CrossMark

ORIGINAL RESEARCH

# Optimal band selection for hyperspectral data with improved differential evolution

Xiang Li[1] · Ge Wang[1]

**Abstract** The optimal band selection has been a hot research topic and one of the difficult problems in the field of hyperspectral remote sensing. To address the issue of hyperspectral data and improve the data processing speed, we propose a new method to improve differential evolution algorithm. We use the ENVI software on the pretreatment of original spectral data and dividing subspace. The experiments are also conducted with the improved DE algorithm. In experiments, different evaluation criteria of band selection are taken as the fitness function value of the DE algorithm. By calculating the optimal band combination in different dimensions of size, we can obtain the optimal selection band of hyperspectral data in this paper. Experiments show that the improved differential evolution algorithm for hyperspectral data dimensionality reduction model not only obtains the optimal band combination, but also greatly improves the classification accuracy.

**Keywords** Hyperspectral · Differential evolution · Band selection space division · Dimensionality reduction

## 1 Introduction

In the last years, the cloud services market experienced an extremely rapid growth, as reported in several market research reports that may lead to severe scalability problems (Christian et al. 2013). In recent decades, digital images are becoming increasingly important in healthcare environment, since they are more and more used in medical and medical-related applications (Castiglione et al. 2015). With the development of the remote-sensing imaging technology, the use of hyperspectral image is becoming more and more widespread. Recently, ambient intelligence for remote sensing images has obtained more attention (Benavente-Peces et al. 2014). Cerezo et al. (2012) dealt with the facial emotion recognition and multimodal fusion of affective information coming from different channels. Due to the dense sampling of spectral signatures of land covers, hyperspectral images have better discrimination among similar ground cover classes than traditional multispectral scanners. At the same time, these images are usually composed of tens of or hundreds of spectral bands with high redundancy and great amount of computation in perspectral image classification. Therefore the most important and urgent issue is how to reduce the number of those bands largely with little loss of information or classification accuracy (Wu et al. 2010). Band selection for hyperspectral image is the process to reduce the band size and identify the most informative bands or further analysis on the hyperspectral image data. Like the feature selection problem, band selection is NP-hard with only explication enumeration approaches known to solve it (Feng et al. 2014). Bands selection techniques generally involve both searching algorithm and defining the bands range. The search algorithm generates some useful features by taking combinations of the bands, those features are reasonably located in the highdimensional feature sets, which reduced the dimensionality on a large scale. Then define the bands range according to the generated features methods (Liu et al. 2010). The optimal band selection of hyperspectral data is an important method to overcome the problem of massive hyperspectral data and improve its processing

✉ Xiang Li
  lixiang@cug.edu.cn; 53222726@qq.com

  Ge Wang
  13545863689@163.com

[1]  School of Computer Science, China University of Geosciences, Wuhan 430074, China

Springer

speed. In the fourth part of section three in this paper, the enumeration method is adopted to get the best three band selection in the studying area of the sampling, according to each evaluation method, the optimal three band combination of the evaluation index is given respectively, so that it can obtain the hyperspectral data imaging in three band and be convenient for visual interpretation in the process of practical application (Huang and He 2005). But there are serious drawbacks on visual interpretation in the process of practical application because of the huge amount of hyperspectral data, and only three band selection doesn't necessarily get more effective classification for hyperspectral data, so this article uses the intelligent algorithm to band selection of hyperspectral data. Differential evolution (DE) algorithm has a good advantage of solving combinatorial optimization problems, therefore, the mentioned algorithm can fully applied to the study of hyperspectral data dimension reduction. This paper puts forward an hyperspectral data dimension reduction model which is based on improved DE algorithm: First, the ENVI software is applied to the preprocessing of original hyperspectral data (including atmospheric correction, radiation correction and geometry correction, etc.) and subspace partition. Second, the idea of Yin-Yang initialization population is introduced to improve the DE algorithm, and then we respectively treat the proposed band selection method suggested in the fourth part of section three as the fitness function value of the DE algorithm to compute the processed data, so we get the band combination of different dimensional sizes. Finally, we obtain the classification accuracy after measuring the stand or fall of band combination though.

In this paper, the main content of the research is a dimension reduction method used in hyperspectral remote sensing image processing. Dimension reduction can be viewed as a transformation from a high order dimension to a low order dimension (Kitti et al. 2012). Firstly, we apply the ENVI software in the original hyperspectral data preprocessing (including atmospheric correction, radiation correction and geometry correction, etc.), so as to get the statistical data of each image and category in different band. And then, dimension reduction model is applied to conduct the experiment over the data and obtain the optimal band combination which can effectively distinguish the main features in the image, and realize the dimension reduction of hyperspectral data.

Our motivation in this paper is use a simple but efficient method to solve a typical combinatorial optimization problem. DE as a very effective novel evolutional algorithm in global search, its genetic operation not only guarantee the results would not fall into local optimal, but also enhance the populations diversity, that it to say, we can search all the possible bands, meanwhile, get the best band combination.

## 2 The principle and improvement of DE algorithm

Differential evolution (DE) algorithm, proposed by Storn and Price (1997), is a simple, yet powerful population-based stochastic search technique for solving global optimization problems in continuous search domain. DE generates offspring by perturbing the solutions with a scaled difference of selected population vectors, and employs parent offspring comparison which allows replacement of a parent only if the offspring is better or equivalent to its parent (Zhao et al. 2011). Recently, DE has become one of the most widely used methods for handling global optimization problems. Furthermore, DE has been successfully applied in many science and engineering fields, such as pattern recognition, signal processing, satellite communications, vehicle routing problem, and so on (Peng et al. 2015). It adopts the real-code, the basic thought of DE is to obtain the intermediate population through recombinant by making use of the difference between individuals in population, and the new generation of population through the competition between parents and offsprings. Since there is no requirement for initial value, DE gets the characteristics of fast convergence, simple structure and robustness, etc (Zhang et al. 2010). DE algorithm uses the individual difference of the current population to generate intermediate population by reorganization, comparing the individual adaptive values of father with that hybrid of son to generate the next generation through competition. Compared with Genetic Algorithm, the most important part of DE algorithm is its mutation. The general process of DE algorithm is as follows: the first step is to select an individual. The variation of individual is produced by adding another two individual weighted differences. Then it is important to get the candidate individuals by Hybrid operation between individuals and parent individuals. The last step is to compare the adaptive values of the candidate individuals with that of the parent individuals to select the superiors into the next generation group. In this way, the algorithm uses differential variations, hybridization and selection to keep the group evolving ongoing until the termination conditions. At the beginning of the algorithm iteration, due to the large differences among the population individuals, the mutation makes the algorithm has a strong global searching ability. In the later period of iteration, the difference among population individuals is small, so the algorithm has stronger local search ability. DE algorithm has a good effect on solving function optimization problems such as high dimension, multi-peak, nonlinear, etc. The main characteristic of DE algorithm can be summarized as follows: (1) Simple structure; (2) High convergence rate; (3) Low probability of local optimum; (4) Strong robustness.

The algorithm is a kind of evolutionary algorithm which is used to solve the real optimization problems originally. At present, it is used to solve the discrete optimization problems and it is widely used in the areas of data mining, pattern recognition and artificial neural network, digital filtering, chemical, mechanical optimization design (Chen et al. 2009).

## 2.1 The basic operation of DE algorithm

DE algorithm is a kind of evolutionary algorithm based on real number encoding whose structural framework is similar to other evolutionary algorithms. For example the initialization of population, individual fitness evaluation, and the process from one generation to the next population should go through the operation mutation, crossover, selection and so on (Liu et al. 2007).

1. Determine the parameters
   Algorithms with suitable parameters can get good solution, while unsuitable parameters can only get bad solution. DE is also sensitive to the parameters setting. Choosing suitable parameter values is, frequently, a problem-dependent task and requires previous experience of the user (Yang et al. 2010). DE algorithm mainly involves the following four parameters: (1) population size N; (2) the individual dimensions D (namely, the length of the chromosome); (3) variants F; (4) crossover probability CR.
   In general, these parameters affect the algorithm to search the optimal solution and the convergence speed, so the setting of parameters is of great significance for the performance of the algorithm. The influence of each parameter on the algorithm performance can be briefly summarized as the following (Chen et al. 2010):

   1. The influence of population size N: A larger population is more diversified, which can increase the likelihood of search for the optimal solution, but at the same time, decline convergence rate; on the contrary, for the smaller group, the search space is less, speeding up the convergence speed of the algorithm, but it is easy to cause local convergence or stagnation of evolution algorithm.
   2. The influence of the individual dimensions D: when D value is small, the calculation process is simple but it is not conducive to individual evolution; When D value is large, though conducive to get the optimal solution of individual evolution, the calculation is complicated.
   3. The influence of variant F: Small F value will speed up the convergence, but it is easy to make the algorithm fall into local convergence, so we should avoid setting smaller F value; Larger F will

increase the possibility of the algorithm jumping out of local optimal solution, but when $F > 1$, the algorithm convergence rate will decline.
   4. The influence of crossover probability CR: we generally choose CR between zero and one, in general, the greater the CR, the faster convergence speed, but it is easy to fall into local optimum.
   Research showed that, population size N is generally between 5 to 10 D (determine the variable dimension of D based on the practical problem), variation factor F is generally between zero and two, generally take $F = 0.5$, crossover probability CR is generally between zero and one, generally take $CR = 0.3$. But when solving practical problems, we select the appropriate parameters after specific analysis.

2. Produce the initial population
   Randomly generated N chromosomes meet the constraint conditions in D dimension space, the specific process is as follows:

$$x_{ij} = rand(0,1) \times (x_j^U - x_j^L) + x_j^L (1 \leq i \leq N, 1 \leq j \leq D) \tag{1}$$

while $x_j^U$ and $x_j^L$ are respectively the upper and lower bounds of the first $j$ variable, rand(0, 1) return the random number in [0,1].

3. Mutation
   In the sight of individual variation selection, mutation operation involves the following three ways:

   1. The rand type (randomly selected individual variation): for individuals $x_{r1}(1 \leq r_1 \leq N)$ in the group, the new individual $x'_{r1}(1 \leq r_1 \leq N)$ generated meets the formula below:

$$x'_{r1}(1 \leq r_1 \leq N) = x_{r1}(1 \leq r_1 \leq N) + F \cdot (x_{r_2} - x_{r3}) \tag{2}$$

while $r_2, r_3 \in [1, N], r_1 \neq r_2 \neq r_3$, and $F > 0$ as the scale factor.
   2. The best type (select the optimal individual for variation): for the best individual $x_{best}(1 \leq r_1 \leq N)$ in the group, the new individual $x'_{r_1}(1 \leq r_1 \leq N)$ generated meet the formula below:

$$x'_{r_1} = x_{best} + F \cdot (x_{r_2} - x_{r_3}) \tag{3}$$

while $r_2, r_3 \in [1, N], r_1 \neq r_2 \neq r_3$, and $F > 0$ as the scale factor. $x_{best}$ was the best individual generated in current.
   3. The rand-to-best type (the compromise method): for individuals $x_{best}(1 \leq r_1 \leq N)$ in the group, the new individual $x'_{r_1}(1 \leq r_1 \leq N)$ generated meet the formula below:

$$x'_{r_1} = x_{r_1} = \lambda \cdot (x_{best} - x_{best}) + F \cdot (x_{r_2} - x_{r_3}) \quad (4)$$

while $r_2, r_3 \in [1, N], r_1 \neq r_2 \neq r_3$, and $F > 0$ as the scale factor. $x_{best}$ was the best individual generated currently, $\lambda$ was an added control variable set the same value with F generally. In view of the number of difference vector, it can also be subdivided into one weighted difference or two weighted difference. The above formula (2), (3) and (4) are the mutation in the case of one weighted difference. Below is the mutation in the case of two weight:

$$x'_{r_1} = x_{r_1} + F \cdot (x_{r_2} - x_{r_3}) + F \cdot (x_{r_4} - x_{r_5}) \quad (5)$$

while $r_2, r_3, r_4, r_5 \in [1, N], r_1 \neq r_2 \neq r_3 \neq r_4 \neq r_5$, and $F > 0$ as the scale factor.

4. The crossover operation

Crossover operation occurs among individuals $x_{r_1}(1 \leq r_1 \leq N)$ in the group and the new individuals $x'_{r_1}(1 \leq r_1 \leq N)$ generated by mutation, the offspring of candidate individuals produced by crossover operation between $x_{r_1}$ and $x'_{r_1}$, individual $x_{r_1}$ or $v$ determined to be reserved to the next generation by selecting operation (Alatas et al. 2008).

Crossover operation mainly has two forms: the bin type and the exp type, Fig. 1 clearly describes these two approaches and the differences between them. The cross of two approaches on a particular gene can be expressed as the following formula:

$$v_j = \begin{cases} x'_{r_1}, & j, \ rand(0,1) \leq CR \\ x_{r_1}, & j, \ rand(0,1) > CR \end{cases} \quad (6)$$

while $rand(0, 1)$ returns the random number in [0, 1], CR is the crossover probability $0 \leq CR \leq 1$.

The main difference between bin type and bxp type is that the each gene in the cross of bin type is independent. Namely, genes of individual $x_{r_1}$ and individual $x'_{r_1}$ are inherited to individual $v$ in a certain probability. But the exp type randomly chooses an initial point, continuing to produce the corresponding random decimal for each following gene. If the values are less than or equal to CR, inheriting the corresponding genes of individual $x'_{r_1}$ to individual $v$, and continuing to do the similar operation for the next gene, until the random decimal is greater than CR. The other genes are genetic from $x_{r_1}$. After concluding different mutation and crossover operations, we know it can form different patterns of DE algorithm, expressed as follows: DE/x/y/z.

Among them, $x$ is the way of choosing individual variation, it can be selected randomly, or choose the best one from the current generation. $y$ represents the number of difference vectors (one weighted difference or two weighted difference). $z$ represents the way of cross.

$x$ may have three kinds of circumstances: randbestrand-to-best. $y$ may have two kinds of circumstances: 1, 2. $z$ may have 3 kinds of circumstances: bin(binary)exp(exponential).

A variety of mutation and crossover patterns can be constituted after combining $x, y, z$, such as:

Scheme DE/rand/1/bin: $x'_{r_1} = x_{r_1} + F \cdot (x'_{r_2} - x'_{r_3})$

Scheme DE/best/2/bin: $x'_{r_1} = x_{best} + F \cdot (x'_{r_2} - x'_{r_3}) + F \cdot (x'_{r_4} - x'_{r_5})$

Scheme DE/rand-to best/1/bin: $x'_{r_1} = x_{r_1} + \lambda \cdot (x_{best} - x_{r_1}) + F \cdot (x'_{r_2} - x'_{r_3})$.
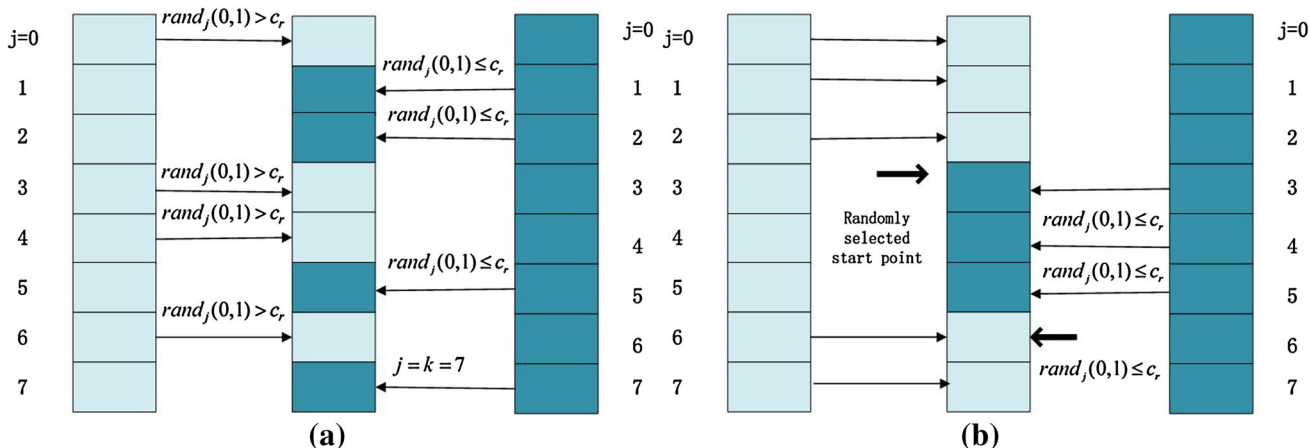


Fig. 1 The contrast of the two kinds of crossover operations. **a** The bin type; **b** The exp type; the three chromosomes showed in each figure from *left* to *right* respectively are $x_{r_1}$, $v$, $x'_{r_1}$

**Table 1** Each band corresponds to the wavelength information table

| Band number | Wave length (nm) | Band number | Wave length (nm) | Band number | Wave length (nm) | Band number | Wave length (nm) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 355.59 | 62 | 976.28 | 123 | 1376.55 | 184 | 1991.96 |
| 2 | 365.76 | 63 | 986.46 | 124 | 1389.64 | 185 | 2001.06 |
| 3 | 375.94 | 64 | 996.63 | 125 | 1396.74 | 186 | 2012.16 |
| … | … | … | … | … | … | … | … |
| 59 | 945.76 | 120 | 1346.25 | 181 | 1961.66 | 242 | 2577. 07 |
| 60 | 955.93 | 121 | 1356.35 | 182 | 1971.76 | | |
| 61 | 966.11 | 122 | 1366.45 | 183 | 1981.86 | | |

5. Select operation

A new individual $v$ is generated after crossover and mutation operation. According to the size of objective function values, an individual chosen from $x_{r_1}$ and $v$ should be inherited to the next generation. The following (taking getting minimum value of function as an example):

$$x_{r_1} = \begin{cases} x_{r_1}, & f(x_{r_1}) \leq f(v) \\ v, & else \end{cases} \tag{7}$$

## 2.2 The basic steps of DE algorithm

Steps of basic DE algorithm:

- Step1: Set the variation factor F, crossover probability CR and the maximum number of iterations $g_{max}$;
- Step2: Initialize the population $p(N \times D)$, set the number of iterations $g = 1$;
- Step3: Terminate the loop when $g = g_{max}$, otherwise, continue to do the next;
- Step4: Perform the following operations for every individual $x_{r_1}(1 \leq r_1 \leq 1)$ in current population;

    Step4.1: Perform the mutation for every gene of individual $x_{r_1}$ in a certain way, producing individual $x'_{r_1}$;
    Step4.2: Crossover operation was carried out on the individual $x_{r_1}$ and $x'_{r_1}$, insuring their genes are inherited to a new individual in a certain probability;
    Step4.3: Perform selection operation on $x_{r_1}$ and $v$, the individual with high fitness values heredity to the next generation;

- Step5: Add 1 to the number of iterations $g$, return to step3.

## 2.3 The improvement of DE algorithm

The ideas of Yin-Yang initialization is to create a mirror image for each individual at the center of the decision space for symmetric point, choosing a better one from the
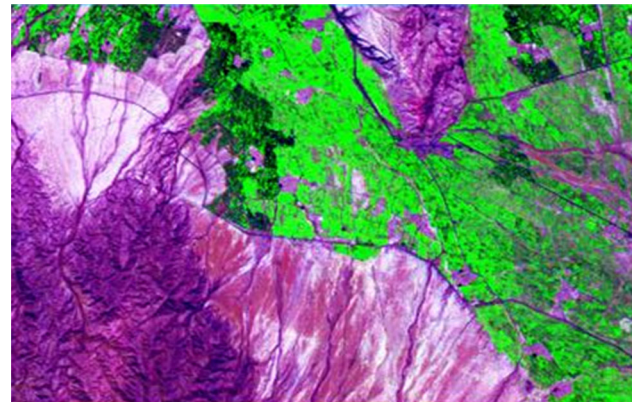


**Fig. 2** Image of hyper spectral data sample selection area

individual and its mirror image, so that it is more close to the global optimal solution. Through a filter for each individual, it will make the initial population more closely to the global optimal solution. In fact, the mirror image of each individual is more excellent than the original individuals in a probability of 50 %, so the initialization of Yin-Yang will accelerate convergence of the population (Das et al. 2008).

For any individual $p(x_1, x_2, \ldots, x_n)$ in population $pop(t)$, assuming each component of the individual satisfy conditions $x_i \in [a_i, b_i](i = 1, 2, \ldots, n)$, the individual $p$ in its decision-making space exists only one mirror individual $\tilde{P} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$, while $\tilde{x}_i = a_i + b_i - x_i(i = 1, 2, \ldots, n)$.

Assuming that $f(x)$ is the target problem which needs to find a minimum. Steps of Yin -Yang initialization are as follows:

- Step1: Randomly generating a population $pop(t) = \{x^1, x^2, \ldots, x^N\}$ in size N, while $x^i = (x_1^i, x_2^i, \ldots, x_n^i)$ $(i = 1, 2, \ldots, N)$, and $n$ is the dimension of decision space;
- Step2: Calculating the mirror $\tilde{x}^i = (\tilde{x}_1^i, \tilde{x}_2^i, \ldots, \tilde{x}_n^i)$ for each individual $x_i = (x_1^i, x_2^i, \ldots, x_n^i)$ in population, if $f(\tilde{x}_i) \leq f(x_i)$;
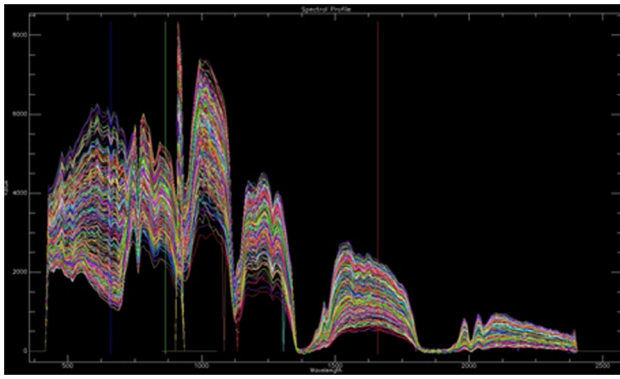
**Fig. 3** The band diagram for part of samples

After the above steps, the Yin-Yang initial population in single objective optimization problem is obtained. In the process of algorithm iteration, it can also introduce the initialization of Yin-Yang, the steps are as follows:

– Step1: Randomly generating a population $pop(t) = \{x^1, x^2, \ldots, x^N\}$ in size $N$, while $x^i = (x^i_1, x^i_2, \ldots, x^i_n)$

$(i = 1, 2, \ldots, N)$, and $n$ is the dimension of decision space;

– Step2: Calculating the mirror $\tilde{x}^i = (\tilde{x}^i_1, \tilde{x}^i_2, \ldots, \tilde{x}^i_n)$ for each individual $x_i = (x^i_1, x^i_2, \ldots, x^i_n)$ in population, and forming a mirror populations $p\tilde{o}p(t) = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\}$;

– Step3: Merging the two groups, that is $pop(t) \cup p\tilde{o}p(t)$, and sorting based on fitness function values, the first N individuals are genetic to the next generations.

In the process of practical application, a certain probability can be set to operate the mirror populations of each generation selectively, in order to improve the executive efficiency of the algorithm.

## 3 Experimental methods and results analysis

In this paper, the improved DE algorithm is adopted to search the optimal band combination, and ten different wavelength selection methods are selected as an evaluation criterion. comparing and analyzing the effects that whether

**Table 2** Table of optimal three band combination for each evaluation criterion based on improved DE algorithm

| Evaluation criterion | Band A | Band B | Band C | Value of fitness function |
|---|---|---|---|---|
| Combination entropy | 120 | 166 | 167 | 0. 5691056499 |
| Eigenvalue of covariance matrix | 32 | 77 | 87 | 1.45249e+17 |
| The best index | 32 | 33 | 77 | 854.4072641284 |
| Discrete degree | 9 | 37 | 129 | 19.0628072483 |
| The B distance | 35 | 39 | 47 | 14.9903921191 |
| Mixing distance | 30 | 32 | 33 | 9400.0585607940 |
| Euclidean distance | 30 | 32 | 33 | 5427.1553350273 |
| Mahalanobis distance | 78 | 93 | 217 | 541994. 4180718781 |
| Spectral angle | 120 | 182 | 183 | 0. 4364923556 |
| Spectral correlation coefficient | 26 | 40 | 93 | 1.1228e−05 |

**Table 3** The classification accuracy for different number of band combination with the evaluation criterion of covariance matrix Eigen values

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| Covariance matrix Eigen values | 5 | 32 40 77 79 84 | 84.52 | 91.90 | 97.86 | 91.43 |
| | 10 | 8 29 44 77 78 79 82 83 92 141 | 90.48 | 94.52 | 99.28 | 94.76 |
| | 20 | 8 9 24 47 52 77 82 88 89 101 108 110 132 136 137 140 184 188 202 212 | 85 | 95.71 | 98.09 | 92.93 |
| | 30 | 9 11 22 28 35 36 38 46 55 57 77 78 81 83 85 86 89 91 102 110 111 114 130 131 147 153 162 166 183 216 | 91. 3 | 96.67 | 99.29 | 95.80 |

**Table 4** The classification accuracy for different number of band combination with the evaluation criterion of the best index

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| The best index | 5 | 30 32 77 78 137 | 75.71 | 91.43 | 96.90 | 88.01 |
| | 10 | 26 29 30 32 33 | 75.95 | 93.33 | 98.57 | 89.28 |
| | | 77 78 90 91 167 | | | | |
| | 20 | 24 26 27 28 30 | | | | |
| | | 31 32 35 45 47 | | | | |
| | | 77 78 79 80 89 | 81.43 | 97.14 | 99.05 | 92.54 |
| | | 90 91 93 139 167 | | | | |
| | 30 | 10 15 23 24 25 | | | | |
| | | 27 28 31 32 33 | | | | |
| | | 39 40 42 52 56 | | | | |
| | | 57 77 78 86 87 | 81.43 | 97.86 | 99.52 | 92.94 |
| | | 88 89 91 93 133 141 | | | | |
| | | 142 167184 207 | | | | |

**Table 5** The classification accuracy for different number of band combination with the evaluation criterion of discrete degree

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| Discrete degree | 5 | 19 21 97 101 166 | 85.48 | 91.67 | 98.57 | 91.91 |
| | 10 | 13 21 86 95 100 | 80.00 | 94.05 | 98.33 | 90.79 |
| | | 134 145 201 202 218 | | | | |
| | 20 | 18 19 21 26 57 | | | | |
| | | 77 92 93 98 101 | | | | |
| | | 133 149 152 159 161 | 79.29 | 95.71 | 99.76 | 91.59 |
| | | 195 196 206 210 218 | | | | |
| | 30 | 15 22 23 36 44 | | | | |
| | | 51 52 53 54 55 | | | | |
| | | 82 87 100 108 120 | | | | |
| | | 128 133 134 147 149 | 81.67 | 95.95 | 99.76 | 92.46 |
| | | 155 160 163 166 183 | | | | |
| | | 204 205 211 216 219 | | | | |

using subspace division selected band combination on classification accuracy.

## 3.1 Encoding and decoding of the algorithm

DE algorithm has a fixed number of genes in the process of operation, namely, the band number of the band subsets. But the number of individual genes of DE algorithm can be modified before operation, and the selected number of band combination can be changed flexibly.

DE algorithm codes in decimal real number. Each individual (that is, the band subset) with m genes (that is, band) and each gene value is between 0 and 1. Taking band 20 for example, supposing that there are a total band number is 166 after preprocessing, the genes value of band 20 is 20/166 = 0.120482. Also, the finally optimal solution obtained by DE algorithm is expressed in real number, so the decoding for the optimum solution of real number encoding is needed, and the decoding process is opposite to the encoding. For example, the best individual gets a gene value as 0.120482, the corresponding band number is 0.120482 × 166 = 20. But for the unification with the original data band number (242 band), the band number must also be converted into one of the 242 band, band used in this paper all have been transformated.

## 3.2 Fitness function of the algorithm

DE algorithm adopts fitness equation to calculate the fitness of each individual, determining the probability of

**Table 6** The classification accuracy for different number of band combination with the evaluation criterion of Bhattacharyya distance

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| Bhattacharyya distance | 5 | 34 35 39 41 81 | 87.62 | 93.10 | 98.33 | 93.02 |
| | 10 | 13 21 86 95 100 | | | | |
| | | 134 145 201 202 218 | 91.43 | 95.24 | 98.57 | 95.08 |
| | 20 | 18 19 21 26 57 | | | | |
| | | 77 92 93 98 101 | | | | |
| | | 133 149 152 159 161 | 84.05 | 95.71 | 98.81 | 92.86 |
| | | 195 196 206 210 218 | | | | |
| | 30 | 15 22 23 36 44 | | | | |
| | | 51 52 53 54 55 | | | | |
| | | 82 87 100 108 120 | | | | |
| | | 128 133 134 147 149 | 82.86 | 96.90 | 99.29 | 93.02 |
| | | 155 160 163 166 183 | | | | |
| | | 204 205 211 216 219 | | | | |

**Table 7** The classification accuracy for different number of band combination with the evaluation criterion of Mahalanobis distance

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| Mahalanobis distance | 5 | 34 41 137 192 216 | 75.24 | 91.90 | 98.57 | 88.57 |
| | 10 | 16 40 81 84 117 | | | | |
| | | 131 141 154 163 190 | 84.52 | 93.81 | 98.81 | 93.18 |
| | 20 | 8 27 32 38 56 | | | | |
| | | 78 90 93 101 105 | | | | |
| | | 111 130 153 157 187 | 84.29 | 96.43 | 98.81 | 93.18 |
| | | 195 199 207 213 215 | | | | |
| | 30 | 10 14 19 20 26 | | | | |
| | | 27 29 45 46 78 | | | | |
| | | 82 84 97 107 108 | 85.71 | 96.43 | 99.05 | 93.73 |
| | | 110 111 114 117 134 | | | | |
| | | 139 148 149 163 167 | | | | |
| | | 197 200 202 212 217 | | | | |

inheritance to the next generation according to the values of the fitness. The bigger the fitness, the greater the probability of inheritance to the next generation is. This paper adopts the eigenvalue of covariance matrix, optimal index, discrete degree, Bhattacharyya distance, mahalanobis distance to calculate and evaluate the fitness, and sets the parameters of DE algorithm as follows: genetic iterations is 200, population size is 30, crossover probability is 0.3 and sacling factor is 0.5, adopt Rand1Bin mutation strategy.

### 3.3 The data source in research area

The experimental data in this paper is hyper spectral remote sensing image data which gets from hyper spectral imager sensor. The image size is that: column is 3408, row

is 256, band is 242, spectral resolution is nearly 10 nm, spectral range from 360 to 2570 nm; Data format is HDF Scientific Data; The storage format of image element is BIL.

Part of the information for band in the following Table 1.

According to the prior knowledge, this paper selects the top left corner (1,3009) and the bottom right corner(256,3408) from study area, size of $256 \times 400$. The main reason of selecting such area as experimental area from the full 256*3408 is that the type of terrain of this region is relatively abundant and easy to identify, as shown in Fig. 2. On this basis, looking at the map for the corresponding latitude and longitude through world wind developed by NASA. In order to observe and confirm the
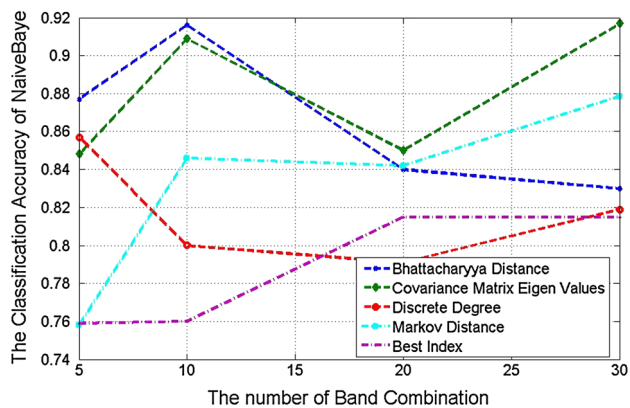
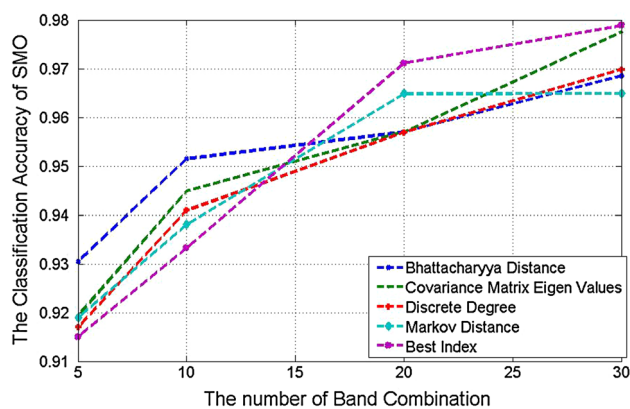**Fig. 4** The classification accuracy graph by using Naive Bayes classifier for each evaluation criteria



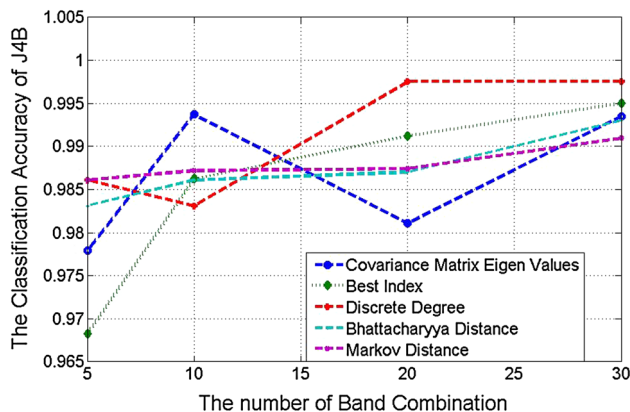**Fig. 5** The classification accuracy graph by using SMO classifier for each evaluation



**Fig. 6** The classification accuracy graph by using J48 classifier for each evaluation criteria



**Fig. 7** The average classification accuracy graph for each evaluation criteria

distribution of mountains, vegetation, water body, surface in the experimental area by three-dimensional terrain model. It is convenient for sampling on the basic.
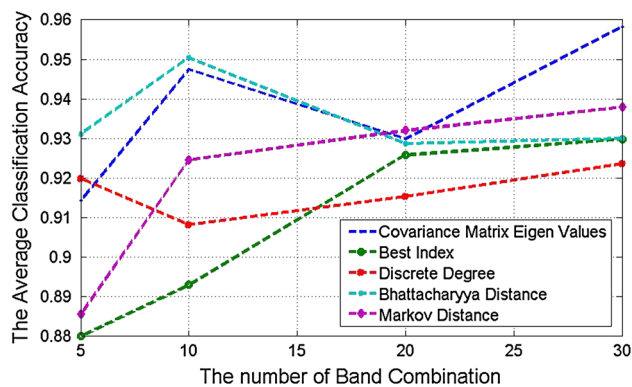
In this paper, 4200 samples were chosen according to four kinds of feature classes such as mountains, vegetation, water body, and surface body, the band diagram for part of samples as shown in Fig. 3:

### 3.4 Result of the experiment and contrastive analysis

1. Results of optimal band selection of three band number

   The experimental results of optimal three band combination that use the improved DE algorithm for each evaluation criterion is respectively shown in Table 2. Comparing the results by the enumeration method for the evaluation criteria, we can find that the algorithm can get the result of optimal band combination with enumeration method in a short time, which further shows that the improved DE algorithm is feasible and effective for band selection of hyperspectral data.

2. Multiband number of optimal band selection in no subspace division

   When enumeration methods are taken for calculating multiband combination (such as 5 band, 10 band even more band), the computation is too large and the operation time is too long. DE algorithm has the advantages of fast convergence in solving combinatorial optimization problems. The following experiment is the selection of multiband combination by using the improved DE algorithm in different evaluation criteria (such as the eigenvalue of covariance matrix, the best index, discrete degree, B distance and mahalanobis distance), and then three kinds of classifiers (Naive Bayes theorem, SMO and J48) are used to verify the optimal band combination. Different classifiers have different classification accuracy, so the average results

**Table 8** The classification accuracy for different number of band combination with the evaluation criterion of covariance matrix Eigen values (with subspace division)

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| Covariance matrix Eigen values | 5 | 21 42 77 106 140 | 90.95 | 94.29 | 98.33 | 94.52 |
| | 10 | 8 24 40 78 79 | | | | |
| | | 100 106 107 131 141 | 91.19 | 95.00 | 99.52 | 95.24 |
| | 20 | 8 9 10 12 29 | | | | |
| | | 52 80 96 97 100 | | | | |
| | | 105 109 110 118 128 | 88.57 | 96.19 | 99.29 | 94.68 |
| | | 140 141 197 205 219 | | | | |
| | 30 | 15 19 24 29 41 | | | | |
| | | 42 43 44 46 54 | | | | |
| | | 79 84 85 86 87 | | | | |
| | | 103 108 109 111 115 | 90.71 | 95.95 | 98.10 | 94.92 |
| | | 117 118 119 151 152 | | | | |
| | | 189 190 194 196 197 | | | | |

**Table 9** The classification accuracy for different number of band combination with the evaluation criterion of the best index (with subspace division)

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| The best index | 5 | 24 29 78 110 138 | 82.62 | 93.33 | 97.86 | 91.27 |
| | 10 | 25 33 35 78 86 | | | | |
| | | 91 107 117 167 196 | 92.86 | 94.05 | 98.10 | 95.00 |
| | 20 | 16 21 26 32 44 | | | | |
| | | 46 77 91 92 93 | | | | |
| | | 105 106 113 115 119 | 91.19 | 95.95 | 98.57 | 95.24 |
| | | 137 139 149 199 216 | | | | |
| | 30 | 12 23 24 26 29 | | | | |
| | | 32 33 40 48 51 | | | | |
| | | 52 84 85 86 91 | | | | |
| | | 92 107 109 110 112 | 90.95 | 96.43 | 99.05 | 95.48 |
| | | 115 117 120 131 135 141 | | | | |
| | | 147 150 159 167 | | | | |

of the three kinds of classification accuracy are treated as the final results.

Tables 3, 4, 5, 6 and 7 shows the different band combination classification accuracy in different evaluation standards without subspace division calculation. Observing the data in the table, we can know that for each evaluation criterion method the classification accuracy is not on the rise with more number of band.

Figure 5 shows the classification accuracy graph by using SMO classifier for each evaluation criteria (abscissa is the band number of band combination, ordinate is the classification accuracy, following figures are the same), the picture shows that the greater the band number, the higher the classification accuracy is Figs. 4 and 6 respectively show the classification accuracy graph by using Naive Bayes classifier and J48 classifier for each evaluation criteria. It seems that the classification accuracy do not increase with the number of band according to the two pictures, and it turns to the fold point when band number is ten. Figure 7 is the graph of average classification accuracy for each evaluation criterion.

**Table 10** The classification accuracy for different number of band combination with the evaluation criterion of discrete degree (with subspace division)

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| Discrete degree | 5 | 15 17 103 111 166 | 88.33 | 93.33 | 98.33 | 93.33 |
| | 10 | 9 42 44 51 97 | | | | |
| | | 102 111 112 167 218 | 87.38 | 98.10 | 98.33 | 93.18 |
| | 20 | 8 24 29 30 42 | | | | |
| | | 48 79 80 89 98 | | | | |
| | | 105 106 107 114 120 | 90.00 | 96.19 | 98.81 | 95.00 |
| | | 139 140 141 199 217 | | | | |
| | 30 | 9 15 16 18 25 | | | | |
| | | 29 38 44 53 54 | | | | |
| | | 55 81 88 89 97 | | | | |
| | | 103 107 109 112 113 | 91.43 | 96.19 | 99.05 | 95.56 |
| | | 114 115 118 131 132 | | | | |
| | | 137 149 189 198 219 | | | | |

**Table 11** The classification accuracy for different number of band combination with the evaluation criterion of Bhattacharyya distance (with subspace division)

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| Bhattacharyya distance | 5 | 9 40 52 112 142 | 90.48 | 93.10 | 97.62 | 93.73 |
| | 10 | 14 34 37 52 97 | | | | |
| | | 100 117 118 149 167 | 91.90 | 93.81 | 98.81 | 94.84 |
| | 20 | 9 17 21 33 44 | | | | |
| | | 47 49 82 100 102 | | | | |
| | | 106 108 109 113 117 | 91.43 | 95.00 | 98.57 | 95.00 |
| | | 134 151 154 161 167 | | | | |
| | 30 | 8 9 11 14 15 | | | | |
| | | 27 31 33 48 52 | | | | |
| | | 54 57 89 90 101 1 | | | | |
| | | 03 109 111 112 113 | 88.81 | 96.90 | 100 | 95.24 |
| | | 115 120 128 131 134 | | | | |
| | | 137 145 152 167 197 | | | | |

Therefore, when selecting band for hyperspectral data, we should choose the band combination with a larger band number as the optimal band combination while using SMO classifier. The band combination with a band number of 10 should be the optimal band combination for the Naive Bayes classifier and J48 classifier application.

3. The optimal band selection with subspace division of multiband

Tables 8, 9, 10, 11 and 12 shows the classification accuracy of different band number of band combination

calculated with each evaluation standards after subspace division. By comparing the experimental results without subspace division above, it can be found that the classification accuracy of band combination of each evaluation criterion is almost improved. Therefore, this group of experiments prove that in the band selection of hyper spectral, the band combination has a good effect on the final classification result after adding subspace division. Figure 8 is the curve contrast figure of the average classification accuracy for the five kinds of evaluation criteria before and after subspace division.

**Table 12** The classification accuracy for different number of band combination with the evaluation criterion of Mahalanobis distance (with subspace division)

| Evaluation criterion | Number of band | Optimal band combination | Naive Bayes (%) | SMO (%) | J48 (%) | Average classification accuracy (%) |
|---|---|---|---|---|---|---|
| Mahalanobis distance | 5 | 14 32 84 113 167 | 90.24 | 95.00 | 98.57 | 94.60 |
| | 10 | 19 23 40 56 96 | | | | |
| | | 97 110 115 133 190 | 90.48 | 94.05 | 98.81 | 94.45 |
| | 20 | 8 13 30 36 42 | | | | |
| | | 47 89 92 100 102 | | | | |
| | | 105 112 114 118 120 | 91.19 | 95.95 | 99.29 | 95.48 |
| | | 138 188 190 210 218 | | | | |
| | 30 | 8 13 15 17 20 | | | | |
| | | 23 37 41 48 51 | | | | |
| | | 55 78 85 87 93 | 92.14 | 96.43 | 99.29 | 95.95 |
| | | 97 108 110 111 112 | | | | |
| | | 113 117 128 132 149 | | | | |
| | | 152 167 184 190 200 | | | | |

## 4 Conclusion

In this paper, an improved DE algorithm is put forward for hyperspectral remote sensing data with characteristics of a large number of band, high spectral resolution, narrow bandwidth, huge amount of data. It implemented a fast and efficient dimension reduction model of hyperspectral remote sensing data. Contrast experiment proves that the algorithm can not only help obtain the optimal 3 bands combination in a short time, but also can contribute to the optimal combination of 5 bands, 10 bands and even a larger number, resulting in more precise classification.
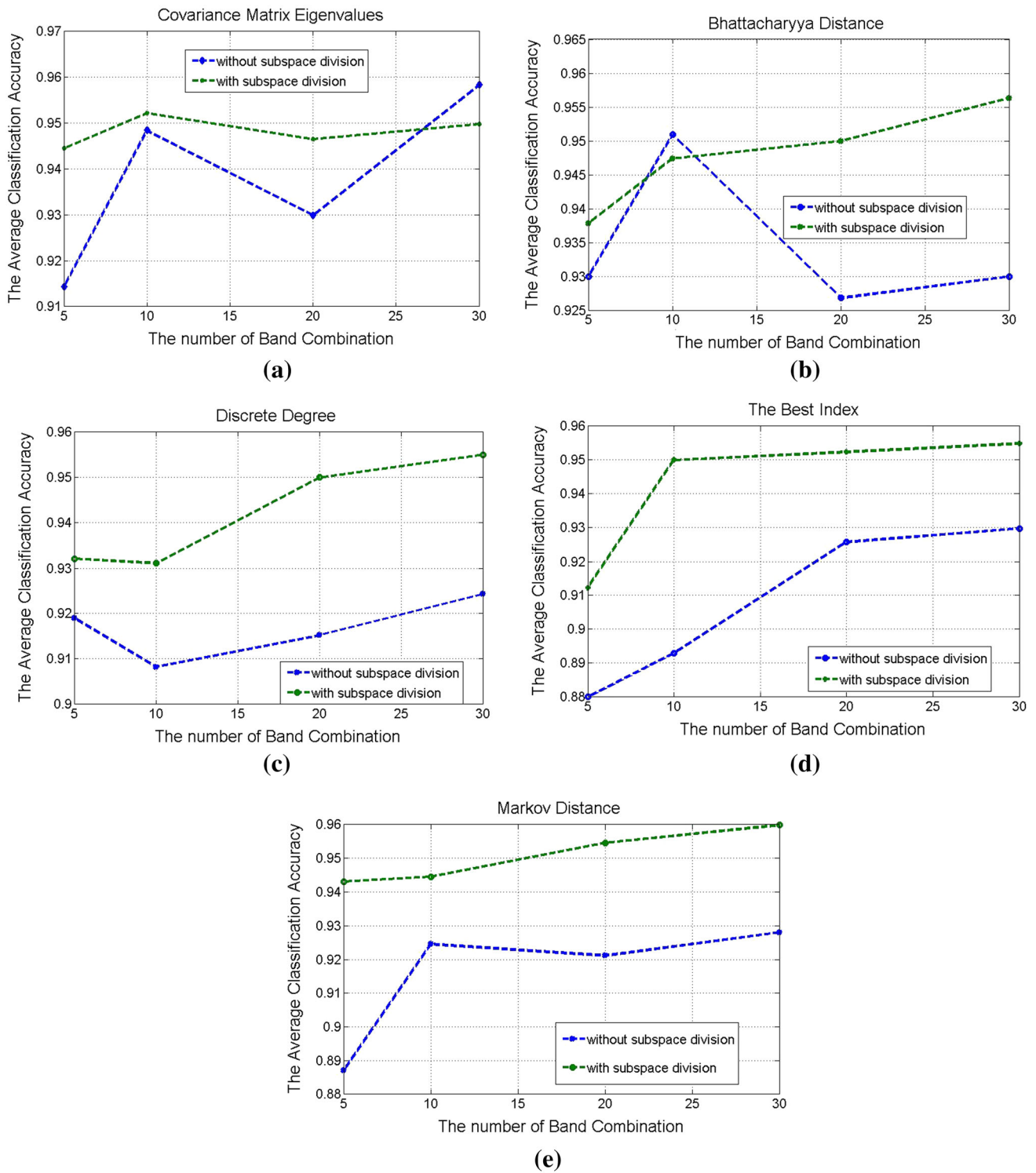
**Fig. 8** The curve contrast figure of the average classification accuracy for the five kinds of evaluation criteria before and after subspace division. **a** Covariance matrix Eigen values; **b** Bhattacharyya distance; **c** discrete degree; **d** the best index; **e** Mahalanobis distance

# References

Alatas B, Akin E, Karci A (2008) Multi-objective differential evolution algorithm for mining numeric association rules. Appl Soft Comput 1(8):646–656. doi:10.1016/j.asoc.2007.05.003

Benavente-Peces C, Ahrens A, Filipe J (2014) Advances in technologies and techniques for ambient intelligence. J Ambient Intell Humaniz Comput 5(5):621–622

Castiglione A, Pizzolante R, Santis A-D, Carpentieri B, Castiglione A, Palmieri F (2015) Cloud-based adaptive compression and secure management services for 3d healthcare data. Futur Gener Comput Syst 43–44(0):120–134. ISSN 0167–739X

Cerezo E, Hupont I, Baldassarri S, Ballano S (2012) Emotional facial sensing and multimodal fusion in a continuous 2d affective space. J Ambient Intell Humaniz Comput 3(1):31–46. ISSN 1868–5137

Chen J, Xin B, Peng Z-H, Pan F (2009) Statistical learning makes the hybridization of particle swarm and differential evolution more efficient a novel hybrid optimizer. Sci China Ser F Inf Sci 52(7):1278–1282. ISSN 1009–2757

Chen J, Xin B, Peng Z-H, Pan F (2010) An adaptive hybrid optimizer based on particle swarm and differential evolution for global optimization. Sci china Inf Sci 53(5):980–989

Christian E, Massimo F, Francesco P, Aniello C (2013) Interconnecting federated clouds by using publish-subscribe service. Clust Comput 16(4):887–903. ISSN 1386–7857

Das S, Abraham A, Konar A (2008) Automatic clustering using an improved differential evolution algorithm. Syst Man Cybern Part A Syst Hum IEEE Trans 38(1):218–237. ISSN 1083–4427

Feng L, Tan A-H, Lim M-H, Jiang S.-W (2014) Band selection for hyperspectral images using probabilistic memetic algorithm. Soft Comput pp 1–9. ISSN 1432–7643

Huang R, He M-G (2005) Band selection based on feature weighting for classification of hyperspectral data. Geosci Remote Sens Lett IEEE 2(2):156–159. ISSN 1545–598X

Kitti K, Chuleerat J, Apisit E (2012) Band selection for hyperspectral imagery with pca-mig. In: Zhifeng B, Gao Y, Gu Y, Guo L, Li Y, Lu J, Zujie R, Wang C, Zhang X (eds) Web-age information management, volume 7419 of lecture notes in Computer Science, pp 119–127. Springer, Berlin. ISBN 978-3-642-33049-0

Liu B, Wang L, Jin Y-H (2007) Advances in differential evolution. Control Decis 22(7):721–729. doi:10.13195/j.cd.2007.07.3.liub.001

Liu X-B, Yu C, Cai Z-H (2010) Differential evolution based band selection in hyperspectral data classification. In: Cai Z, Hu C, Kang Z, Liu Y (eds) Advances in computation and intelligence, volume 6382 of lecture notes in computer science, pp 86–94. Springer, Berlin. ISBN 978-3-642-16492-7

Peng Z-W, Liao J-L, Cai Y-Q (2015) Differential evolution with distributed direction information based mutation operators: an optimization technique for big data. J Ambient Intell Humaniz Comput pp 1–14. ISSN 1868–5137

Wu H, Zhu J-L, Li S-J, Wan D-S, Lin L (2010) A hybrid evolutionary approach to band selection for hyperspectral image classification. In: Zeng Z and Wang J (eds) Advances in neural network research and applications, volume 67 of lecture notes in electrical engineering, pp 329–336. Springer, Berlin. ISBN 978-3-642-12989-6

Yang M, Guan J, Cai Z-H, Wang L (2010) Self-adapting differential evolution algorithm with chaos random for global numerical optimization. In: Cai Z, Hu C, Kang Z, Liu Y (eds) Advances in computation and intelligence, volume 6382 of lecture notes in computer science, pp 112–122. Springer, Berlin. ISBN 978-3-642-16492-7

Zhang D-M, Liu W, Xu X, Deng Q (2010) A novel interpolation method based on differential evolution-simplex algorithm optimized parameters for support vector regression. In: Cai Z, Hu C, Kang Z, Liu Y (eds) Advances in computation and intelligence, volume 6382 of lecture notes in computer science, pp 64–75. Springer, Berlin. ISBN 978-3-642-16492-7

Zhao S-Z, Suganthan P-N, Swagatam D (2011) Self-adaptive differential evolution with multi-trajectory search for large-scale optimization. Soft Comput 15(11):2175–2185. ISSN 1432–7643