As the feature dimension of data increases, the classification accuracy will gradually decrease, that is, the curse of dimensionality problem; The spectrum itself is also susceptible to factors such as noise and heterogeneity [1]. Therefore, dimensionality reduction of hyperspectral data is a very important step, which can not only reduce computational burden but also achieve high resolution in high-dimensional data. There are roughly two approaches to dimensionality reduction for hyperspectral data, namely feature extraction and band selection. The first method is to use mapping to transform high-dimensional data into low-dimensional data, while the second method is to refine appropriate bands from the original data to form a subset. The method of band selection can preserve the original features to a greater extent without damaging their physical significance, therefore, this method has also been more widely applied [2]. The selection of hyperspectral bands is actually an optimization problem, which involves using corresponding evaluation rules to select the optimal subset of bands. The problem with this condition is usually solved using more efficient meta heuristic optimization algorithms, including particle swarm optimization and genetic algorithms. But these algorithms all have some unresolved shortcomings, such as the insufficient convergence speed and search ability of genetic algorithms; The particle swarm optimization algorithm has weak optimization ability and is prone to falling into local optimization problems. So, the optimization algorithm still needs further improvement and innovation [3]. The spotted hyena optimization algorithm has high stability, but it also has problems such as local optimization. Therefore, the study introduced the golden sine method, chaotic strategy, and tournament strategy to improve it. The golden sine algorithm can improve the global search ability of the algorithm, and the combination of the two strategies improves the local optimization and initial cluster problem of the original algorithm. This provides a data foundation for subsequent algorithm work. The innovation of this study is to reduce the dimensionality of hyperspectral data through feature extraction and band selection, while selecting hyperspectral bands; The chaotic Spotted hyena optimization algorithm is combined with the golden sine algorithm to improve the local optimization and initial clustering problem. The study combines two more advanced algorithms to solve the local optimization problem and initial clustering problem of hyperspectral images, which is a bold and novel method. In the pixel classification of hyperspectral images, a large amount of parameter information needs to be adjusted, and previous research has had limited effectiveness in this area. Therefore, this research work successfully solved the problem of parameter optimization.

II. RELATED WORKS

Remote sensing technology is widely used in natural disaster prediction, Geological survey, agriculture and other fields. Among them, hyperspectral remote sensing technology with high resolution is a popular research direction. Hyperspectral remote sensing technology is a remote sensing technology used to obtain and analyze the details of the earth surface. Compared with traditional remote sensing technology, hyperspectral remote sensing technology can obtain more and narrower spectral band data. Traditional remote sensing technology usually only uses a few bands, while hyperspectral remote sensing technology can obtain data from dozens or even hundreds of bands. Peyghambari S et al. applied remote sensing technology to geological detection, introduced the basic process of information extraction, and verified the effective application of HiR-HRST in different cases through experiments [4]. Qin et al. believed that reasonable extraction of water bodies is the foundation for ensuring water resources. They have studied the combination of HiR-HRST and U-Net neural networks for feature extraction of small water bodies. Scholars have taken measures such as strengthening network depth to further improve it, and the reliability of this method has been verified through regression experiments [5]. Liew et al. proposed an algorithm for measuring water parameters, which combines HiR-HRST for image classification to achieve object recognition function. The effectiveness of this method was verified in subsequent experiments. Although HiR-HRST is widely used, there are still some unresolved issues [6]. Li et al. believed that environmental factors such as atmosphere and location, as well as the characteristics of sensors themselves, can affect the accuracy of hyperspectral remote sensing images. Therefore, scholars have proposed an HSI improvement technique to achieve noise elimination and improve defogging, and verified the effectiveness of this strategy through simulation experiments. In addition to this technology, optimizing algorithms is also one of the ways to improve hyperspectral remote sensing [7]. Yang et al. believed that the high correlation of bands can increase computational pressure and limit the accuracy of hyperspectral imaging. Thus, refining bands was necessary. They used the K-means algorithm to achieve this goal. Due to its high sensitivity to the initial cluster centers, randomly selecting the initial cluster centers may cause a decrease in algorithm accuracy. Therefore, the study introduced the Mahmin distance algorithm to select the initial clustering centers, then cluster the bands, and output the bands with the maximum variance [8]. Zhang et al. chose a common spatial pattern (CSP) algorithm that can achieve filtering and feature extraction, but this algorithm has a high dependence on time and frequency bands, making it difficult to implement. Thus, quantum particle swarm optimization algorithm was introduced to improve algorithm accuracy, and support vector machines were subsequently used for information classification. The simulation results of the improved algorithm and the control algorithm using known datasets showed that the improved algorithm has significantly improved classification accuracy [9]. Rato and Reis realized that spectral data has a wide range of applications in the industrial field, and due to the large amount of spectral data, only a small portion of information is actually useful. The efficiency

of the algorithm is closely related to the ability to select wavebands. In previous studies, a selection method using interval partial least squares has been proposed. However, this method has significant limitations, so the author chose a multi resolution algorithm for improvement. This algorithm ensures optimal data resolution while selecting bands, and its performance is significantly better than interval partial least squares method [10]. Luo et al. proposed a Feature selection method based on the maximum information coefficient of conditional refinement to improve the consistency of correlation measures and the accuracy of redundant estimation under limited samples. The test of the final prediction model of the actual Fluid catalytic cracking (FCC) process proved the extensive correlation between variables and objectives. Only a few variables are essential for modeling, while the rest are redundant. Compared with other methods, CRMICM has the best dimensionality reduction effect on FCC process data in terms of feature number and model prediction accuracy, demonstrating its good applicability to chemical processes [11]. Bharathi et al. proposed a hybrid Feature selection algorithm based on ant colony optimization (ACO) and fast branch and bound (QBB). Ant colony optimization algorithms realizes Feature selection by observing real ants looking for food resources. First, use the QBB algorithm to divide the Big data set into two parts. The combination of these two algorithms realizes data dimensionality reduction based on Feature selection. According to a set of performance indicators, such as accuracy and recall, the algorithm is simulated and compared with existing Feature selection algorithms. The F-score, classification accuracy, and size of selected features demonstrate the effectiveness of the proposed hybrid algorithm [12]. Parlak et al. believe that it is very necessary to apply feature selection method as a dimensionality reduction step in the field of text classification, so a new FS method is proposed. It benefits from corpus-based and class-based probabilities in computation. Comparative analysis was performed on four benchmark datasets, and the experimental results showed that the EFS method outperformed the other nine methods in most cases, based on micro and macro F1 scores [13]. Yu et al. proposed an efficient method for reducing dimension of trajectory and a method for setting initial values of DBSCAN hyperparameters. Trajectory dimensionality reduction algorithm processes trajectories of different lengths into the same dimension (same number of feature points). Retention points retain spatial and temporal information about these trajectories as much as possible. The DBSCAN hyperparameter initialization algorithm obtains the effective initial values of eps and MinPts, which is convenient for subsequent adjustment. Finally, they verify these methods on two real scene trajectory data sets, and the experimental results are satisfactory and effective [14]. In order to efficiently cluster and reduce dimension of data, Guo et al. proposed a prior dependence graph (PDG) construction method to model and discover complex relationships of data. The developed PDG model is then applied to two typical data analysis tasks, namely unsupervised data clustering and dimensionality reduction. Experimental results on several benchmark databases show that compared with some existing graph learning models, PDG model can achieve remarkable performance, and can be deployed in edge computing module to provide efficient solutions for massive data management and application in AIoT [15]. Xiang et al. proposed a new variable neighborhood search golden sine cosine SALP group algorithm, which is also proposed as a new shape matching optimization method. As a relatively new branch, the atomic potential matching model is inspired by potential field gravity. Compared with traditional edge potential function (EPF) models, APM has been proven to be less sensitive to complex backgrounds in test images and more cost-effective in the calculation process. The experimental results of four real examples show that the GSCSSA-VNS algorithm can provide very competitive results, outperforming other algorithms [16]. Luo et al. designed a spotted hyena optimizer to train the Feedforward neural network. Their goal is to apply meta heuristic optimization algorithm to solve this problem better than mathematical and deterministic methods. To demonstrate that using SHO to train FNN is more effective, the performance of the proposed method was tested using five classification datasets and three function approximations. The experimental results show that the proposed SHO algorithm for optimizing FNN has the best comprehensive performance, and has more outstanding performance than other advanced meta heuristic algorithms in terms of performance indicators [17]. Baset et al. proposed a new algorithm that incorporates the characteristics of the sine cosine algorithm and the Simpson method (SCA-SM). This method consists of two stages: in the first stage, the sine cosine algorithm is used to find the optimal segmentation point on the integration interval of the integrand. In the second stage, the approximate integral value of the integrand is calculated using the Simpson method. The numerical simulation results show that the algorithm provides an effective way for the numerical calculation of definite integral, with fast Rate of convergence, high precision and robustness [18].

To sum up, the pixel classification of hyperspectral images will face difficulties such as high-dimensional data, noise and clutter information, data imbalance, Feature selection and extraction, model complexity and computation. At the same time, there are also studies on related problems, and some results have been achieved. For example, the classification accuracy of hyperspectral images has been improved and the classification time has been optimized. However, there is still much room for improvement in parameter optimization, operation efficiency and computational complexity. In order to solve the parameters, this study combined gold sine method with chaotic spotted hyena optimization algorithm, which not only enhanced the global search ability of the algorithm, but also solved the local optimization problem of the initial algorithm, and fully reduced the dimension of hyperspectral image data to achieve the maximum efficiency of the algorithm.

III. A PIXEL CLASSIFICATION METHOD FOR HYPERSPECTRAL IMAGES BASED ON GSSHO

A. INTRODUCTION TO OPTIMIZATION ALGORITHM AND HYPERSPECTRAL TECHNOLOGY FOR SPOTTED HYENAS

The spotted hyena is a large social animal that typically inhabits its tropical or desert grasslands in Asia and Africa. During hunting, it uses its calls and movements to transmit signals to its companions, forming a structural network to pursue and surround its prey. The spotted hyena is the only carnivorous dog that can compete with lion packs, with a strong body and intelligent mind, and belongs to a highly socialized group [19]. Indian scholars Dhiman and Kumarl proposed a simulation algorithm called Spotted Hyena Optimizer (SHO) based on this high hunting rate form of predation method. It includes four stages: search, siege, hunting, and attack, and the mathematical model established based on this principle can achieve the goal of optimization [20]. The principle is to constantly approach and surround the prey after identifying the target location. SHO is expressed in the form of a mathematical model, where the individual who searches for prey is the best search point, while other search points update their location in real-time. Formula (1) is the distance formula between spotted hyenas and prey [21].

$$D_h = |B \cdot P_P(t) - P(t)| \tag{1}$$

In formula (1) above, D_h represents the distance between the searched individual and the captured prey. B represents the coefficient vector. t represents the number of iterations. P_P and P represent the location of the prey searched and the location of the individual searched at the time of iteration t. When the number of iterations is $t + 1$, the search for individual position points is formula (2):

$$P(t+1) = P_P(t) - E \cdot D_h \tag{2}$$

In equation (2), E represents the coefficient vector. The search individual position in iteration $t + 1$ is related to the prey point and the distance between the two. Both formulas (1) and (2) contain coefficient vectors, and their expressions are formula (3):

$$\begin{cases} B = 2 \cdot r_{d1} \\ E = 2h \cdot r_{d2} \end{cases} \tag{3}$$

In formula (3), both r_{d1} and r_{d2} are any number within [0,1]; h represents a control factor that linearly decreases between 5 and 0. The calculation of control factors is related to the number of iterations, as shown in formula (4):

$$|h| = 5 - [I_{inter}(5/M_{inter})] \tag{4}$$

In equation (4) above, M_{inter} represents the maximum number of iterations; I_{inter} represents any natural number other than zero. Spotted hyenas often engage in group warfare, relying on reliable communication networks within the group to surround their prey. Assuming that the best searched individual is infinitely close to the prey, the other individuals determine the position of the best searched individual as the prey position, forming a cluster and jointly moving towards the best point position. Save the best value updated at that time. The calculation formula is obtained from the following equation (5):

$$\begin{cases} D_h = |B \cdot P_P - P_k| \\ P_k = P_h - E \cdot D_h \end{cases} \tag{5}$$

In equation (5), P_h is the optimal position for the class of hyenas; P_k represents the position of the remaining hyenas. According to equation (5), the search for individuals is based on a reliable population network and the location of the prey found for hunting.

$$\begin{cases} C_h = P_k + P_{k+1} + \ldots + P_{k+N} \\ N = C_{nos}(P_h, P_{h+1}, P_{h+2}, \ldots, P_h + M) \end{cases} \tag{6}$$

In equation (6), N is the number of all spotted hyenas; C_h represents the set of N optimal solutions; C_{nos} is the number of alternative solutions obtained by increasing the random quantity ME. The process of spotted hyenas attacking prey in a cluster is also the process of local search in the mathematical model. In this process, the control factor h gradually decreases and the coefficient vector E also constantly changes. When the absolute value of E is less than 1, it is the moment of attack. Otherwise, the search for prey continues. The calculation formula for this process is formula (7):

$$P(t+1) = \frac{C_h}{N} \tag{7}$$

Most spotted hyenas are concentrated in the optimal search individual set C_h, where individuals disperse and pursue their prey. The critical condition is whether $|E|$ is greater than 1, which forcibly limits the distance between spotted hyenas and their prey. Expanding the search phase can help find the best hunting location and ensure the smooth implementation of global searches. The research aims to achieve the goal of image classification by improving algorithms to achieve hyperspectral band selection. Figure 1 shows the working principle of HiR-HRST:

Hyperspectral remote sensing is a technology based on electromagnetic spectrum and spectral imaging. The sensor will classify and record the waves at the same position according to the frequency. The data is recorded in the form of a cube, which contains a large amount of band data. Therefore, external forces are needed for image classification to achieve imaging purposes [22].

B. IMPROVED AND OPTIMIZED SPOTTED HYENA ALGORITHM

Although the SHO algorithm has high adaptability and parameter optimization ability, it may encounter defects such as weakened convergence ability and local optimization when facing high-dimensional optimization events. Therefore, the study introduces the Golden Sine Algorithm (Golden-SA) algorithm to improve SHO, which is a new meta heuristic algorithm based on sine function, with equal distribution

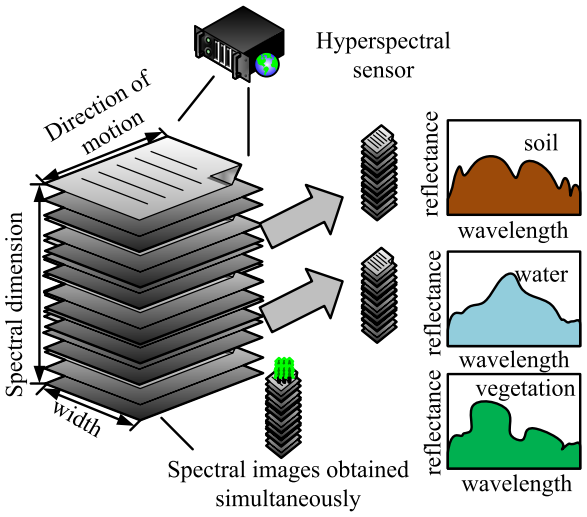

FIGURE 1. Principle of hyperspectral remote sensing imaging.

in all dimensions [23]. The biggest difference between this algorithm and traditional optimization problem solving algorithms is the update mode of position. The algorithm initially randomly generates a number of individual positions of s. Assuming that the solution values of the optimization problem all exist at the corresponding individual positions $X_i^{(t)}$, the individual position calculation formula for the t and $t+1$ iterations is formula (8) [24]:

$$\begin{cases} X_i^{(t)} = (X_{i,1}, X_{i,2}, \ldots, X_{i,d}) \\ X_i^{(t+1)} = X_i^{(t)} \left|\sin R_1\right| + R_2 \sin R_1 \cdot \left|x_1 P_i^{(t)} - x_2 X_i^{(t)}\right| \end{cases} \tag{8}$$

In equation (8) above, d represents the dimension quantity; i represents an individual point; t represents the number of iterations; $P_i^{(t)}$ represents the optimal position. R_1 is a random number of $[0, 2\pi]$, which is related to the length of motion of subsequent individuals during iteration. R_2 is a random number of $[0, \pi]$, which is related to the direction of motion of subsequent individuals during iteration. x_1 and x_2 are the coefficient of the golden section number, which can reduce the search space and quickly find the optimal value. Equation (9) is its expression:

$$\begin{cases} x_1 = -\pi + 2\pi(1-\tau) \\ x_1 = -\pi + 2\pi\tau \end{cases} \tag{9}$$

In equation (9), the value of the golden section number τ is $(\sqrt{5}-1)/2$. The working principle of the Golden-SA algorithm is listed in Figure 2:

It is known that SHO takes the best search individual's position point as the moving target, and obtains position update information by reducing the linear distance between the two. This type of algorithm lacks data communication. Furthermore, using parameter vector E to control the search range cannot accurately simulate the real search process. Thus, the SHO algorithm is not competent for complex

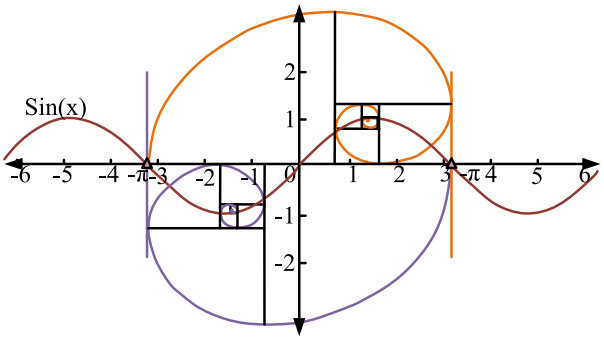

FIGURE 2. Working principle of Golden-SA algorithm.

optimization problems, otherwise it may result in decreased accuracy and local optimization. The study introduces the concept of golden sine to improve the SHO algorithm, and the position update calculation formula of this method is formula (10):

$$P_{h_i}^{(t+1)} = P_{h_i}^{(t)} \left|\sin R_1\right| + R_2 \sin R_1 \cdot \left|x_1 P_{k_i}^{(t)} - x_2 P_{h_i}^{(t)}\right| \tag{10}$$

Comparing equation (10) with equation (8) above, it is only necessary to know the optimal individual position P_h and the position data of the remaining spotted hyena P_k to achieve the golden sine strategy. It is worth noting that these data are easily available in SHO. In each iteration process, the remaining individuals communicate with the best individual to make the positional difference data between the two more accurate. According to the relationship between the sine function and the unit circle, the algorithm can traverse all values on the unit circle, thus possessing higher global search ability. The golden section coefficient not only grasps the direction of motion and distance of the position, but also continuously reduces the search space. This greatly improves the convergence ability and optimization efficiency of the model, allowing the spotted hyena to move faster to the optimal individual point position and obtain better optimization results. From the previous text, the SHO algorithm obtains the optimal solution through continuous iteration, so the optimization efficiency and accuracy of the algorithm are greatly affected by the initial cluster. However, initially, clusters were randomly initialized, and such clusters were often not average enough to cover the entire search space. This study utilizes chaotic algorithms to assist in initializing the original algorithm [25]. Scholars have proposed a chaotic algorithm based on the extremely high regularity and sensitivity of chaotic motion, which has been applied in multiple fields such as mechanics and engineering. The study utilizes chaotic mapping functions for cluster initialization, greatly enhancing the diversity and randomness of the initial cluster. The commonly used function is the Logistic function. Due to the distribution limitations of the generated sequence, there is often a phenomenon of high, medium, and low ends, which cannot achieve a symmetrical arrangement of data. The initial cluster obtained in this way is still unevenly distributed,

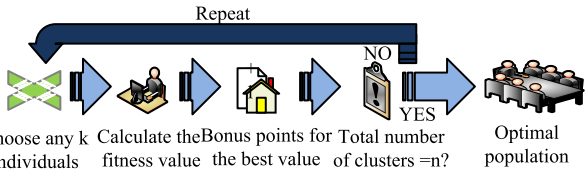

FIGURE 3. Tournament strategy flow.

resulting in a decrease in computational accuracy. Therefore, the study utilizes a uniformly arranged and faster Tent sequence to achieve initialization operations. The specific calculation formula is expressed as formula (11) [26]:

$$P_{t+1} = \begin{Bmatrix} 2P_t & 0 \le P_t < 0.5 \\ 2(1-P_t) & 0.5 \le P_t \le 1 \end{Bmatrix} \tag{11}$$

In equation (11), P_t represents a chaotic sequence, with an initial P_t value of 0.6. By using the Bernoulli shift to convert the formula, the calculation formula shown in formula (12) can be obtained:

$$P_{t+1} = (2P_t) \bmod 1 \tag{12}$$

To address the drawbacks of local optimization in the spotted hyena algorithm, a tournament strategy was selected to improve it. Its principle is a local selection mechanism, which uses the form similar to the wheel race to obtain the position with the highest fitness. And the data communication between various points has also been strengthened, further avoiding the occurrence of local optimization phenomena, and the algorithm accuracy is also higher. Fig.3 shows the specific steps:

From Fig. 3 that the tournament strategy will first select k individuals randomly from the cluster, usually two, and then compare the fitness of k individuals one by one to add one point to the best "class hyena". Repeat the above operations to concentrate the individuals with the top n scores and ultimately form an optimal initial cluster. This method is adopted for initialization to ensure that all data can be compared and filtered, and to obtain the best initial cluster, laying a solid foundation for subsequent optimization efficiency. It should be noted that this strategy is not based on the fitness value to achieve selection, but to detect the relationship between individuals' fitness, which is to suppress the rapid convergence problem caused by super individuals. Individuals determine whether they can enter the initial cluster based on their scores. Compared to traditional algorithms, this selection system can achieve more outstanding clusters and provide guarantees for the efficiency and accuracy of optimization problems.

C. APPLICATION OF IMPROVED SHO IN HYPERSPECTRAL BAND SELECTION

The image classification method combining spectral data and ground feature data often accompanies phenomena such as high computational intensity, high band correlation, and complex dimensionality when dealing with data with abundant band information, greatly reducing classification efficiency. Based on this, scholars propose a band selection technique that reduces high-dimensional data to easily computable low-dimensional forms, and treats each band as independent features. Therefore, band selection is essentially the selection of features. The study combines Golden-SA, chaos strategy, and tournament selection mechanism to improve SHO, to reduce computational burden, improve algorithm accuracy, and ultimately achieve the selection of hyperspectral bands. Fig. 4 shows the overall algorithm flow:

The goal of band selection is to select the optimal band from multiple known bands. Its characteristic is that it does not consider the specific features of each band, and only requires the band to match the required features in quantity. A common method is to use multiple models to predict data, then compare the predicted value with the true value, and select the optimal result as the final result. In practical applications, the band selection problem can be transformed into a mathematical programming problem, with the optimization goal of maximizing the objective function under given constraints [27]. The selection of frequency bands is relatively complex, and ordinary decimal programming usually cannot solve this discontinuous composite optimization problem. Therefore, binary algorithms are studied to express all individuals as logical numbers of 0 or 1. The algorithm used in the study is consistent with the original hyperspectral in terms of encoding dimension, and the retention and removal of bands are represented by logical values. In binary encoding, 0 represents discarding the band, 1 represents retaining the band and classifying it, and its mapping form is simple [28]. This study uses the sigmoid function for binary transformation, as shown in formula (13):

$$\begin{cases} Sigmoid(P) = 1/[1+\exp(-P)], \\ P_b = \begin{cases} 1, rand() < Sigmoid(P) \\ 0, rand() \ge Sigmoid(P), \end{cases} \end{cases} \tag{13}$$

In equation (13) above, P represents the location of the individual; $P_b$is the data after binary conversion; $rand()$ represents any number within [0,1]. Fitness function is a way to evaluate the quality of individuals, so it greatly affects the optimization results. The selection of hyperspectral bands pursues the effect of achieving twice the result with half the effort, that is, obtaining better classification accuracy in a small number of bands. This study uses the maximum and minimum method to achieve the purpose, that is, to structure the fitness function by improving the classification accuracy and reducing the number of bands. The fitness function of individual i is formula (14):

$$F(i) = \alpha[1 - OA(i)] + \beta\frac{R}{N} \tag{14}$$

Equation (14) above combines the two requirements and finally decides to take the minimum value of the fitness function as the optimal value. Where α and β both represent weight parameters; R and N represent the number of selected bands and the overall number of bands; $OA(i)$ represents the

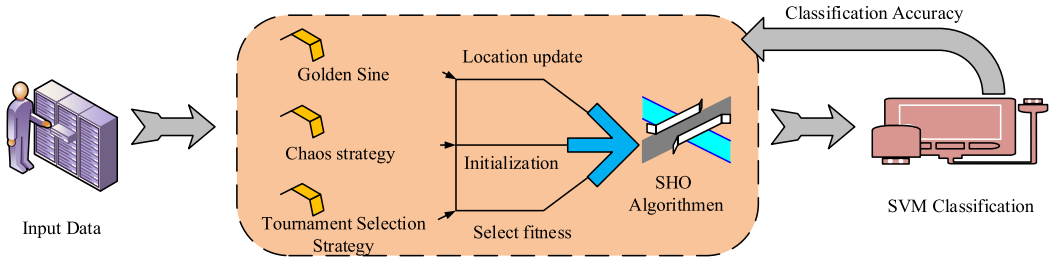

FIGURE 4. Overall classification flow chart.

classification accuracy of the current band. Formula (15) is its expression:

$$OA(i) = F_p/(F_p + F_N) \tag{15}$$

In equation (15), F_p and F_N represent the correct and incorrect classification numbers of the band subset, respectively. The above two methods further improve the efficiency of band selection. This study incorporates the idea of hunting in groups of hyenas into the optimization problem of band selection. When the number of bands in the hyperspectral image is n, the algorithm constructs an independent n as the spatial vector and selects an initialization cluster with a number of m. By searching for prey in the space and keeping up with the individual position with the fitness function, the algorithm stops working after judging the best result or meeting the preset iteration times. Fig. 5 shows the specific algorithm flow:

From Fig. 5, the initial cluster of search individuals needs to be selected through the chaotic strategy algorithm first, and then the initial parameter values are selected to calculate the fitness value $F(i)$ of the comparison individuals. Golden-SA is used to select the search individual with the best location, define the optimal solution according to formula (6), and track the location information of each search individual in real-time to detect whether there are any out of bounds individuals. Subsequently, appropriate adjustments will be made, and the optimal solution will be selected based on the tournament strategy. The cluster values and optimal search individual positions will be updated until the required conditions are met. On the contrary, repeat step five and once the stopping conditions are met, return to the optimal solution.

IV. SIMULATION EXPERIMENT AND RESULT ANALYSIS

A. EXPERIMENTAL DATASET AND ALGORITHM PARAMETER SETTINGS

This study selected two known hyperspectral remote sensing image datasets. One is the image data of Salinas Canyon in California, USA, captured using AVIRIS spectral imaging equipment, excluding bands with obvious defects (such as being absorbed by water vapor), leaving 204 bands in a total of 16 categories; The second is the image data located in Pavia City, Italy, which is shot by the ROSIS spectral imaging equipment to remove the noise polluted bands. There are 102 available band images, a total of 9 categories, and the image ratio is 1096 pixels × 715pixel. The experiment allocates the training and testing sets in three proportions: 2:8, 5:5, and 8:2. The repeated experiments show that the proportion division has little impact on the algorithm accuracy and can be ignored. Considering that the working time of the algorithm is positively correlated with the capacity of the training set, in order to save time and cost, a 2:8 allocation was selected for the dataset in the study. In summary, approximately 61 band data were used as the training set, while the remaining 245 data were all divided into test sets. This experiment selected a 3GHz central processing unit and a 4GB memory Windows 7 system, and implemented simulation on the MATLAB platform. The study used GSSHO, and the selected feature data was classified using Support Vector Machine (SVM). To demonstrate the reliable performance of this algorithm, four algorithms, Particle Swarm Optimization (PSO), Simulated Annealing Algorithm (SA), Genetic Algorithm (GA), and Grey Wolf Algorithm (GWO), were selected for comparative analysis in the experiment. The cluster capacity of the five algorithms is set to 10, and the maximum number of iterations is 15. Each type of algorithm needs to be repeated for 30 times to eliminate experimental errors caused by contingency. The average value of each type of algorithm will be used as comparative data. The control factor of the GSSHO algorithm is set to a linearly decreasing value within [0,5], with M candidate values within the range of [0.5,1]. Each time two values are selected for championship selection; The learning factors c1 and c2 of PSO are both 1.49, with weights ω between [0.2,0.9], the velocity is within the range of [0,6]; Annealing coefficient αTof SA is 0.99, and the initial temperature T0 is 0.1; The crossover probability C and mutation probability M of GA are both 0.4; Collaboration coefficient α of GWO isin the interval of [0,2]. The fitness value of each algorithm is taken as the first evaluation criterion, and the Overall Classification Accuracy (OA), the total number of selected bands N, Kappa coefficient and the operation duration of the algorithm are also selected as the auxiliary evaluation criteria. According to formula (15), classification accuracy refers to the ratio of the number of accurate classifications to the total number of classifications. The larger this value, the higher the classification accuracy. The Kappa coefficient is an indicator of the consistency between the classification value and the true value of the evaluation algorithm model. Similarly, the larger the value,

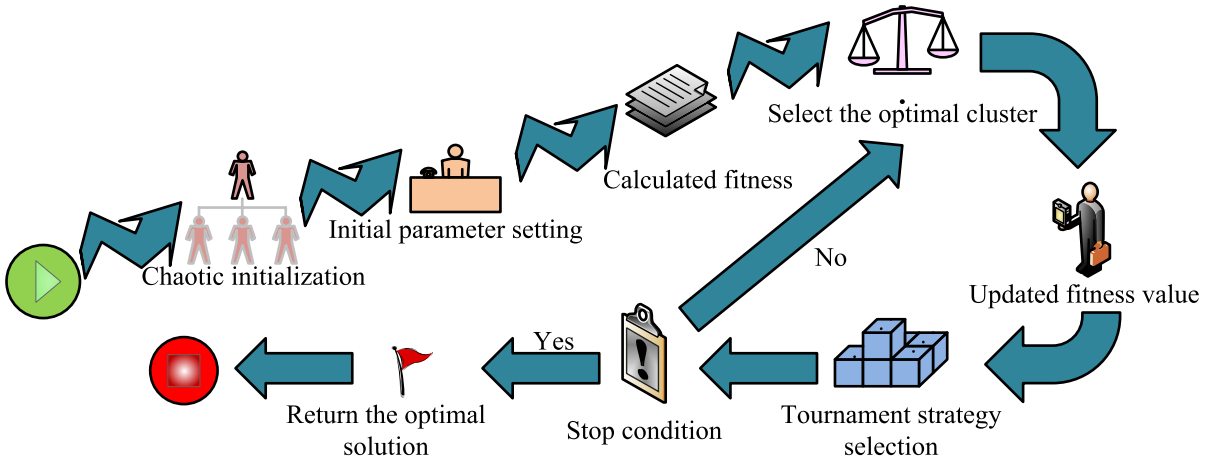

FIGURE 5. Overall process of algorithm.

the better the classification performance. The most important fitness value has a great impact on whether the algorithm can be realized and the convergence efficiency, and also directly reflects the classification accuracy and dimension reduction results. When the value of fitness is small, the performance of the algorithm is better. The experimental results on the Salinas dataset are displayed in Fig. 6:

From Fig. 6, the GA algorithm has the highest average classification accuracy, reaching 93.64%; The GSSHO algorithm used in the study is second only to this, reaching 93.33%, which is 0.59% higher than the PSO algorithm with the worst accuracy. Although there is a slight difference in classification accuracy between GSSHO and GA algorithms; However, considering the total number of bands selected by the algorithm, the GA with the best classification accuracy is clearly at a disadvantage, with the highest number among the five algorithms, reaching 121; GSSHO only has 32 bands, which is the least selected band in the experimental algorithm. The classification accuracy in Figure 6 (a) shows significant changes, which may be due to the imbalance between data quality and categories. Hyperspectral image data may be affected by factors such as noise, cluttered information, and changes in lighting. If the data quality is poor or contains a large amount of interference, accuracy may be affected; If the number of samples in certain categories is too small, classification algorithms may tend to misclassify these pixels into larger categories, thereby affecting accuracy. Therefore, in terms of classification accuracy and the number of selected bands, the GSSHO algorithm has the best overall performance. Fig.7 shows the experimental results of the fitness value, Kappa coefficient and running time of each algorithm:

As Fig.7, the average fitness function value of GSSHO algorithm is the lowest, only 0.0836, which is 0.0783 lower than that of GA algorithm; The Kappa coefficient of this algorithm is also relatively optimal, at 0.9283, which is an improvement of 0.0082 compared to the PSO algorithm; Moreover, the running time of the GSSHO algorithm is relatively short, only 10852 seconds, which is 1846 seconds

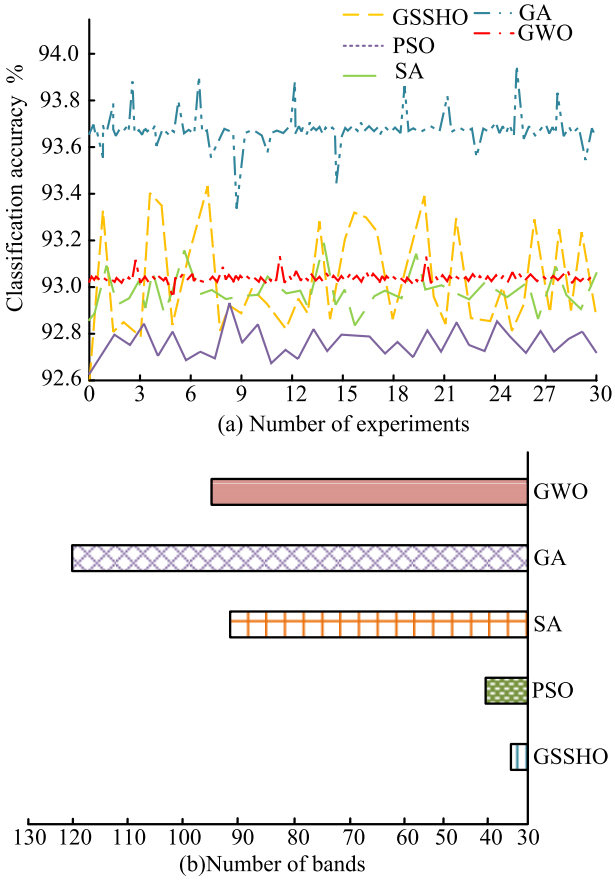

FIGURE 6. Salinas data set experiment results.

less than the SA algorithm. It can be seen from the comprehensive comparison of the three indicators that GSSHO is not necessarily the optimal value of a single indicator, but in general, GSSHO is the best option, which can guarantee low fitness value and short running time at the same time. Based on Figures 5 and 6, the comprehensive performance

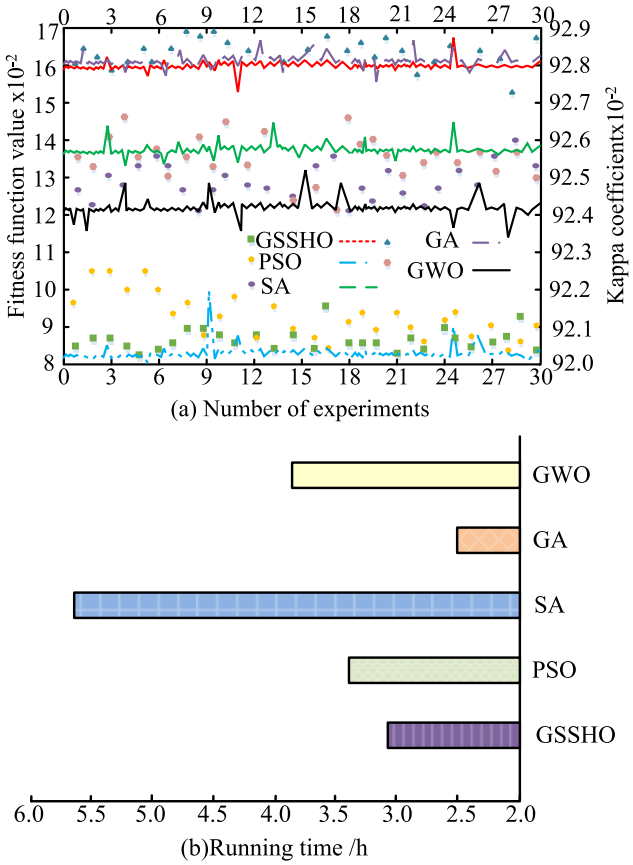

FIGURE 7. Salinas data set experiment results.

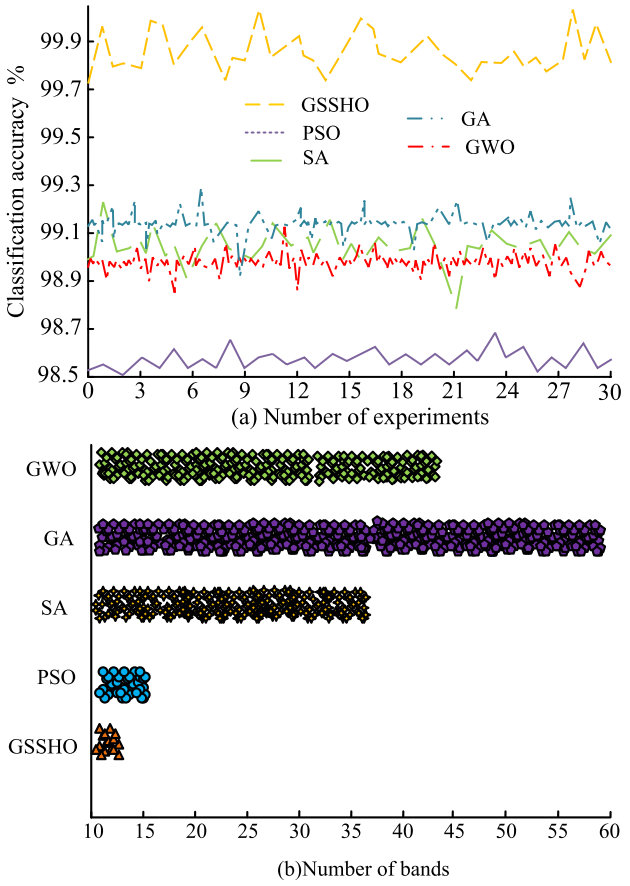

FIGURE 8. Pavian centre data set experiment results.

of GSSHO is the best among the five algorithms. Although the classification accuracy is slightly lower than that of the GA algorithm, its results demonstrate a lower total number of bands, approximately one quarter of that of the GA algorithm. Therefore, GSSHO can improve its dimensionality reduction ability while also ensuring algorithm accuracy. The study conducted the same experiment on another Pavian Centre dataset (Figure 8):

According to Fig. 8, the classification accuracy of GSSHO algorithm is significantly better than other algorithms, with an average of 99.12%, which is 0.51% higher than the PSO algorithm with the worst accuracy; Not only that, the number of bands selected by the algorithm has also been significantly reduced, to only 11, equivalent to one-fifth of the GA algorithm. Therefore, GSSHO is the best in terms of computational accuracy and dimensionality reduction ability. Fig. 9 shows the results of three other indicators compared with algorithms:

As Fig. 9, the average value of the fitness function of the GSSHO algorithm is 0.0315, which is 0.0944 lower than that of the GA algorithm. Moreover, the Kappa coefficient and runtime of the algorithm remain within a relatively optimal range of 0.9856 and 7038s, respectively. Although it is 2108s more than GA, its fitness function value and Kappa coefficient are far better than GA algorithm. Overall, GSSHO is still the best of the five algorithms. It is worth noting that in the two dataset experiments, the total number of bands for GSSHO was 32 and 11, respectively, while the bands for the original data were 204 and 102, respectively. GSSHO has excellent data dimensionality reduction capabilities, simplifying a large amount of complex data and laying a solid foundation for image classification. The values of its fitness function are all infinitely close to 0, indicating that the convergence of the algorithm is extremely strong. To further observe the classification results of each algorithm, the images of its wavelet subset were analyzed in the experiment. The results of the Salinas dataset are Figure 10:

From Fig. 10, the classification accuracy of each algorithm for Class 8 and Class 15 is relatively low, that is, the classification of the parts circled in the figure is relatively turbid. The support vector machine classifier is easy to confuse these two types of data, because the spectral characteristics between the data are similar, and the information of the band subset obtained by the algorithm is not sufficient, so it can not be classified accurately. Transforming the image into comparable data can yield experimental results as listed in Table 1:

The data in Table 1 shows that the classification accuracy of GSSHO is only slightly lower than GA, which can be ignored, but it is significantly better than other algorithms. The classification accuracy for Class 8 data reaches 90.38%,

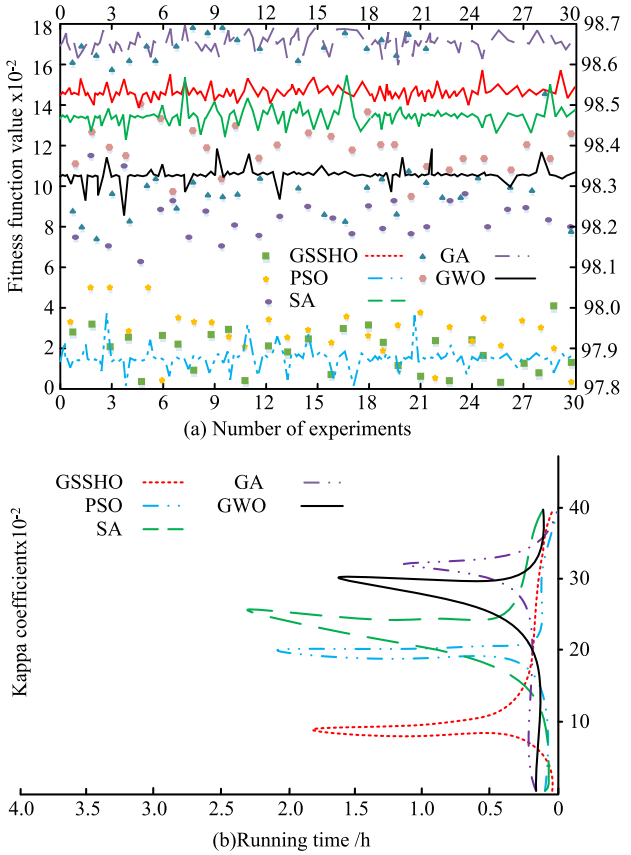

FIGURE 9. Pavian centre data set experiment results.

TABLE 1. Classification accuracy of two categories with large errors in Salinas data set.

	Category 8	Category 15
GSSHO	90.38%	68.07%
PSO	89.06%	64.89%
SA	90.03%	68.28%
GA	90.55%	68.56%
GWO	89.85%	67.56%

which is 1.32% higher than the PSO algorithm; The classification accuracy of GSSHO for class 15 data is 3.18% higher than that of PSO algorithm. Therefore, when faced with data that is difficult to recognize, GSSHO still outperforms other algorithms in terms of performance. Fig. 11 shows the band subset classification results of the Pavian Centre dataset:

Fig.11 shows that the decomposition results of each algorithm for the Pavian Centre dataset are good, and there is basically no confusion. And the algorithm accuracy is infinitely close to 1, and the classification effect is very excellent. To eliminate other factors and further understand the performance of GSSHO algorithm, this study selected three different classifiers for further control experiments. They are: K-Nearest Neighbor (KNN), Decision Tree (DT), and Naive Bayes (NB). The total number of bands, the value of fitness function and the classification accuracy are selected

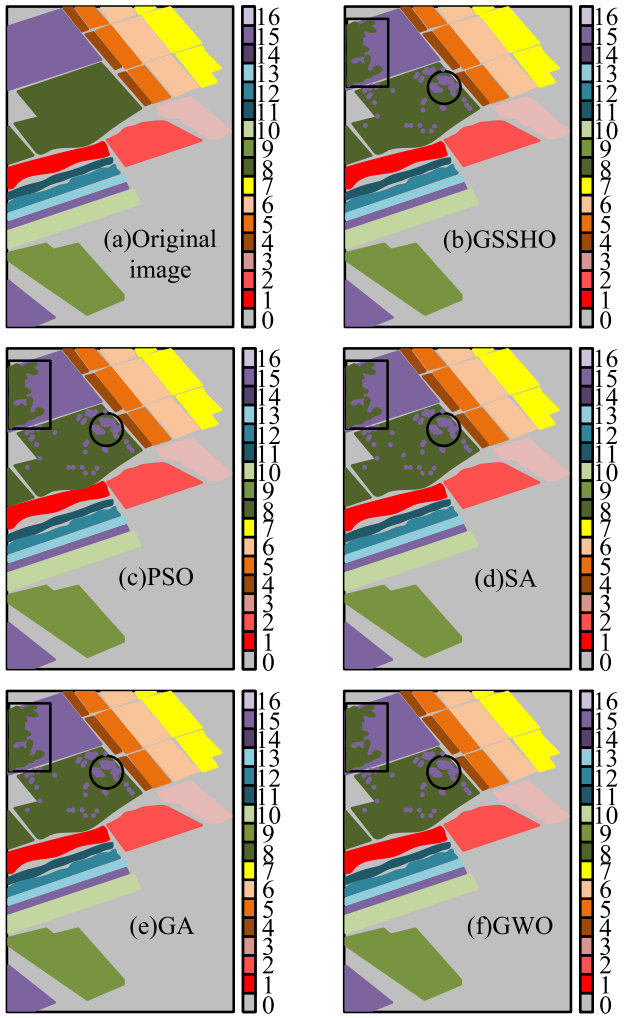

FIGURE 10. Band subset diagram of salinas data set.

TABLE 2. Results of different classifiers in the Salinas dataset.

	SVM	KNN	DT	NB
Classification accuracy	93.33%	90.53%	88.04%	76.94%
Number of bands	32	42	40	36
Fitness function value	0.0836	0.1201	0.1342	0.2214

to evaluate the selected subset of bands. Table 2 shows the experimental results in the Salinas dataset:

In Table 2, GSSHO has the best performance when using SVM classifiers, and can also achieve good performance in KNN and DT classifiers. However, the performance in NB classifiers is not ideal, and the performance of all indicators has significantly decreased. The classification results of GSSHO algorithm are poor. The reason is that the performance of the NB classifier will decline when facing data with more complex dimensional features, while the SVM classifier is better at handling small sample data with discretization and high dimensions. The comparison results of each classifier in the dataset are exhibited in Table 3:

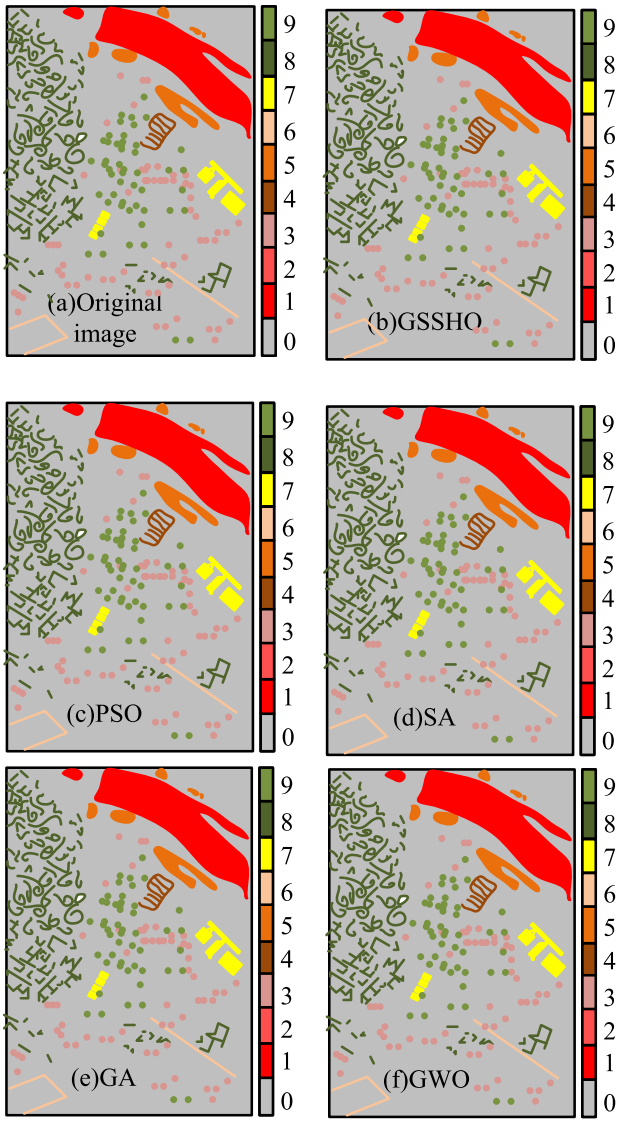

FIGURE 11. Band subset diagram of pavian centre data set.

TABLE 3. Results of different classifiers in the Pavian Centre dataset.

	SVM	KNN	DT	NB
Classification accuracy	99.12%	98.20%	97.22%	88.69%
Number of bands	11	38	23	14
Fitness function value	0.0315	0.0939	0.0631	0.1123

The data in Table 3 shows that GSSHO performs extremely well in SVM classifiers, with a classification accuracy close to 1, which is the same as the Salinas dataset; The performance of the NB algorithm is unsatisfactory, but the accuracy also reaches 88.69%. Overall, GSSHO performs better than in the Salinas dataset. Through the above experiments, it can be concluded that using the GSSHO algorithm for hyperspectral pixel classification can achieve more efficient classification results. This algorithm performs well in data dimensionality reduction and convergence, greatly reducing computational intensity and achieving twice the result with half the effort.

Based on the analysis of the above results, the golden sine algorithm and the chaotic Spotted hyena algorithm both adopt a global search strategy. By introducing randomness and chaos strategies, they can better traverse the solution space and find the global optimal solution. Chaotic Spotted hyena algorithm introduces tournament selection strategy, which can retain the optimal solution, and uses chaos strategy to introduce greater randomness in each iteration. This can avoid falling into the local optimal solution and improve the ability of local search. The golden sine algorithm and the chaotic Spotted hyena algorithm comprehensively utilize the characteristics of golden ratio, sine function and chaotic mapping, giving full play to the advantages of these attributes. By combining these methods, it is possible to better balance exploration and utilization of search space in both global and local searches. Therefore, the proposed algorithm has better performance in hyperspectral image classification.

V. CONCLUSION

This study improves the original SHO by introducing three methods: Golden SA, chaos strategy, and tournament strategy. The improved SHO improves the algorithm's global search ability for data, solving the problems of random selection of initial clusters and local optimization in the original algorithm. To further validate the performance of the improved algorithm, two datasets, Salinas and Pavian Centre, were used for simulation experiments, while PSO/SA/GA/GWO algorithms were added for comparative experiments. The experimental data shows that the algorithm has the best comprehensive performance in both datasets, with classification accuracy reaching 93.33% and 99.12%, respectively; The total number of selected bands has been reduced by 6/1 and 9/1 respectively compared to the original data. The experiment also makes a comparative analysis of the performance of the algorithm in different classifiers: GSSHO algorithm performs best in SVM, but performs poorly in NB classifier, and its performance in all aspects is significantly reduced. Its fitness function values rise to 0.2214 and 0.1123 respectively. This is because the NB algorithm itself is not good at handling data with complex dimensional features. In summary, the GSSHO algorithm has extremely strong global search ability, fast convergence ability, data dimensionality reduction ability, and image classification ability compared to other algorithms. It has significant advantages in operational efficiency, data processing, and classification accuracy, and also has strong stability. Able to handle complex feature data and achieve more efficient hyperspectral image classification. Although the research has achieved certain results, the GSSHO algorithm still has shortcomings. The search process of the golden sine algorithm and the chaotic Spotted hyena algorithm involves a large number of parameter updates and fitness evaluation, which may lead to high computational complexity of the algorithm and may face a long computational time when processing large-scale hyperspectral images. Therefore, subsequent research can further improve the operational efficiency of the algorithm

by introducing methods such as parallel computing, pruning, and pruning.

REFERENCES

[1] L. I. Hai, C. Yang, W. Hui, S. Zhu, and Q. Zhang, "Changes of glaciers and glacier lakes in Alpine and extremely Alpine regions using remote sensing technology: A case study in the Shisha Pangma area of southern Tibet," *Chin. J. Geol. Hazard Control*, vol. 32, no. 5, pp. 10–17, 2022.
[2] T. Mao, L. Yu, Y. Zhang, and L. Zhou, "Modified Mahalanobis–Taguchi system based on proper orthogonal decomposition for high-dimensional-small-sample-size data classification," *Math. Biosci. Eng.*, vol. 18, no. 1, pp. 426–444, 2021.
[3] M. A. Elsabagh, M. S. Farhan, and M. G. Gafar, "Meta-heuristic optimization algorithm for predicting software defects," *Expert Syst.*, vol. 38, no. 8, pp. 12768–12769, Dec. 2021.
[4] S. Peyghambari and Y. Zhang, "Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: An updated review," *J. Appl. Remote Sens.*, vol. 15, no. 3, Jul. 2021, Art. no. 031501.
[5] P. Qin, Y. Cai, and X. Wang, "Small waterbody extraction with improved U-Net using Zhuhai-1 hyperspectral remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 3047918–3047933, 2021.
[6] S. C. Liew, C. K. Choo, J. W. M. Lau, W. S. Chan, and T. C. Dang, "Monitoring water quality in Singapore reservoirs with hyperspectral remote sensing technology," *Water Pract. Technol.*, vol. 14, no. 1, pp. 118–125, Mar. 2019.
[7] J. Li, H. Shen, H. Li, M. Jiang, and Q. Yuan, "Radiometric quality improvement of hyperspectral remote sensing images: A technical tutorial on variational framework," *J. Appl. Remote Sens.*, vol. 15, no. 3, Sep. 2021, Art. no. 031502.
[8] Y. Yang, X. Wang, M. Huang, Q. Zhu, Y. Guo, L. Xu, and Z. Zhou, "Hyperspectral band selection based on dual evaluation measures and improved nondominated sorting genetic algorithm," *J. Appl. Remote Sens.*, vol. 15, no. 2, Jun. 2021, Art. no. 028504.
[9] L. Zhang, Q. Wei, and Z. Lu, "Combined optimization of frequency band and time segment using quantum particle swarm algorithm for brain-computer interfaces," *J. Circuits, Syst. Comput.*, vol. 30, no. 13, pp. 2150234–2150255, Oct. 2021.
[10] T. J. Rato and M. S. Reis, "Multiresolution interval partial least squares: A framework for waveband selection and resolution optimization," *Chemometric Intell. Lab. Syst.*, vol. 186, no. 15, pp. 41–54, Mar. 2019.
[11] L. Luo, G. He, C. Chen, X. Ji, L. Zhou, Y. Dai, and Y. Dang, "Adaptive data dimensionality reduction for chemical process modeling based on the information criterion related to data association and redundancy," *Ind. Eng. Chem. Res.*, vol. 61, no. 2, pp. 1148–1166, Jan. 2022.
[12] B. Bharathi and M. D. A. Praveena, "A hybrid feature selection algorithm for big data dimensionality reduction," *Int. J. Adv. Intell. Paradigms*, vol. 19, nos. 1–2, pp. 380–392, Jan. 2021.
[13] B. Parlak and A. K. Uysal, "A novel filter feature selection method for text classification: Extensive feature selector," *J. Inf. Sci.*, vol. 49, no. 1, pp. 59–78, Feb. 2023.
[14] X. Yu, W. Long, Y. Li, L. Gao, and X. Shi, "Trajectory dimensionality reduction and hyperparameter settings of DBSCAN for trajectory clustering," *IET Intell. Transp. Syst.*, vol. 16, no. 5, pp. 691–710, May 2022.
[15] T. Guo, K. Yu, M. Aloqaily, and S. Wan, "Constructing a prior-dependent graph for data clustering and dimension reduction in the edge of AIoT," *Future Gener. Comput. Syst.*, vol. 128, pp. 381–394, Mar. 2022.
[16] Z. Xiang, G. Zhou, Y. Zhou, and Q. Luo, "Golden sine cosine SALP swarm algorithm for shape matching using atomic potential function," *Expert Syst.*, vol. 39, no. 2, pp. e12854.1–e12854.32, Feb. 2022.
[17] Q. Luo, J. Li, Y. Zhou, and L. Liao, "Using spotted hyena optimizer for training feedforward neural networks," *Cognit. Syst. Res.*, vol. 65, pp. 1–16, Jan. 2021.
[18] M. A. Baset, Y. Zhou, and I. Hezam, "Use of a sine cosine algorithm combined with Simpson method for numerical integration," *Int. J. Math. Oper. Res.*, vol. 14, no. 3, pp. 307–318, Apr. 2019.
[19] J. Prasanth Ram, D. S. Pillai, D. Mathew, J. Ha, and Y.-J. Kim, "A simple, reliable and adaptive approach to estimate photovoltaic parameters using spotted hyena optimization: A framework intelligent to predict photovoltaic parameters for any meteorological change," *Sol. Energy*, vol. 236, no. 1, pp. 480–498, Apr. 2022.
[20] A. Naderipour, Z. Abdul-Malek, M. Hajivand, Z. M. Seifabad, M. A. Farsi, S. A. Nowdeh, and I. F. Davoudkhani, "Spotted hyena optimizer algorithm for capacitor allocation in radial distribution system with distributed generation and microgrid operation considering different load types," *Sci. Rep.*, vol. 11, no. 1, pp. 2728–2732, Feb. 2021.
[21] F. Martinez-Rios and A. Murillo-Suarez, "Multi-threaded spotted hyena optimizer with thread-crossing techniques," *Proc. Comput. Sci.*, vol. 179, no. 2, pp. 432–439, 2021.
[22] V. Manoharan, and S. Tamilperuvalathan, "Prediction on enhanced electrochemical discharge machining behaviors of zirconia-silicon nitride using hybrid DNN based spotted hyena optimization," *Int. J. Energ. Res.*, vol. 46, no. 7, pp. 9221–9441, 2012.
[23] J. Zan, "Research on robot path perception and optimization technology based on whale optimization algorithm," *J. Comput. Cognit. Eng.*, vol. 1, no. 4, pp. 201–208, 2022.
[24] Y. Guo, Z. Mustafaoglu, and D. Koundal, "Spam detection using bidirectional transformers and machine learning classifier algorithms," *J. Comput. Cognit. Eng.*, pp. 5–9, Apr. 2022.
[25] Y. Lu, X. Wu, L. Yao, T. Zhang, and X. Zhou, "Multi-channel data aggregation scheduling based on the chaotic firework algorithm for the battery-free wireless sensor network," *Symmetry*, vol. 14, no. 8, p. 1571, Jul. 2022.
[26] S. Y. D. Nezhad, N. Safdarian, and S. A. H. Zadeh, "New method for fingerprint images encryption using DNA sequence and chaotic tent map," *Optik*, vol. 224, no. 7, Dec. 2020, Art. no. 165661.
[27] C. He, P. Zhong, M. Li, L. Jie, and Z. Li, "An evolutionary many-objective algorithm based on a novel tournament selection strategy," *J. Comput. Commun.*, vol. 09, no. 6, pp. 238–248, 2021.
[28] B. Cai, L. Liu, J. Wu, X. Che, and Y. Wang, "Three-dimensional shape measurement based on spatial–temporal binary-coding method," *Meas. Sci. Technol.*, vol. 32, no. 9, pp. 95014–95025, 2021.

XIPING YANG was born in Henan, China, in 1978. She received the bachelor's degree from Henan Normal University, in 2001, and the master's degree from the Beijing Institute of Technology, in 2007. She is currently with the Zhengzhou University of Aeronautics. Her research interest includes combinatorial optimization.

LIFANG CHENG was born in Hebei, China, in 1978. She received the bachelor's degree from Hebei Normal University, in 2002, the master's degree from Shanxi University, in 2005, and the Ph.D. degree from Beijing Jiaotong University, in 2018. She has published a total of ten articles. Her research interest includes nonlinear dynamics.