A modified Cuckoo Search algorithm based optimal band subset selection approach for hyperspectral image classification

Shrutika S. Sawant,[a] Manoharan Prabukumar[a,*] and Sathishkumar Samiappan[b]

[a]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

[b]Geosystems Research Institute, Mississippi State University, Starkville, Mississippi State, USA

Contact

S.S. Sawant: shrutika.sawant1@gmail.com,
https://orcid.org/0000-0002-1532-947X
M. Prabukumar: mprabukumar@vit.ac.in,
https://orcid.org/0000-0002-8127-653X

S. Samiappan: sathish@gri.msstate.edu,
https://orcid.org/0000-0002-8443-883X

Band selection is an effective way to reduce the size of hyperspectral data and to overcome the "curse of dimensionality" in ground object classification. This paper presents a band selection approach based on modified Cuckoo Search (CS) optimisation with correlation-based initialisation. CS is a popular metaheuristic algorithm with efficient optimisation capabilities for band selection. However, it can easily fall into local optimum solutions. To avoid falling into a local optimum, an initialisation strategy based on correlation is adopted instead of random initialisation to initiate the location of nests. Experimental results with Indian Pines, Salinas and Pavia University datasets show that the proposed approach **obtains overall accuracy of 82.83%, 94.83% and 91.79%, respectively,** which is higher than the original CS algorithm, Genetic Algorithm (GA), Particle Swarm Optimisation (PSO) and Gray Wolf Optimisation (GWO).

Keywords: hyperspectral image, band selection, Cuckoo Search optimisation, correlation-based initialisation

Introduction

Hyperspectral imagery contains hundreds of contiguous spectral bands, which enhances the ability to discriminate very similar spectral objects.[1] However, such high dimensionality brings highly correlated and redundant band information which not only increases computational complexity but also may result in the Hughes phenomenon, leading to challenges in classification.[2–4] Therefore, dimensionality reduction becomes an essential task in hyperspectral image processing.[5] Dimensionality reduction can be done in two ways: band (feature) selection[6–8] and feature extraction.[9–12] Band selection is the process of selecting the set of useful bands which preserves the physical properties of the original hyperspectral data. On the other hand, feature extraction is the process of mapping the data from high-dimensional space to lower-dimensional space which provides complementary

information rather than preserving the physical characteristics of the original hyperspectral data.[13–16] To reduce dimensionality while preserving the class discrimination capability of hyperspectral imagery, band selection methods are widely used. According to the availability of class labels, band selection can be further divided into two types: supervised and unsupervised. Supervised band selection approaches use prior information to select the most discriminative bands.[17,18] Without any prior information of class labels, unsupervised band selection approaches select the most informative features to preserve the information of the original bands as much as possible.[19–21]

Several studies have been proposed for band selection such as clustering-based band selection,[22,23] and ranking-based methods.[24,25] Clustering-based band selection consists of two steps. First, bands are grouped into clusters, in which the inter-cluster variance is maximised and the intra-cluster variance is minimised. In the second step, the bands with the highest average correlations from their corresponding clusters are chosen as the best. Clustering is the most commonly used method of selecting discriminative bands and the selected discriminative bands are considered as the cluster centres. However, clustering-based band selection methods focus mainly on redundancy among the bands. Consequently, the most suitable and informative bands may be discarded in the selection process. In the ranking-based band selection method, the bands are selected based on the ranking, in which the rank of each band is first computed according to a definite evaluation criterion, and then the top-ranked bands are sorted in a sequence to form the subset. The main disadvantage of ranking methods is that correlation among bands is ignored while evaluating the discriminating ability of a band. As a result, in most cases, ranking-based methods select redundant bands.

Over the years, in the literature, numerous band selection approaches have been presented based on nature-inspired optimisation (also called metaheuristic) algorithms, including Artificial Bee Colony (ABC), Genetic Algorithm (GA), Particle Swarm Optimisation (PSO), Gray Wolf Optimisation (GWO), Wind Driven Optimisation, Cuckoo Search (CS) and so on.[26–31] These methods consider the band selection problem as a combinatorial optimisation problem which is solved by formulating an appropriate fitness function or objective function. The objective function evaluates the band subsets and returns the degree of their goodness. The objective

function needs to be defined carefully as it influences the performance of the system. It can be dependent or independent of the learning algorithm. Hence, the objective function can be modelled by independent evaluation criteria such as information measures (divergence, entropy or mutual information), distance measures [Bhattacharya distance, Kullback–Leibler divergence, Jeffries–Matusita distance, Hausdorff distance and Spectral Angle Mapper (SAM)] and dependency measures (correlation measures, similarity measures). Selection of an effective search strategy is very important in band selection. To optimise the objective function, an appropriate optimisation algorithm must be chosen to converge to the global optimum solution and not to get stuck in a local optimum. GA is an evolutionary optimisation algorithm inspired by Darwin's theory and PSO is inspired by the social behaviour of birds. Both algorithms search for the best possible combination of individuals in a population based on an objective function (fitness function). The search for each algorithm is determined based on the best of populations in each iteration. CS is a nature-inspired heuristic algorithm and widely used in solving many optimisation problems including hyperspectral imagery band selection. To enhance the performance of the standard CS, several variations have been proposed. Shortly after the introduction of basic CS, Yang and Deb[32] extended the basic CS to a multi-objective CS algorithm intended for solving design optimisation problems. Many binary CS variants have also been proposed for solving binary optimisation problems.[31,33,34] Medjahed et al.[31] proposed a new framework for the band selection problem based on binary CS. Recently, a lot of new CS variants have been proposed using hybridisation of other algorithms.[35–37] Incorporating adaptations to CS has also become a popular topic for further research. For example, recent adaptive CS variants are widely studied.[38–40] These variations may result in improvements; however, they increase the complexity. According to the "No Free Lunch Theorem", there is no optimisation algorithm which is sufficiently appropriate to solve all types of optimisation problems.[41] One of the major drawbacks of the CS algorithm is the initialisation strategy. The CS algorithm randomly initialises the location of the nest, which may repeat, and this increases the chance of falling into a local optimum solution. Therefore, in this paper, we propose a modified CS algorithm which uses a correlation-based initialisation strategy to avoid falling into a local optimum.

Background

In this section, the technical background of the standard CS algorithm is presented.

Standard CS algorithm

CS is a nature-inspired metaheuristic algorithm which mimics the breeding behaviour of some cuckoo species. It was proposed by Yang and Deb[42] to solve global optimisation problems. The CS algorithm combines the breeding behaviour of cuckoos with the Lévy flights observed in several species of birds. Lévy flight follows the searching behaviour of animals with a random walk, and the next walk always depends on the current step and the switching probability for the next step. The cuckoo lays its eggs in the nests of other birds. If the host bird discovers the cuckoo's egg, they will either throw the foreign eggs away or abandon the nest and build a new nest. In the standard CS algorithm, each egg of the host birds in a nest denotes a solution, and a cuckoo egg represents a new solution. If a new solution is better than the one in the nest, the worse one will be replaced with the new best solution. The development of the CS algorithm considers three guidelines. (1) Each cuckoo puts one egg in an arbitrarily selected nest. (2) The best nest with good quality eggs will be transferred to the next generations. (3) The number of available hosts' nests is fixed. The host bird determines the cuckoo's egg by a probability of $P \in [0, 1]$. If a host bird determines a cuckoo's egg, then the host bird can abandon the nest or throw the egg away and construct an entirely new nest using Lévy flights. Based on these guidelines, the pseudo-code of standard CS is presented as Algorithm 1.

In the CS algorithm, each egg can be regarded as a solution. In the initial step, each solution, x_i is generated randomly. A new solution x_{i+1} is generated by adopting a Lévy flight random walk as follows:

$$x_{i+1} = x_i + \alpha \oplus \text{levy}(\lambda) \qquad (1)$$

where $\alpha > 0$ is a step-size scaling factor; the product of $\oplus$ means entry-wise multiplication; $\text{levy}(\lambda)$ is a random walk with the random step length which is drawn from a Lévy distribution. A random walk is a Markov chain whose next location depends only on the current location and the transition probability. The $\text{levy}(\lambda)$ is calculated using Mantegna's algorithm as follows:

$$\text{levy}(\lambda) = \frac{\mu}{|\nu|^{1/\lambda}} \qquad (2)$$

where λ is a fixed parameter with value 3/2. Furthermore, μ and ν are drawn from a normal distribution with zero means and an associated variance, as follows:

$$\sigma_\mu = \left[\frac{\Gamma(1+\lambda)\sin(\pi\lambda/2)}{\Gamma\left[(1+\lambda/2)\lambda 2^{(\lambda-1)/2}\right]} \right] \qquad (3)$$

where $\Gamma(\cdot)$ is the standard Gamma function.

The standard CS considers the band selection problem as a combinatorial optimisation problem which is solved by formulating an appropriate fitness function or objective function. The objective function assesses the band subsets and provides the degree of their goodness. The objective function needs careful determination as it influences the performance of the system. It can be dependent or independent of the learning algorithm. Hence, the objective function can be modelled by a dependent or independent evaluation criterion. Independent evaluation criteria include information measures (divergence, entropy or mutual information), distance measures (Bhattacharya distance, Kullback–Leibler divergence, Jeffries–Matusita distance, Hausdorff distance and SAM) and dependency measures (correlation measures, similarity measures). On the other hand, dependent evaluation criteria require a predefined learning algorithm. The selection of an effective search strategy is a crucial task in the band selection problem. To optimise the defined objective function, an appropriate optimisation algorithm must be chosen to converge to the global optimal solution and escape from the local optimum. As standard CS is applied to solve the hyperspectral band selection problem, each cuckoo represents a band of hyperspectral data and each nest represents a solution to the problem (set of selected bands). Discovery probability p_a indicates that each band has a probability of p_a of being successful. An objective function is denoted as f and the value of the objective function is treated as the fitness value.

Algorithm 1: Band selection with standard CS

Input: User specified random band set b_i, max_iteration T, number of nests n

Output: Selected band set

    Step 1: Initialise population for each host n nests;
        Initialise random solution $b_i = \{b_1, b_2, \ldots, b_m\}$.
        Choose the current best nest by evaluating the objective function f_i

Step 2: *While* ($t < T$)

    Get a cuckoo egg arbitrarily via Lévy flights

    Evaluate the objective function f_j

    Select a nest from n random nests

    If ($f_i > f_j$)

        Replace j with the new solution

    Select worst nests according to probability P

        Replace them for new solution by Lévy flights

    Preserve the best solutions

    Rank the solutions and determine the current best solution

    end while

Standard CS may fall into a local optimum solution. The reason is that the standard CS algorithm initialises the initial solution or population or location of the nests randomly. Sometimes, the population will be the same, and sometimes the population is not adequately dispersed in a given dimension. Therefore, it causes repeated calculations and may result in local optima. Moreover, random initialisation may result in the selection of less informative bands as well as redundant bands. Hence, in this paper, a modified CS algorithm is proposed which uses initialisation based on correlation instead of random initialisation.

Materials and methods

This section describes the materials and methods used for the experimentation performed.

Experimental setup

To assess the effectiveness of the proposed band selection approach, a series of experiments are conducted on three standard datasets, Indian Pines, Salinas and Pavia University.[43] All the experiments are conducted using MATLAB platform version R2018b on an Intel Xeon processor 2.90 GHz CPU with 128 GB RAM in a Windows 10 (64 bit) environment. To have a fair comparison, the performance of the proposed method is compared with the other four metaheuristic band selection approaches, i.e. GA,[27] PSO,[28] GWO[29] and standard CS (SCS). The performance of the proposed method is compared with other competing methods using three well-known quality metrics, i.e. overall accuracy, average accuracy and kappa coefficient. Overall accuracy (OA) is the percentage of correctly classified pixels in the whole scene. Average Accuracy (AA) is the mean of percentage of correctly

labeled pixels for each class. Kappa Coefficient (κ) is a robust measure of the degree of agreement that integrates diagonal and off-diagonal entries of the confusion matrix.

To perform the classification task using selected bands, support vector machine (SVM) is employed as a classifier.[44] In SVM, Radial Basis Function is used as the kernel function. The parameters C and γ in SVM are determined by a five-fold cross-validation method ($\gamma = 2^{-3}, 2^{-2},..., 2^2, 2^3; C = 2^1, 2^2,..., 2^8$). In all experiments, to evaluate the effectiveness of the proposed method with less amount of labeled data, 20 % of samples for each class from the reference data of Indian Pines and Pavia University dataset are randomly chosen as training samples, and the remaining samples in each class are used for testing. The parameters of each algorithm are summarised in Table 1. The experiments are repeated ten times with a random split of training and testing samples and the corresponding average of OA, AA and k is recorded.

Experimental datasets

The Pavia University dataset was captured by the Reflective Optical System Imaging Spectrometer (ROSIS) over Pavia, Northern Italy in July 2002. This dataset contains nine different classes. The size of the dataset is 610 × 340 pixels with 1.3 m spatial resolution over the 430–860 nm range. This scene contains 103 spectral reflectance bands.

Table 1. Parameters setting of GA, PSO, GWO, SCS and CSCI.

Algorithm	Parameters	Value
GA	Number of genetics	20
	Selection ratio	0.8
	Crossover ratio	0.9
	Mutation ratio	0.01
PSO	Population size	20
	Number of iterations	100
	Cognitive constant (c1)	2
	Social constant (c2)	2
	Inertia weight	0.2
GWO	Number of search agents	20
	Number of iterations	100
SCS and CSCI	Population size	20
	Number of iterations	100
	Discovery probability p_a	0.25

The Indian Pines dataset was captured by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over the agriculture region of North-Western Indiana in June 1992. The spatial dimension of this dataset is 145 × 145 pixels with 20 m spatial resolution and 10 nm spectral resolution over the 400–2500 nm range. This scene contains 220 spectral reflectance bands, but only 200 bands will be available for the experiment after removal of 20 water absorption bands. This dataset comprises 16 classes of vegetation species. Due to the similar spectral signature of these vegetation species, distinguishing them is a challenging task.

The Salinas dataset was captured by AVIRIS over Salinas Valley, California. This dataset contains 16 different land cover classes. The size of the dataset is 512 × 217 pixels with a 3.7 m spatial resolution over the 400–2500 nm range. This scene contains 224 spectral reflectance bands.

Proposed approach

To overcome the drawback of the standard CS algorithm, the location of the nest is initialised based on the correlation between spectral bands. This ensures that the location of the nest will not repeat and hence accelerates the speed of convergence. Hyperspectral imagery bands are highly correlated. Initially, the hyperspectral image is partitioned into m subgroups or clusters using a k-means clustering algorithm. The partitioning reduces the statistical dependency between subgroups by increasing the class separation. The partitioning is done by finding the correlation coefficient among the original bands. The correlation coefficient indicates the closeness between the bands, which ranges from –1 to 1. Let the hyperspectral image cube be represented as, $\chi \in R^{H \times W \times N}$, where H and W are the height and width of the hyperspectral image cube and N is the total number of spectral bands. Let $I(x,y,i)$ and $I(x,y,j)$ be the ith and j^{th} band images of hyperspectral data. The correlation coefficient between two spectral bands is calculated by:

$$CC = \frac{Cov\left[I(x,y,i),I(x,y,j)\right]}{\sigma_i \sigma_j} \tag{4}$$

where *Cov* is the covariance between the two bands, σ_i and σ_j are standard deviations of the respective bands. If the correlation coefficient is –1, the two bands are correlated inversely. If the correlation coefficient is 0, they are completely uncorrelated and if the correlation

coefficient is 1, they are identical. The resulting clusters consist of a group of highly correlated bands, and bands from two different clusters will have less correlation. Figure 1 illustrates the absolute values of the correlation coefficients of the popular Indian Pines data, Pavia University data and Salinas data. The correlation matrix of the Indian Pines dataset with 200 bands is presented in Figure 1(a). From Figure 1(a) it can be observed that the spectral bands are partitioned into five non-uniform groups, so the bands in each subgroup are greatly correlated with one another. Similarly, the correlation matrix of the Pavia University dataset with 103 bands is presented in Figure 1(b). From Figure 1(b), it can be observed that the spectral bands are partitioned mainly into two non-uniform groups, so the bands in each subgroup are greatly correlated with one another. The correlation matrix of the Salinas dataset with 224 bands is presented in Figure 1(c). From Figure 1(c), it can be observed that the spectral bands are partitioned into eight non-uniform groups, so the bands in each subgroup are greatly correlated with one another.

Once clusters are formed, one representative band is chosen from each cluster which ensures the maximum separability among the bands of the different clusters and less separability among the bands of the same cluster. Equation (5) ensures the maximum separability among the bands of the different clusters and less separability among bands of the same cluster.

$$\{b_1,b_2,\ldots b_m\} = \arg\max\left\{\sum_m \sum_{m \neq n} d(b_m,b_n) - \sum_m \sum_{b_x \in m} d(b_x,b_m)\right\} \tag{5}$$

where b_m is selected the representative band from the mth cluster, $d(b_m,b_n)$ is the spectral distance between a representative band from the mth cluster and a representative band from the nth cluster using Bhattacharya distance as in Equation (6).

$$d(b_m,b_n) = \frac{1}{8}(\mu_m - \mu_n)^T \left(\frac{\Sigma_m + \Sigma_n}{2}\right)^{-1}(\mu_m - \mu_n)$$
$$+ \frac{1}{2}\ln\left(\frac{|(\Sigma_m + \Sigma_n)/2|}{|\Sigma_m|^{\frac{1}{2}}|\Sigma_n|^{\frac{1}{2}}}\right) \tag{6}$$

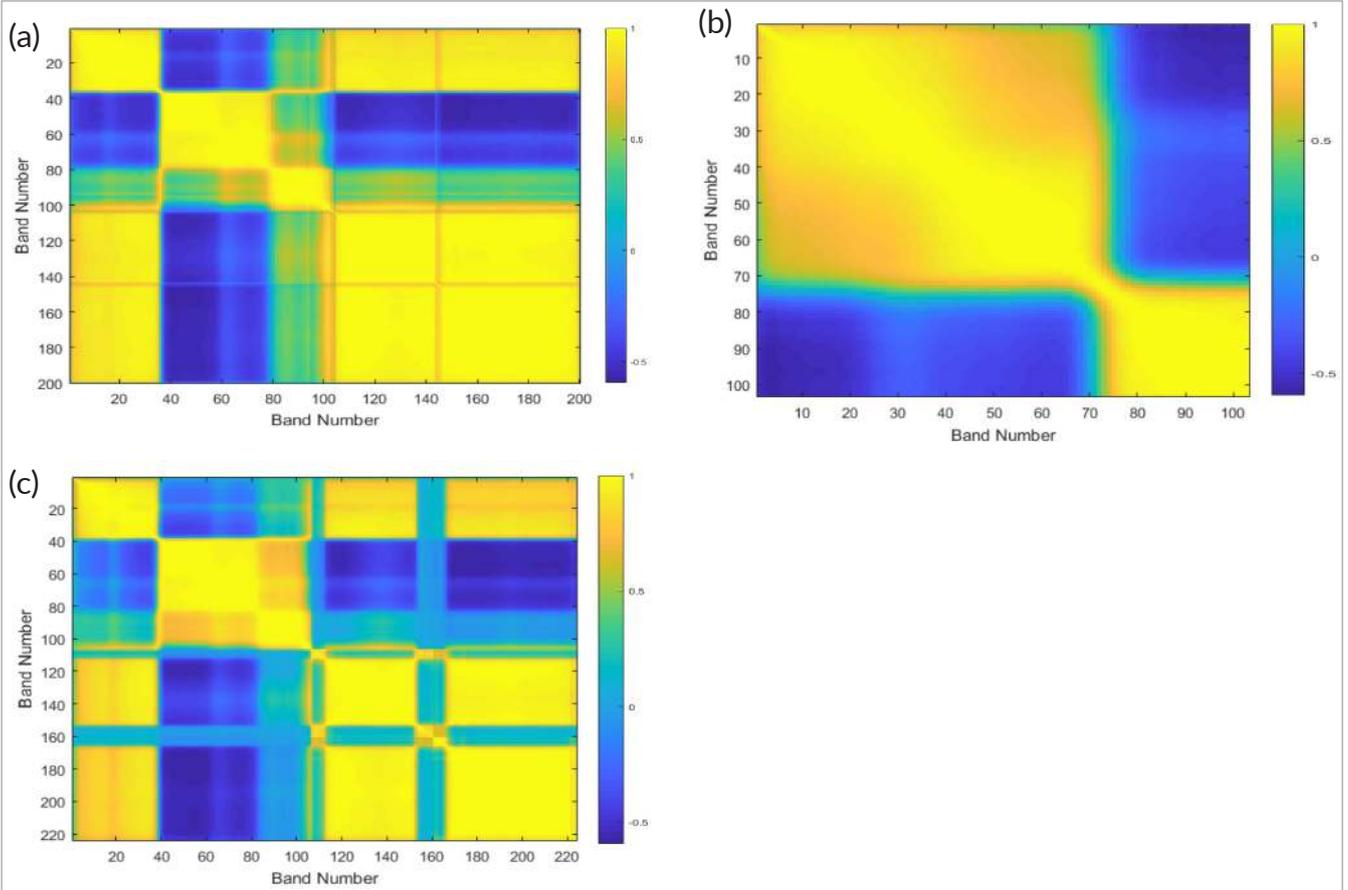Therefore, the selected bands will be less redundant and less correlated as only representative bands will be

Figure 1. (a) Correlation image of Indian Pines scene; Yellow = ±1, Blue = 0. (b) The spectral signature for various classes of Indian Pines scene. (c) Correlation image of Pavia University scene; Yellow = ±1, Blue = 0.

picked from each cluster. In this study, the overall classification accuracy of the classifier is used as an objective function f and the value of the objective function is treated as the fitness value. Classification accuracy is defined as follows:

$$f = Classification\,Accuracy = \frac{\sum_{i=1}^{total\,number\,of\,pixels} evaluate(p_i)}{total\,number\,of\,pixels}$$

(7)

where assess(p_i) is the function used to classify pixels p_i.

$$evaluate(p_i) = \begin{cases} 1 & if\ classify(p_i) = \Omega \\ 0 & otherwise \end{cases}$$

(8)

classify(p_i) is the function that gives the class of p_i. For the pixel p_i with true class, the function assess(p_i) = 1 and 0 otherwise.

The pseudo-code of the modified CS is presented as Algorithm 2.

Algorithm 2: Band selection using modified CS algorithm

Input: Correlation-based band set b_i, max_iteration T and number of nests n

Output: Selected band set

Step 1: Initialise population for each host n nests;
Initialise correlation-based band set which satisfies the condition in Equation (5),
$b_i = \{b_1, b_2,..., b_m\}$;
Choose the current best nest by evaluating the objective function f_i using Equation (7)

Step 2: **while** ($t < T$)
Get a cuckoo egg arbitrarily via Lévy flights
Evaluate the objective function f_j using Equation (7)
Select a nest from n random nests
If ($f_i > f_j$)
Replace j with the new solution
Select worst nests according to probability P
Replace them with new solution by Lévy flights

Preserve the best solutions
Rank the solutions and determine the current best solution
end while

Results and discussion

In this section, each hyperspectral dataset is analysed with the following comparisons: classification results, analysis of the number of selected bands and fitness curve analysis.

Classification results

In this subsection, classification results obtained for the three datasets are discussed.

ROSIS Dataset: Pavia University dataset
The proposed method is compared with the band selection method based on standard CS, GA and PSO. The proposed method, CSCI, uses a new initialisation strategy based on correlation to enhance the exploration ability of the standard CS algorithm. The obtained results are reported in Figure 2, which shows the

proposed approach attaining the best results among all competing approaches in terms of OA, AA and κ. The performance of CSCI is more prominent, probably because they select the most informative bands while reducing redundant information. This shows that the optimisation ability of the proposed approach is enhanced and reaches the best global optimum solution. The classification maps produced for all the competing methods on the Pavia University dataset are shown in Figure 3. The proposed approach shows obtaining a classification map smoother than the other competing methods.

AVIRIS dataset: Indian Pines dataset
The proposed method is compared with the band selection method based on standard CS, GA and PSO. The proposed method, CSCI, uses a new initialisation strategy based on correlation to enhance the exploration ability of the standard CS algorithm. The obtained results are reported in Figure 4 and classification maps are depicted in Figure 5.

As shown in Figure 4, the CS algorithm shows better performance than GA and PSO, whereas the results obtained by the proposed CSCI approach are better than
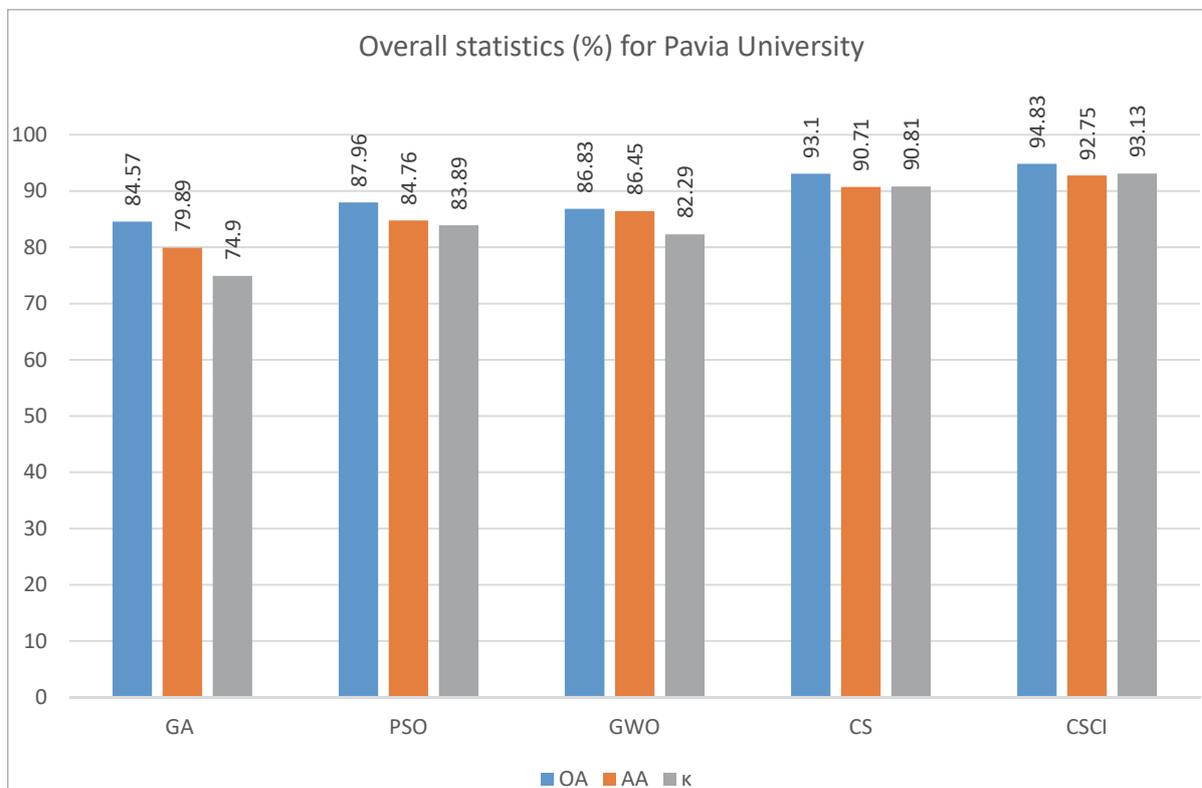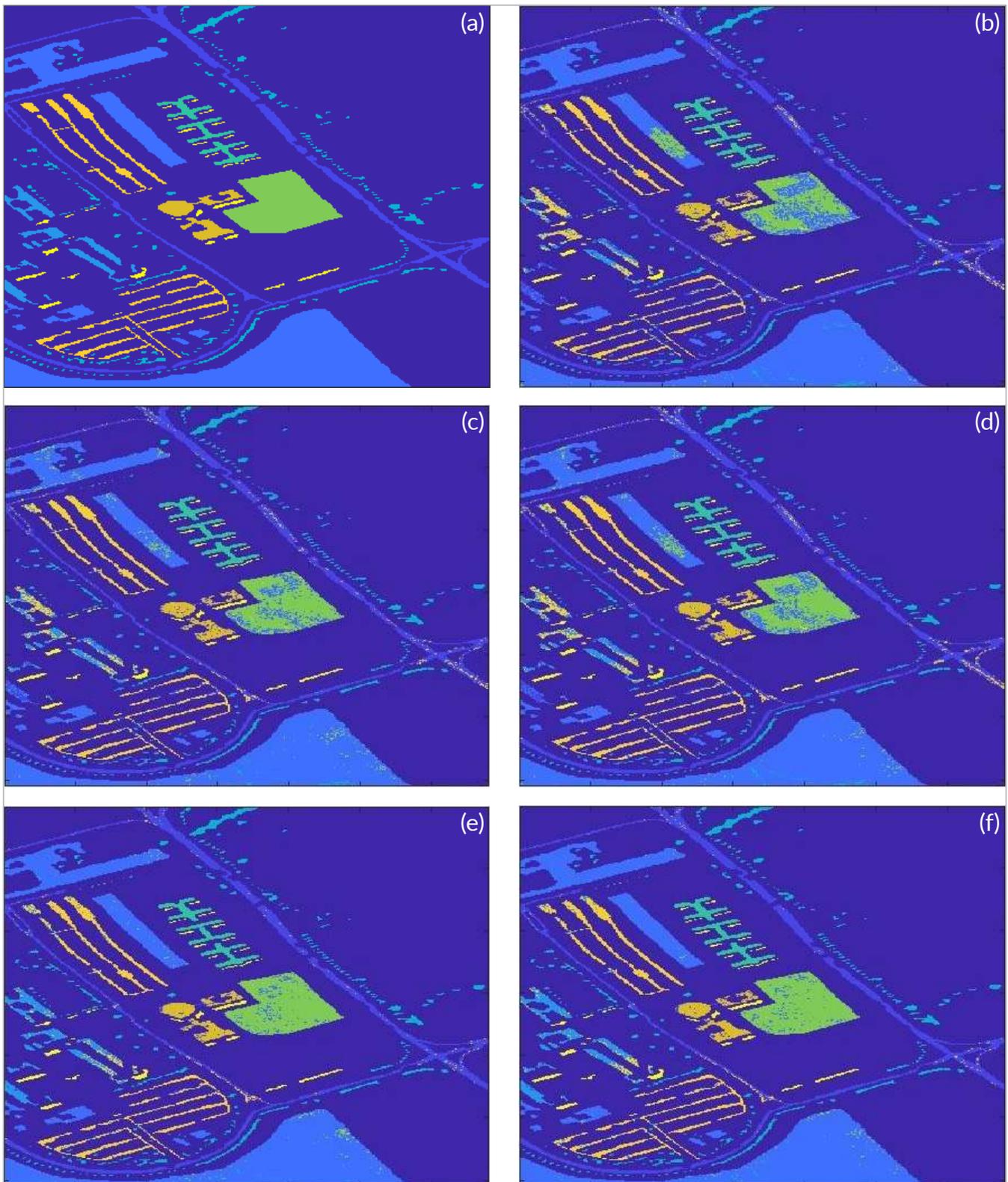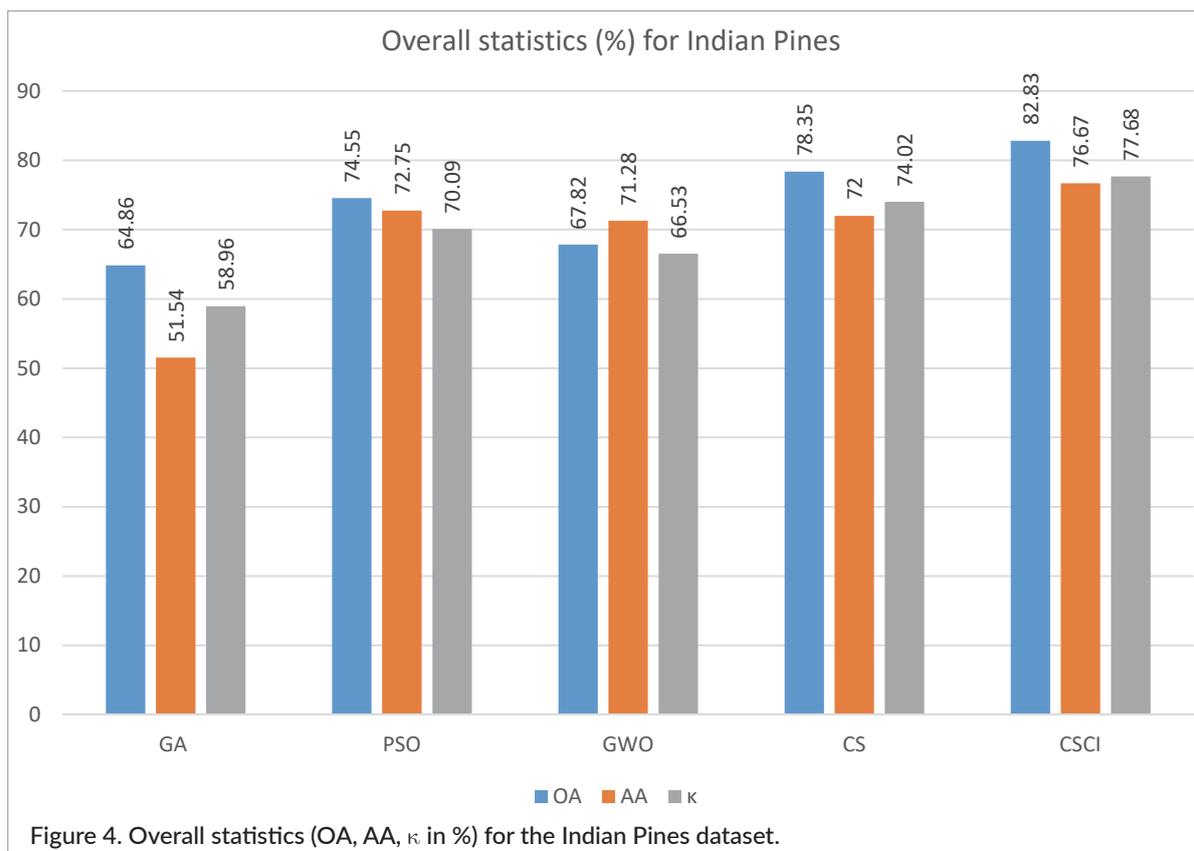

Figure 2. Overall statistics (OA, AA, κ in %) for Pavia University dataset.

Figure 3. Classification map obtained by various methods for Pavia University dataset. (a) Ground truth, (b) GA, (c) PSO, (d) GWO, (e) CS and (f) CSCI.

Figure 4. Overall statistics (OA, AA, κ in %) for the Indian Pines dataset.

the standard CS approach. By selecting only 20 % of samples per class, the proposed method obtains an OA of about 82.83 % and AA of 76.67 %. Compared to other competing approaches, the proposed approach improves the classification accuracy significantly, as shown in Figure 4.

AVIRIS dataset: Salinas dataset
The average classification results obtained by all approaches are summarised in Figure 6. The classification maps for all six band selection approaches are shown in Figure 7.

Figures 6 and 7 show that the proposed approach achieves the highest classification performance in terms of OA, AA, class-wise accuracy as well as κ. The spatial resolution of the Salinas data is 3.7 m per pixel, therefore, the classification accuracies are generally higher than the other two data sets. It is seen that the OA and AA obtained by all five approaches exceed 88 %, since the ground truth in Salinas is relatively easy to classify. The CSCI approach achieves the best classification performance.

Influence of the number of band clusters on classification results

In the proposed method, representative bands are chosen from each band cluster as discussed above. Accordingly, the number of representative bands is the number of clusters. To show the convincing result of the proposed method, the selected number of bands is set to 20. The reason behind this particular number is that our main purpose is to use an optimal number of bands with plenty of information, which reduces the burden of the subsequent hyperspectral image processing tasks, for example, image segmentation or image classification. So, the number of bands must be as low as possible. The band selection process is very effective if only a few bands are selected, however, the classification performance is only satisfactory or acceptable. We can see from Figure 8 that when a lower number of clusters are selected, i.e. $m = 5$, the overall performance of the proposed technique is not satisfactory. This is because when there are fewer clusters, the hyperspectral bands in each contain more complementary information and the
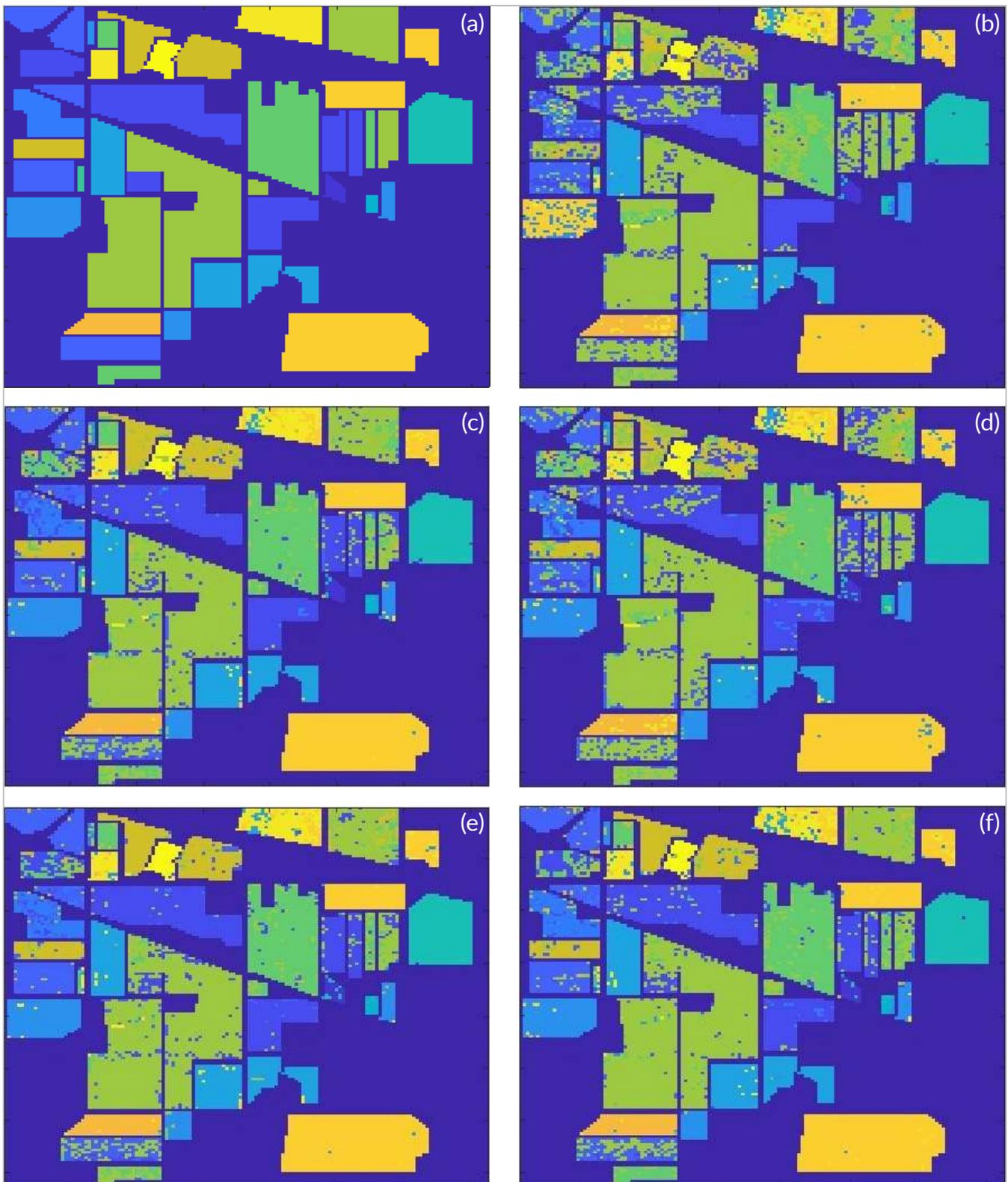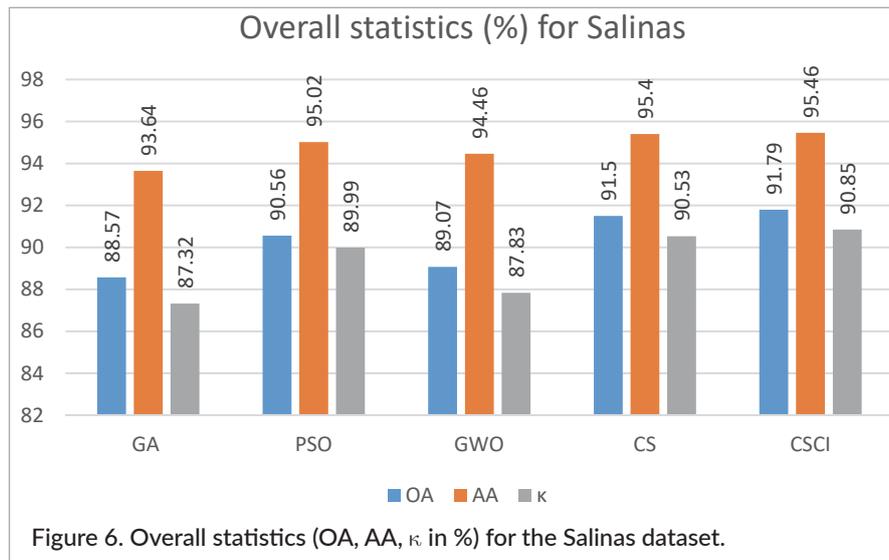
Figure 5. Classification map obtained by various methods for the Indian Pines dataset. (a) Ground truth, (b) GA, (c) PSO, (d) GWO, (e) CS and (f) CSCI.

Figure 6. Overall statistics (OA, AA, κ in %) for the Salinas dataset.

useful discriminative information will be lost in the fusion process. The curves of the proposed method rise abruptly with the number of clusters in the range of 5–25. Then, the curve shows a sluggish escalation when the number of clusters exceeds 25. Lastly, the graph is stable after the number of clusters exceeds 40. The reason is that when there are fewer band clusters, the hyperspectral bands in each subgroup contain a greater amount of useful discriminative information and, during the averaging operation, that useful information will be lost. Alternatively, when more band clusters are considered, band clusters show a very high redundancy and barely offer additional information for better classification.

Analysis of bands

Figure 9 shows the selected bands by CS and CSCI approaches for Indian Pines, Pavia University and Salinas datasets. In Figure 9(a,b), the 16 categories of the Indian Pines image are represented by 16 curves in different colours. In Figure 9(c,d), the nine categories of the Pavia University image are represented by nine curves in different colours. Similarly, in Figure 9(e,f), the 16 categories of the Salinas image are represented by 16 curves in different colours. Each curve is a reflectance spectrum showing the reflectance of this category across a range of wavelengths. In Figure 9, the 20 bands selected by the CS and CSCI approaches are represented by dotted lines. It can be seen that the bands are spread further apart by the CSCI approach [Figure 9(b,d,f)]. In hyperspectral data, spectral bands

are contiguous. Hence, the more the bands are spread out, the better will the different categories be discriminated. The standard CS algorithm uses a random number to initiate the location of nests. Sometimes the locations of these nests are inappropriately dispersed in a defined area and sometimes the location of these nests will be identical. Therefore, it causes repeated calculations and there is a high chance that it may fall into a local optimal solution. However, CSCI uses correlation-based initialisation instead of random initialisation and reaches the global best solution. This is why the bands selected by the CS algorithm are not the same as those selected by the CSCI algorithm.

Fitness curve analysis

Figure 10 shows the convergence rate as a function of the fitness value of the CS and CSCI approaches on the Indian Pines, Pavia University and Salinas datasets. The convergence behaviour is determined by the plot of the number of iterations vs the fitness value. The fitness value of the solution is calculated using the classification accuracy. The performance of CSCI is superior to that of the CS algorithm. With the same number of iterations, the classification accuracy of CSCI is superior to that of the CS algorithm. Although CS possesses a global search ability, the CS algorithm suffers from loss of population diversity as a result of random initialisation of the nest location. Therefore, CS fails to obtain a global optimum solution and gets stuck into a local optimum solution. On the other hand, the CSCI algorithm utilises maximum iterations and reaches the global best solution. The
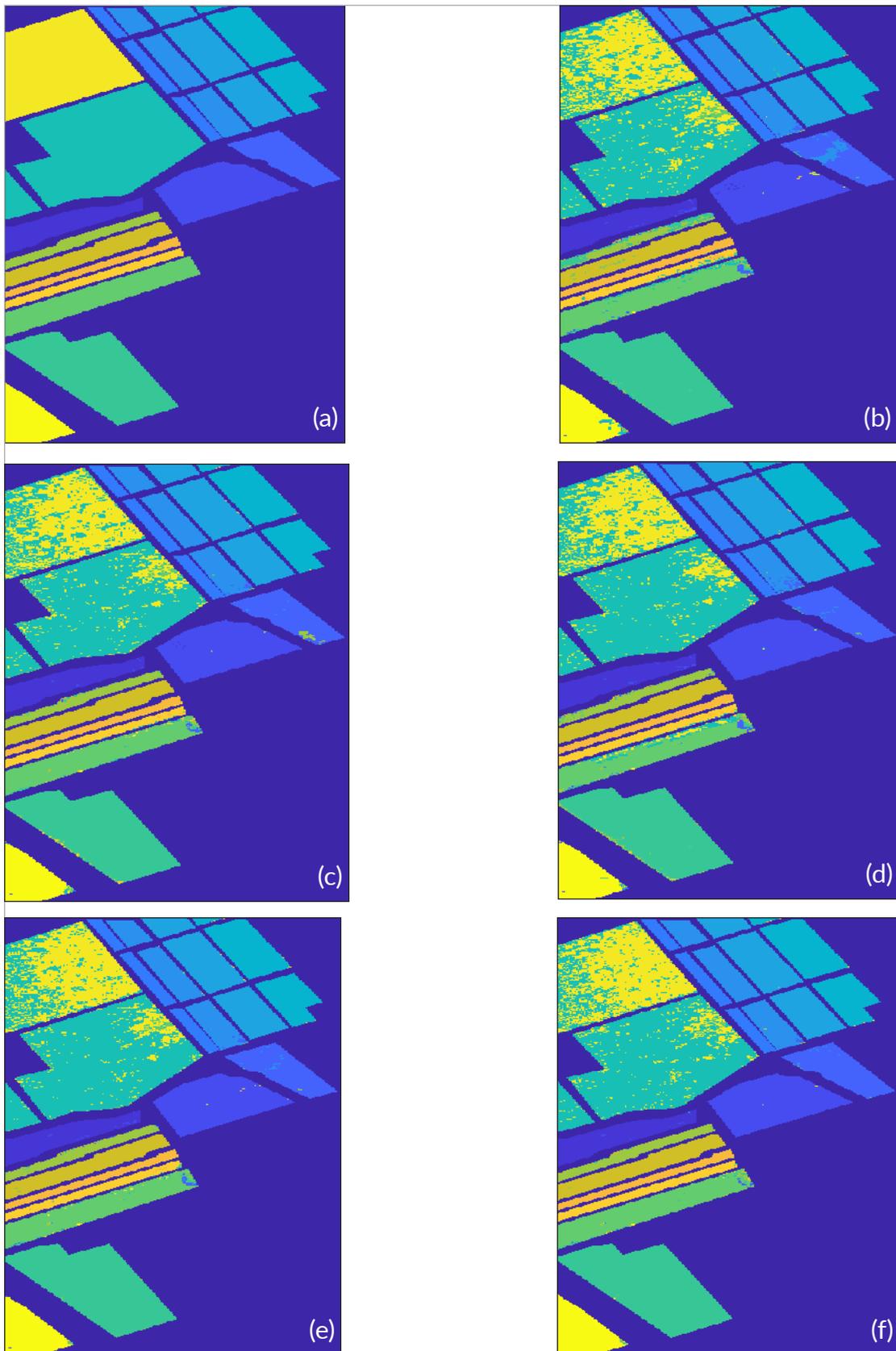
Figure 7. Classification map obtained by various methods for the Salinas dataset. (a) Ground truth, (b) GA, (c) PSO, (d) GWO, (e) CS and (f) CSCI.
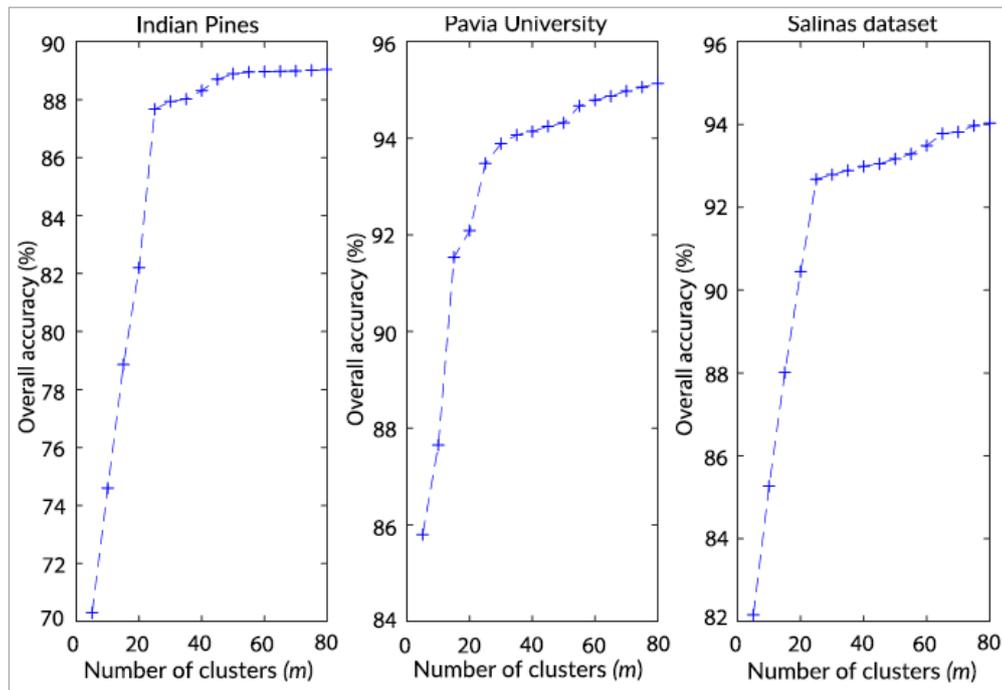
Figure 8. Influence of number of clusters on OA for (a) Indian Pines, (b) Pavia University and (c) Salinas scene.
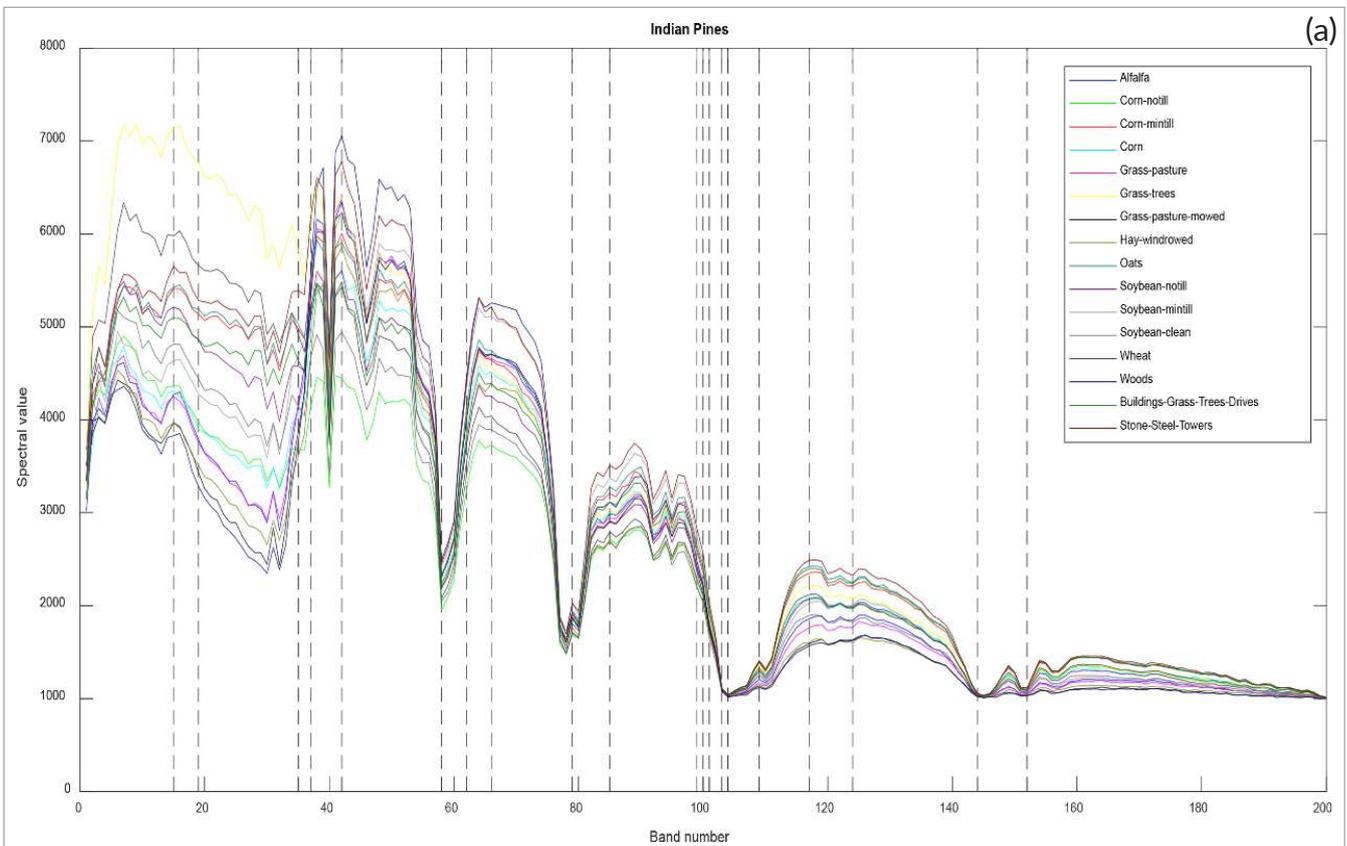

Figure 9. Spectral response of different land cover classes of the dataset. Dotted lines represent the representative bands. (a) Bands selected by CS for the Indian Pines dataset.

Figure 9. Spectral response of different land cover classes of the dataset. Dotted lines represent the representative bands. (b) Bands selected by CSCI for the Indian Pines dataset. (c) Bands selected by CS for the Pavia University dataset.
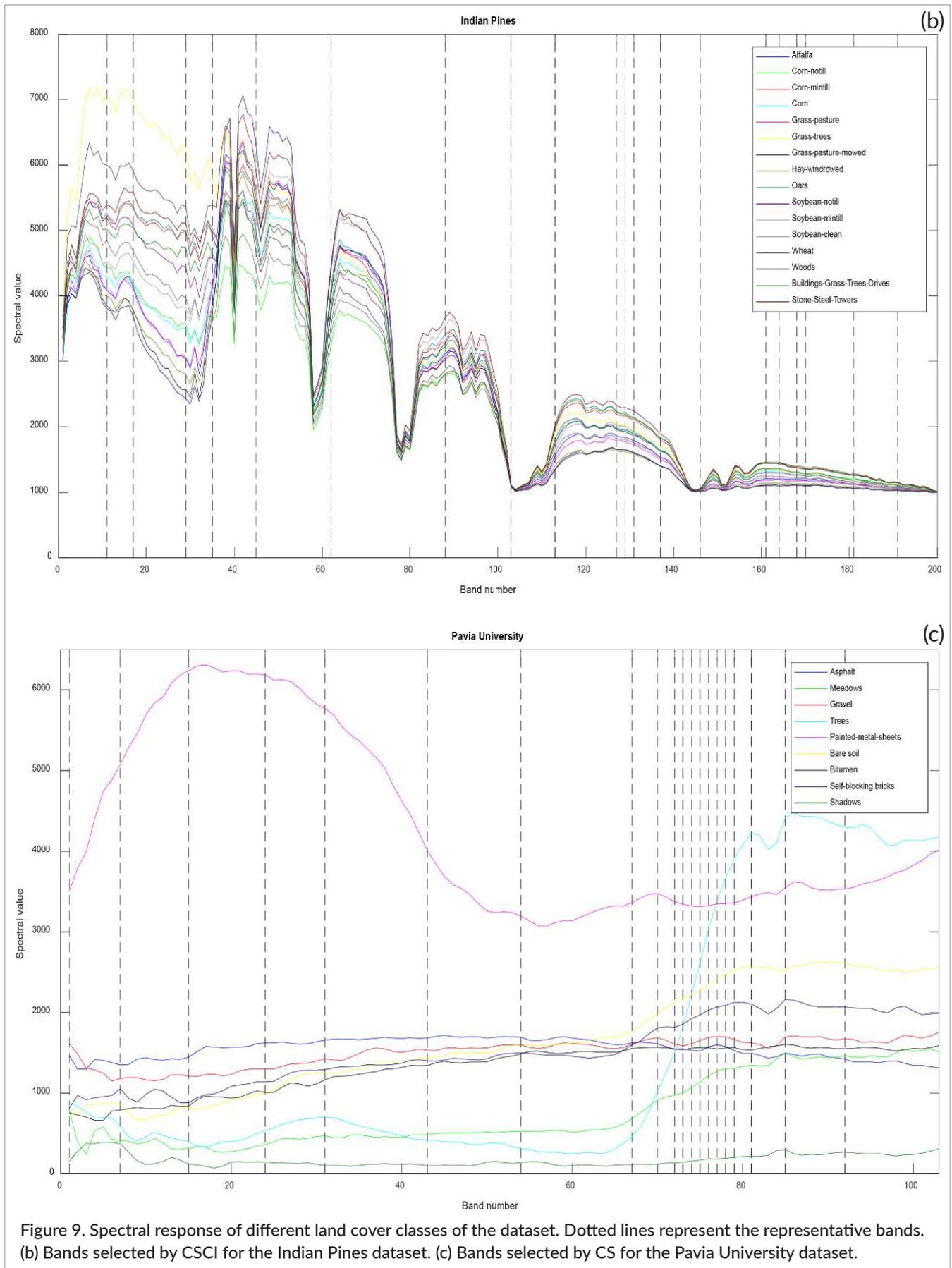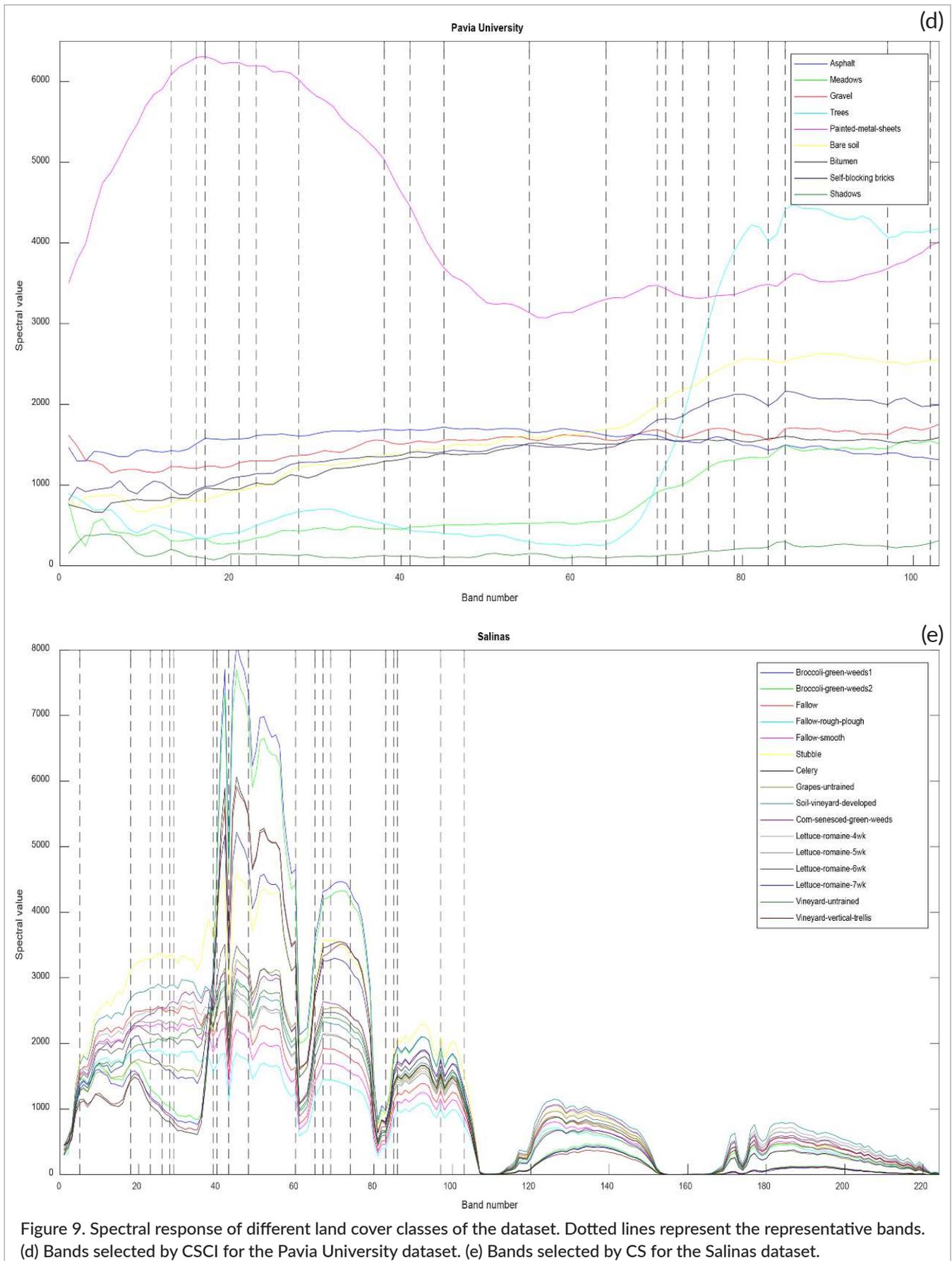
Figure 9. Spectral response of different land cover classes of the dataset. Dotted lines represent the representative bands. (d) Bands selected by CSCI for the Pavia University dataset. (e) Bands selected by CS for the Salinas dataset.
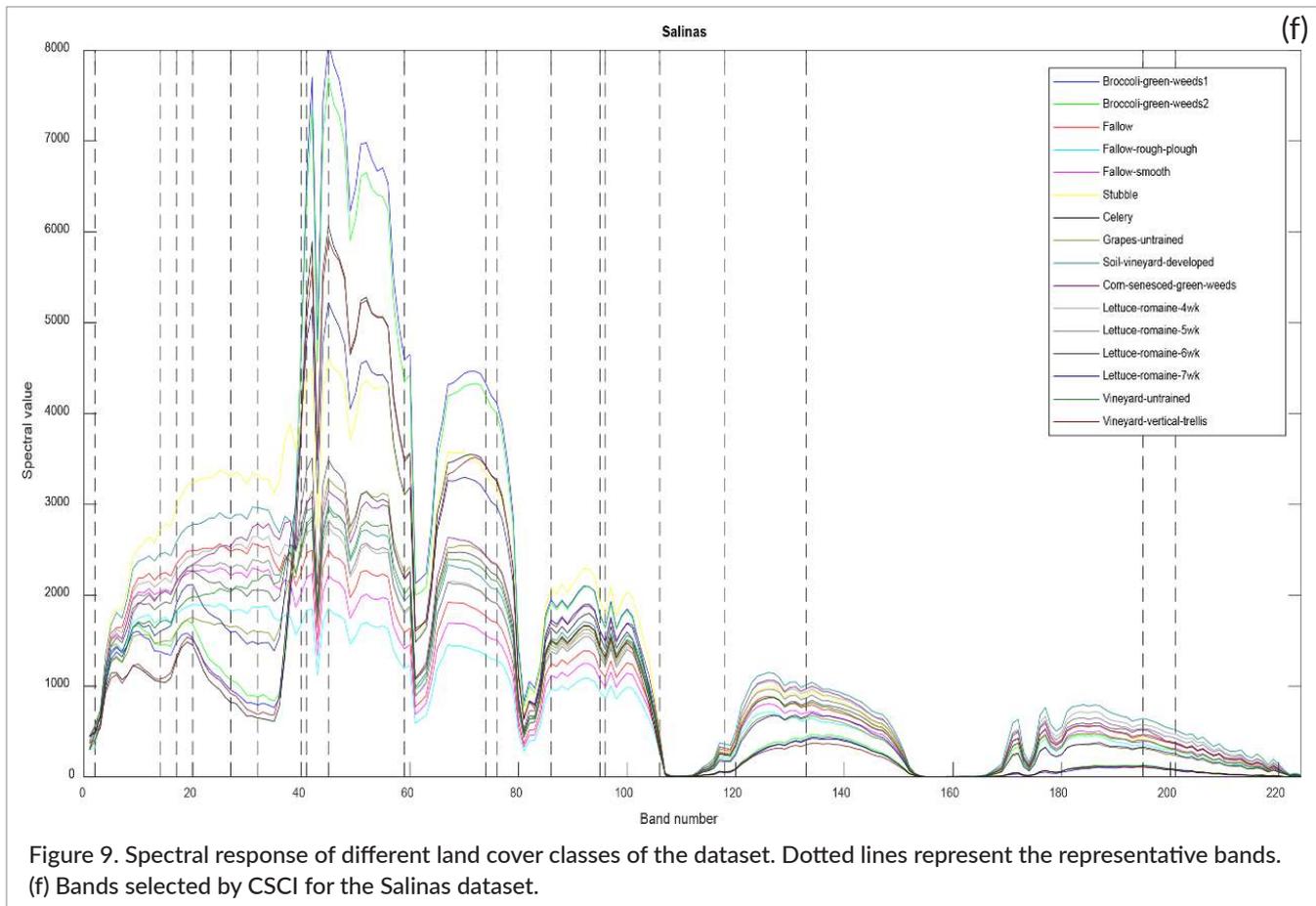
Figure 9. Spectral response of different land cover classes of the dataset. Dotted lines represent the representative bands. (f) Bands selected by CSCI for the Salinas dataset.

convergence speed of the CSCI approach is faster than the CS approach, because, in the middle of the iteration, the CS approach easily falls into a local optimum solution and needs time to jump out of that. This shows that the population initialisation strategy greatly affects the performance of the approach.

Conclusion and future work

Several studies in the past have shown that a diverse set of features need not come from a uniform random feature selection and a non-uniform strategy can be used instead. This is especially true with multi-classification systems. In this work, we studied one such approach (initialisation based on correlation) using band groups. The findings of this study will influence future research to study other metaheuristic band selection approaches under this framework. This study shows the merit of the proposed approach when tested with three benchmark hyperspectral datasets and compared against three

state-of-the-art feature selection algorithms. In this paper, we have proposed a modified CS algorithm for the band selection of hyperspectral images. The approach used an initialisation strategy based on correlation for the initial population instead of a random initialisation strategy. Correlation-based initialisation enhances the exploration ability of the standard CS algorithm and exhibits better performance for band selection of hyperspectral images. Furthermore, the CSCI algorithm can prevent itself from falling into a local optimum solution. The experimental results have been compared with other band selection techniques optimised with GA, PSO, GWO and standard CS algorithms. Among these algorithms, CSCI shows excellent optimisation performance with the ability to obtain the global optimal solution. For better comparison, analysis of selected bands and fitness curve analysis have also been performed. The experimental results on three standard datasets verify the robustness of the proposed approach in solving the problem of band selection.

Future work will focus on formulating the objective function to optimise the band selection problem
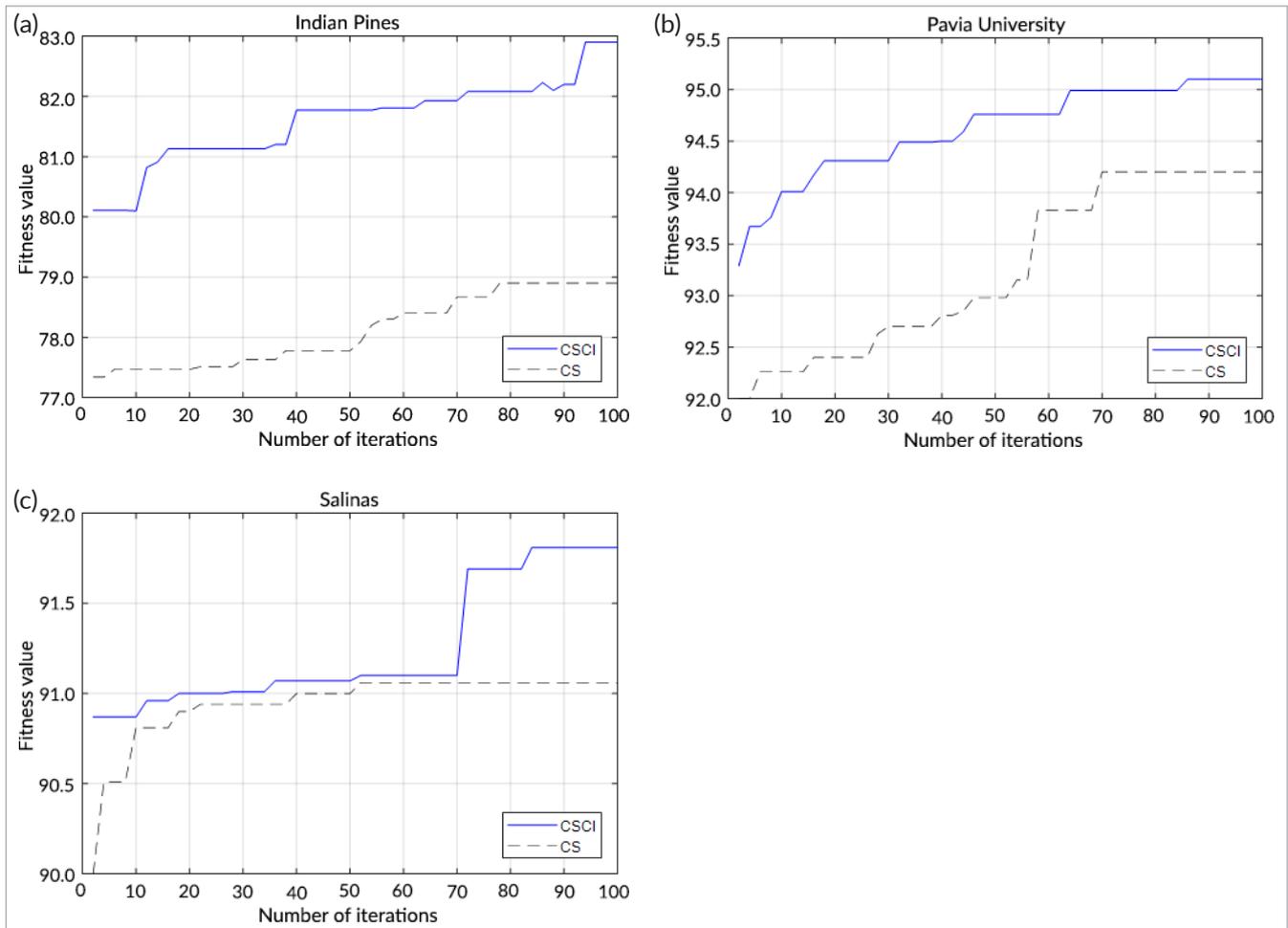
Figure 10. Convergence behaviour of the CS and CSCI algorithms on (a) Indian Pines, (b) Pavia University and (c) Salinas datasets.

to further improve the performance of hyperspectral imagery classification and also compare the final bands selected from different initialisation strategies such as entropy, mutual information and other distance measures along with the correlation-based initialisation strategy. This will help us to understand the mechanics of the different initialisation strategies. Also, the proposed method will be tested on a number of small and large datasets.

Acknowledgements

References

1. B.K. Mohan and A. Porwal, "Hyperspectral image processing and analysis", *Curr. Sci.* **108,** 833 (2015).
2. H. Pu, Z. Chen, B. Wang and G.M. Jiang, "A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery", in *IEEE Trans. Geosci. Remote Sens.* **52(11),** 7008 (2014). https://doi.org/10.1109/TGRS.2014.2306687
3. S.S. Sawant and M. Prabukumar, "A review on graph-based semi-supervised learning methods for hyperspectral image classification", *Egypt. J. Remote Sens. Sp. Sci.* in press (2018). https://doi.org/10.1016/j.ejrs.2018.11.001
4. L.N.P. Boggavarapu and M. Prabukumar, "Survey on classification methods for hyper spectral remote sensing imagery", in *2017 International Conference on Intelligent Computing and Control Systems*

(ICICCS), Madurai, pp. 538–542 (2017). https://doi.org/10.1109/ICCONS.2017.8250520

5. H. Huang, F. Luo, J. Liu and Y. Yang, "Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding", *ISPRS J. Photogramm. Remote Sens.* **106,** 42 (2015). https://doi.org/10.1016/j.isprsjprs.2015.04.015

6. A.L.I. Moghimi, C.E. Yang and P.M. Marchetto, "Ensemble feature selection for plant phenotyping: a journey from hyperspectral to multispectral imaging", *IEEE Access* **6,** 56870 (2018). https://doi.org/10.1109/ACCESS.2018.2872801

7. S. Samiappan, S. Prasad and L.M. Bruce, "Non-uniform random feature selection and kernel density scoring with SVM based ensemble classification for hyperspectral image analysis", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6(2),** 792 (2013). https://doi.org/10.1109/JSTARS.2013.2237757

8. S.S. Sawant, M. Prabukumar and S. Samiappan, "A band selection method for hyperspectral image classification based on Cuckoo Search algorithm with correlation based initialization", in *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS),* Amsterdam, Netherlands, pp. 1–4 (2019). https://doi.org/10.1109/WHISPERS.2019.8920950

9. M. Prabukumar, S. Sawant, S. Samiappan and L. Agilandeeswari, "Three-dimensional discrete cosine transform-based feature extraction for hyperspectral image classification", *J. Appl. Remote Sens.* **12(4),** 046010 (2018). https://doi.org/10.1117/1.JRS.12.046010

10. J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li and S. Marshall, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging", *IEEE Trans. Geosci. Remote Sens.* **53(8),** 4418 (2015). https://doi.org/10.1109/TGRS.2015.2398468

11. M. Prabukumar and S. Shrutika, "Band clustering using expectation–maximization algorithm and weighted average fusion-based feature extraction for hyperspectral image classification", *J. Appl. Remote Sens.* **12(4),** 046015 (2018). https://doi.org/10.1117/1.JRS.12.046015

12. R. Vaddi and M. Prabukumar, "Comparative study of feature extraction techniques for hyper spectral remote sensing image classification: A survey", in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS),* Madurai, pp. 543–548 (2017). https://doi.org/10.1109/ICCONS.2017.8250521

13. L.N.P. Boggavarapu and P. Manoharan, "Classification of hyper spectral remote sensing imagery using intrinsic parameter estimation", in *Intelligent Systems Design and Applications. 2018. Advances in Intelligent Systems and Computing,* Ed by A. Abraham, A. Cherukuri, P. Melin and N. Gandhi. Springer (2020). https://doi.org/10.1007/978-3-030-16660-1_83

14. R. Vaddi and P. Manoharan, "Probabilistic PCA based hyper spectral image classification for remote sensing applications", in *Intelligent Systems Design and Applications. ISDA 2018 2018. Advances in Intelligent Systems and Computing,* Ed by A. Abraham, A. Cherukuri, P. Melin and N. Gandhi. Springer (2020). https://doi.org/10.1007/978-3-030-16660-1_84

15. L.N.P. Boggavarapu and M. Prabukumar, "Hyperspectral image classification fuzzy embedded hyperbolic sigmoid nonlinear principal component and weighted least square approach", *J. Appl. Remote Sens.* **14(2),** 0245011 (2020). https://doi.org/10.1117/1.JRS.14.024501

16. R. Vaddi and P. Manoharan, "Hyperspectral image classification using CNN with spectral and spatial features integration", *Infrared Phys. Technol.* **107,** 103296 (2020). https://doi.org/10.1016/j.infrared.2020.103296

17. B. Guo, R.I. Damper, S.R. Gunn and J.D.B. Nelson, "Improving hyperspectral band selection by constructing an estimated reference map", *J. Appl. Remote Sens.* **8(1),** 083692 (2014). https://doi.org/10.1117/1.JRS.8.083692

18. H. Yang, Q. Du, H. Su and Y. Sheng, "An efficient method for supervised hyperspectral band selection", *IEEE Geosci. Remote Sens. Lett.* **8(1),** 138 (2011). https://doi.org/10.1109/LGRS.2010.2053516

19. J. Feng, L. Jiao, F. Liu, T. Sun and X. Zhang, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images", *Pattern Recognit.* **51,** 295 (2016). https://doi.org/10.1016/j.patcog.2015.08.018

20. X. Luo, Z. Shen, R. Xue and H. Wan, "Unsupervised band selection method based on importance-assisted column subset selection", *IEEE Access* **7,** 517 (2019). https://doi.org/10.1109/ACCESS.2018.2885545

21. S.S. Sawant and P. Manoharan, "Unsupervised band selection based on weighted information entropy and 3D discrete cosine transform for hyperspectral image classification", *Int. J. Remote Sens.* **41(10),** 3948 (2020). https://doi.org/10.1080/01431161.2019.1711242

22. Y. Yuan, J. Lin and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis", *IEEE Trans. Geosci. Remote Sens.* **54(3),** 1431 (2016). https://doi.org/10.1109/TGRS.2015.2480866

23. H. Zhai, H. Zhang, L. Zhang and P. Li, "Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection", *IEEE Trans. Geosci. Remote Sens.* **57(3),** 1723 (2019). https://doi.org/10.1109/TGRS.2018.2868796

24. L. Xie, G. Li, L. Peng, Q. Chen, Y. Tan and M. Xiao, "Band selection algorithm based on information entropy for hyperspectral image classification", *J. Appl. Remote Sens.* **11(2),** 026018 (2017). https://doi.org/10.1117/1.JRS.11.026018

25. S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection", *IEEE Trans. Geosci. Remote Sens.* **54(1),** 88 (2016). https://doi.org/10.1109/TGRS.2015.2450759

26. F. Xie, F. Li, C. Lei, J. Yang and Y. Zhang, "Unsupervised band selection based on artificial bee colony algorithm for hyperspectral image classification", *Appl. Soft Comput.* **75,** 428 (2019). https://doi.org/10.1016/j.asoc.2018.11.014

27. P. Ghamisi and J.A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization", *IEEE Geosci. Remote Sens. Lett.* **12(2),** 309 (2015). https://doi.org/10.1109/LGRS.2014.2337320

28. H. Su, Q. Du, G. Chen and P. Du, "Optimized hyperspectral band selection using particle swarm optimization", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7(6),** 2659 (2014). https://doi.org/10.1109/JSTARS.2014.2312539

29. S.A. Medjahed, T. Ait Saadi, A. Benyettou and M. Ouali, "Gray Wolf Optimizer for hyperspectral band selection", *Appl. Soft Comput.* **40,** 178 (2016). https://doi.org/10.1016/j.asoc.2015.09.045

30. S.S. Sawant and P. Manoharan, "New framework for hyperspectral band selection using modified wind-driven optimization algorithm", *Int. J. Remote Sens.* **40(20),** 7852 (2019). https://doi.org/10.1080/01431161.2019.1607609

31. S.A. Medjahed, T.A. Saadi, A. Benyettou and M. Ouali, "Binary cuckoo search algorithm for band selection in hyperspectral image classification", *IAENG Int. J. Comput. Sci.* **42,** 1 (2015). http://www.iaeng.org/IJCS/issues_v42/issue_3/IJCS_42_3_03.pdf

32. X. Yang and S. Deb, "Multiobjective cuckoo search for design optimization", *Comput. Oper. Res.* **40(6),** 1616 (2013). https://doi.org/10.1016/j.cor.2011.09.026

33. D. Rodrigues, "BCS: A Binary Cuckoo Search algorithm for feature selection", in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, Beijing, pp. 465–468 (2013). https://doi.org/10.1109/ISCAS.2013.6571881

34. S.S. Sawant and P. Manoharan, "Hyperspectral band selection based on metaheuristic optimization approach", *Infrared Phys. Technol.* **107,** 103295 (2020). https://doi.org/10.1016/j.infrared.2020.103295

35. S. Suresh, S. Lal, C.S. Reddy and M.S. Kiran, "A novel adaptive Cuckoo Search algorithm for contrast enhancement of satellite images", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10(8),** 3665 (2017). https://doi.org/10.1109/JSTARS.2017.2699200

36. M. Shehab, A.T. Khader and M.A. Al-Betar, "A survey on applications and variants of the cuckoo search algorithm", *Appl. Soft Comput.* **61,** 1041 (2017). https://doi.org/10.1016/j.asoc.2017.02.034

37. X. Chen and K. Yu, "Hybridizing cuckoo search algorithm with biogeography-based optimization for estimating photovoltaic model parameters", *Sol. Energy* **180,** 192 (2019). https://doi.org/10.1016/j.solener.2019.01.025

38. U. Mlakar, I. Fister and I. Fister, "Hybrid self-adaptive cuckoo search for global optimization", *Swarm Evol. Comput.* **29,** 47 (2016). https://doi.org/10.1016/j.swevo.2016.03.001

39. L. Liu, X. Liu, N. Wang and P. Zou, "Modified cuckoo search algorithm with variational parameters and logistic map", *Algorithms* **11(3),** 30 (2018). https://doi.org/10.3390/a11030030

40. M. Mareli and B. Twala, "An adaptive cuckoo search algorithm for optimisation", *Appl. Comput. Inform.* **14(2),** 107 (2018). https://doi.org/10.1016/j.aci.2017.09.001

41. D.H. Wolpert and W.G. Macready, "No free lunch theorems for optimization", *IEEE Trans. Evol. Comput.* **1(1),** 67 (1997). https://doi.org/10.1109/4235.585893

42. X. Yang and Suash Deb, "Cuckoo Search via Lévy flights", in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, Coimbatore, pp. 210–214 (2009). https://doi.org/10.1109/NABIC.2009.5393690

43. M. Graña, M.A. Veganzons and B. Ayerdi, "Hyperspectral Remote Sensing Scenes" [Online]. Available: http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

44. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines", *IEEE Trans. Geosci. Remote Sens.* **42(8),** 1778 (2004). https://doi.org/10.1109/TGRS.2004.831865