Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization

Pedram Ghamisi, *Student Member, IEEE*, and Jon Atli Benediktsson, *Fellow, IEEE*

***Abstract*—A new feature selection approach that is based on the integration of a *genetic algorithm* and *particle swarm optimization* is proposed. The overall accuracy of a *support vector machine* classifier on validation samples is used as a fitness value. The new approach is carried out on the well-known Indian Pines hyperspectral data set. Results confirm that the new approach is able to automatically select the most informative features in terms of classification accuracy within an acceptable CPU processing time without requiring the number of desired features to be set *a priori* by users. Furthermore, the usefulness of the proposed method is also tested for road detection. Results confirm that the proposed method is capable of discriminating between road and background pixels and performs better than the other approaches used for comparison in terms of performance metrics.**

***Index Terms*—Attribute profile, feature selection, hybridization of genetic algorithm (GA) and particle swarm optimization (PSO), hyperspectral image analysis, road detection, support vector machine (SVM) classifier.**

I. Introduction

SUPERVISED classification techniques classify the input data by partitioning the feature space into decision regions, by using a set of training samples for each class. These samples are usually obtained by manual labeling of a small number of pixels in an image or based on some field measurements. Thus, the collection of these samples is expensive and time demanding. As a result, the number of available training samples is usually limited, which is a challenging issue in supervised classification.

In [1], it was shown that, after a few features, while the number of training samples is kept constant, the classification accuracy actually decreases as the number of features increases. For the purpose of classification, this is referred to as the *curse of dimensionality* [2]. To address this issue, the use of feature selection/extraction techniques is of importance.

Feature extraction/selection techniques can be grouped into two categories: unsupervised and supervised approaches. For the purpose of image classification, the latter techniques are preferred since they try to reduce the dimensionality of the data while maximizing the separability between classes. Nonparametric weighted feature extraction (NWFE) and parametric decision boundary feature extraction (DBFE) have been extensively used for this purpose. On the other hand, divergence, transformed divergence, Bhattacharyya distance, and Jeffries–Matusita distance are well-known feature selection techniques that have been widely used in remote sensing. For more information on the aforementioned techniques, please see [1].

Conventional feature selection techniques usually demand many samples to estimate statistics accurately. In addition, they are usually based on an exhaustive process for finding the best set of features, and in this case, they are time demanding, and their CPU processing time exponentially increases as the number of bands (features) increases. To this extent, a new generation of feature selection techniques is based on evolutionary optimization methods, since they are not based on an exhaustive process and can lead to a conclusion in a faster way. In addition, by considering an efficient fitness function for these methods, they can handle high-dimensional data with even a limited number of training samples (ill-posed situations). In particular, the *genetic algorithm* (GA) and *particle swarm optimization* (PSO) have gained significant attention from researchers. There is an extensive literature regarding the use of the GA and PSO for the purpose of feature selection. For example, in [3], Bazi and Melgani proposed a support vector machine (SVM) classification system that allows for detecting the most distinctive features and estimating the SVM parameters by using a GA. In [4], Daamouche *et al.* proposed the use of PSO to select for classification the most informative features obtained by morphological profiles. However, both PSO and the GA suffer from a few shortcomings. The main shortcoming of PSO is the premature convergence of a swarm. The key reason behind this shortcoming is that particles try to converge to a single point, which is located on a line between the global best and the personal best positions. This point is not guaranteed for a local optimum [5]. Another reason could be the fast rate of information flow between particles, which leads to the creation of similar particles. This results in a loss in diversity. Furthermore, the possibility of being trapped in local optima is increased [6]. The main advantage of using PSO is its simple concept along with the fact that it can be implemented in a few lines of code. Furthermore, PSO also has a memory of past iterations. On the other hand, in the GA, if a chromosome is not selected, the information contained by it is lost. However, without a selection operator as in the GA, PSO may waste resources on inferior individuals [6]. PSO may enhance the search capability for finding an optimal solution. However, the GA has difficulty in finding an exact solution [7].

In this letter, to address the main shortcomings of GA- and PSO-based feature selection techniques and to take the advantage of their strength, a new feature selection approach is proposed, which is based on the integration of the GA

Manuscript received May 12, 2014; revised June 17, 2014; accepted July 7, 2014. Date of publication July 28, 2014; date of current version August 21, 2014. This work was supported in part by the Icelandic Research Fund for Graduate Students.

The authors are with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: peg6@hi.is; benedikt@hi.is).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2014.2337320

and PSO. To find the most discriminative features in terms of classification accuracies, the overall accuracy (OA) of the SVM classifier over validation samples is investigated as a fitness value. The SVM is selected due to the fact that it is capable of providing acceptable classification accuracies for high-dimensional data when even a limited number of training samples is available. To evaluate the efficiency of the proposed method, two different scenarios are drawn.

i) In the first scenario, the proposed feature selection approach is performed on a well-known hyperspectral data set, i.e., the AVIRIS Indian Pines. Results demonstrate that the new method can significantly increase the classification accuracy of the raw data in an acceptable CPU processing time.
ii) In the second scenario, the proposed feature selection technique is applied on a set of features derived by attribute profiles [8], to select the most discriminative features for detecting roads from a background. Results infer that the new feature selection approach is able to detect roads in a complex urban image with acceptable accuracy.

This letter is organized as follows: The proposed methodology is discussed in Section II. Section III is devoted to experimental results. Finally, Section IV outlines the main conclusions.

II. Methodology

Here, first, the concept of two well-known optimization techniques, namely, GA and PSO, will be recalled. Then, the proposed feature selection technique, which is based on the hybridization of the GA and PSO (HGAPSO + SVM), will be described.

A. GA

The GA is inspired by the genetic process of biological organisms. The GA consists of several solutions called chromosomes or individuals. Each chromosome in a binary GA includes several genes with binary values 0 and 1, which determines the attributes for each individual. A set of the chromosomes is made up to form a population. The merit of each chromosome is evaluated by using a fitness function. Fit chromosomes are selected for the generation of new chromosomes. In that step, two fit chromosomes are selected and combined through a crossover step to produce a new offspring (or solution). Then, mutation is applied on the population to increase the randomness of individuals for decreasing the possibility of getting stuck in local optimum [9].

B. PSO

PSO is a biologically inspired technique derived from the collective behavior of bird flocks, first introduced by Kennedy and Eberhart [10]. PSO consists of a set of solutions (particles) called population. Each solution consists of a set of parameters and represents a point in multidimensional space. A group of particles (population) makes up a swarm. Particles move through the search space with a specified velocity for finding the optimal solution. Each particle has a memory that helps it in keeping the track of its previous best position. The positions of the particles are distinguished as *personal best* and *global best*. The velocities of particles are adjusted according to the historical behavior of each particle and its neighbors, while they fly through the search space. Each move of particles is deeply influenced by its current position, its memory of previous useful parameters, and the group knowledge of the swarm [10]. Therefore, the particles have a tendency to fly toward improved search areas over the course of the search process.

The velocity of the ith particle in the $(k+1)$th iteration is mathematically defined as

$$V_i^{k+1} = WV_i^k + C_1 r_1 \left(pb_i^k - X_i^k\right) + C_2 r_2 \left(gb_d^k - X_i^k\right) \quad (1)$$

where C_1 and C_2 are acceleration constants, r_1 and r_2 are random values in the range of 0 and 1, W is the inertia weight (predefined by the user), X_i^k shows the position of each particle in d-dimensional search space, pb_i^k is the best previous position of each particle named particle best position, and gb^k is the best position of all the particles (called the global best particle). The position of the ith particle is updated by

$$X_i^{k+1} = X_i^k + V_i^{k+1}. \quad (2)$$

The PSO was originally introduced for the optimization of problems in continuous multidimensional search space. To extend that concept to feature selection, it needs to be developed to deal with binary data, in which 0 and 1 demonstrate the absence and presence of a band, respectively. In [10], *Kennedy* and *Eberhart* applied the sigmoid transformation on the velocity component to develop a binary discrete PSO to control the range of velocity between 0 and 1 according to

$$\Delta X_i^{k+1} = \frac{1}{1 + \exp\left(-V_i^{k+1}\right)}. \quad (3)$$

For updating the position of each particle, ΔX_i^{k+1} is compared with r_x, which is a random d-dimensional vector in which each component is, in general, a uniform random number between 0 and 1 according to

$$X_i^{k+1} = \begin{cases} 1, & \Delta X_i^{k+1} \geq r_x \\ 0, & \Delta X_i^{k+1} < r_x. \end{cases} \quad (4)$$

C. HGAPSO + SVM

GA and PSO can be combined in different ways. However, in the proposed feature selection approach, hybridization is obtained through integrating the standard velocity and update rules of PSO with selection, crossover, and mutation from the GA.

Fig. 1 shows the block diagram of the proposed approach. To investigate the hybridization of the GA and PSO for the purpose of feature selection, the dimension of each particle needs to be equal to the number of features. In this case, that velocity dimension, i.e., $\dim V_i^k$, as well as the position dimension, i.e., $\dim X_i^k$, correspond to the total number of bands (l bands) in the input data $(\dim V_i^k = \dim X_i^k = l)$. In that case, each particle's velocity is represented as a l-dimension vector. In addition, as one wishes to use the algorithm for band selection, each particle represents its position in binary values, i.e., 0 or 1, where 0 and 1 demonstrate the absence and the presence of the corresponding feature, respectively.

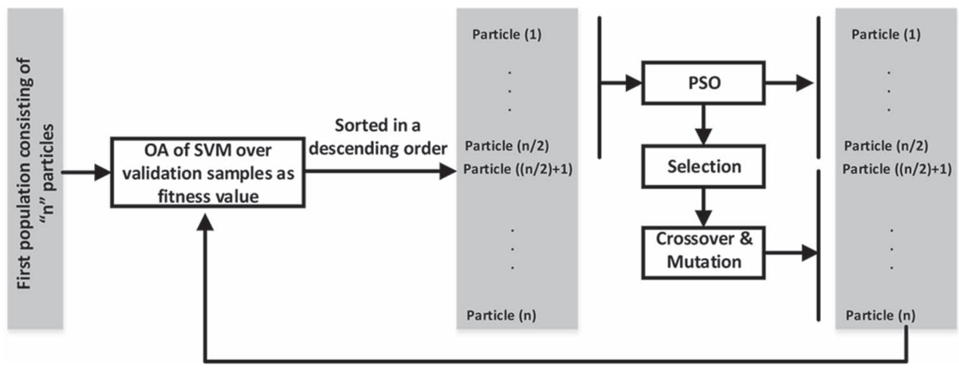

Fig. 1. Flowchart of the proposed method.

In this letter, a random population is initially generated. The individuals in the population may be regarded as chromosomes with respect to the GA or as particles with respect to PSO. Then, a new population for the next generation is produced through enhancement, crossover, and mutation as described below.

Enhancement: In each generation, after the fitness values for all the individuals in the same population are calculated (the OA of SVM on validation samples), the top half of the best-performing particles is selected. These individuals are regarded as elites. Then, the elites are enhanced by PSO. By using these enhanced elites as parents, the generated offsprings usually achieve better performance than using the elites directly [11]. Furthermore, (1) is applied to the elites. In each iteration, the range of velocity is regulated between 0 and 1 with the sigmoid function [see (3)] and compared with a random chromosome between 0 and 1 to update the position in binary format [see (4)]. By performing PSO on the elites, the search ability of the algorithm may increase. Half of the population in the next generation consists of the enhanced individuals, and the rest is generated by the crossover operation.

Crossover: To produce well-performing individuals, the crossover operation is only performed on selected individuals produced by PSO. To select parents for the crossover operation, a tournament selection scheme is used, in which two enhanced elites are selected at random, and their fitness values are compared. The individual with the better fitness value is selected as a parent and inserted into the mating pool. Then, the two individuals are moved back to the population. In the same way, the other parent is chosen and moved to the mating pool. Two offsprings are created by performing crossover on the selected parents. A two-point crossover operation is used. The produced offsprings make up the other half of the population in the next generation.

Mutation: This operation occurs along with the crossover operation. Here, uniform mutation is adopted. In our case, a constant mutation probability that is equal to 0.01 is used.

III. Experimental Results

A. Description of Data Sets

1) Indian Pines: The hyperspectral data set used in experiments is the well-known AVIRIS data captured of Indian Pines (NW Indiana) in 1992 comprising 16 classes, mostly related to different land covers. The data set consists of 145×145 pixels with a spatial resolution of 20 m/pixel. In this letter, 220 data channels (including all noisy and atmospheric absorption bands) are used. Training samples are available for 16 classes, and the total number of training and test samples is 695 and 9691, respectively. The same training and test samples for all 16 classes as in [12] are chosen, and a half of the training samples is selected for validation.

2) Toronto: The RGB Toronto Roads data set is captured at the resolution of 1.2 m/pixel. This data set contains three bands consisting of 1500×1500 samples. Fig. 2(a) and (b) shows this data set and its corresponding digitized samples [13]. For this data set, 0.01 of the total samples are randomly chosen as training samples (1052 samples for class Road and 21448 for class No-road) and the rest as test samples (10 2007 samples for class Road and 2 125 493 for class No-road). Then, a half of the training samples is chosen for validation.

B. General Information

The proposed method was implemented in *MATLAB*, on a computer having Intel(R) Core(TM) i7 CPU 2.40 GHz and 16 GB (15.9 GB usable) of memory.

The number of populations in the first and second scenarios was set as 20 and 10, respectively. The same set of parameters for both data sets was chosen, which infers that the proposed method is data set distribution independent, and there is no need to set any parameters for it, and the method can automatically choose the most informative bands in terms of classification accuracies.

The hybridization of GA-PSO will automatically stop, when the difference between the OA of the best solution and the average value of fitness values in a swarm is less than a predefined threshold value.

For the first scenario, to compare the capability of the proposed methodology, four well-known feature selectors, namely, divergence, transformed divergence, Bhattacharyya distance, and Jeffries–Matusita distance, have been taken into account. In addition, two frequently used supervised feature extraction techniques, namely, DBFE and NWFE, have been considered. In the case of NWFE and DBFE, features with cumulative eigenvalues above 99% are retained and classified with SVM. This way of choosing features has been widely used in the literature (e.g., in [14] and [15]). In addition to the aforementioned techniques, GA + SVM and PSO + SVM are investigated to be compared with the proposed approach. Since the first scenario is related to feature selection and image classification,

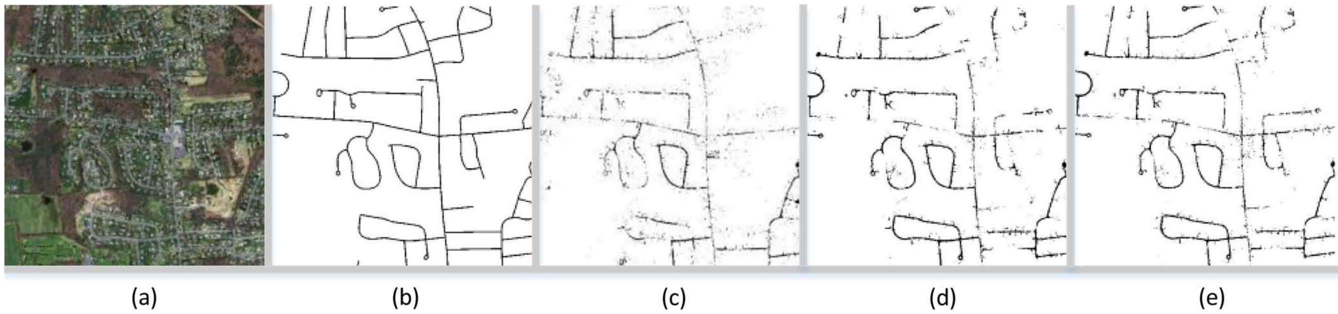

Fig. 2. (a) Input data. (b) Manually produced reference map. (c) Map obtained by SVM_{RGB}. (d) Map obtained by SVM_{AP}. (e) Map obtained by HGAPSO + SVM_{AP}.

TABLE I
FIRST SCENARIO: CLASSIFICATION ACCURACIES AND CPU PROCESSING TIME IN SECONDS. THE BEST RESULT IN EACH ROW IS SHOWN IN BOLDFACE. THE NUMBER OF FEATURES IS SHOWN IN BRACKET. SINCE PSO + SVM, GA + SVM, AND HGAPSO + SVM ARE THE AVERAGE OF TEN RUNS, THE NUMBER OF FEATURES IS NOT GIVEN

Index	SVM (220)	DBFE (17)	NWFE (120)	PSO +SVM	GA +SVM	HGAPSO +SVM
AA	76.02	73.36	60.13	74.94±2.54	73.96±4.37	**77.92**±1.42
Kappa	0.6119	0.6055	0.5533	0.7281±0.025	0.7141±0.040	**0.7495**±0.0069
OA	65.41	64.96	68.44	74.69±2.38	73.39±3.81	**76.68**±0.64
Time(s)	70	72	132	293±39	121±19	201±58

OA, average accuracy, kappa coefficient, and CPU processing time are considered for the evaluation of the final results. Since the PSO+SVM, GA+SVM, and HGAPSO+SVM are based on evolutionary techniques and their results can be different in different runs, all aforementioned approaches have been run ten times, and the average results are reported in Table I.

For the second scenario, since it is related to road detection, the root mean square error (RMSE) is taken into account as it was suggested that the RMSE is the most solid index [16]. For this scenario, since the Toronto data consist of only three components (RGB), to produce extra features, an attribute profile is used. A morphological attribute profile is considered as the generalization of morphological profile, which simplifies the input image by using the sequential stricter thresholds to model spatial information of the input image. For a detailed description of the attribute profile, refer to [8] and [14]. In this letter, three attributes, i.e., area ($\lambda_a = (1000/v)\ \{1,3,5,7\}$, where v is the resolution of the input data), standard deviation ($\lambda_s = (\mu_i/100)\ \{30,40\}$, where μ is the mean of the ith feature), and the diagonal of the box bounding the regions ($\lambda_d = \{25,50,100\}$), are used. However, other types of attributes with different ranges can be used. In this case, 19 features for each component (including itself) were produced. Since we have three components (R, G, and B), the total number of produced features is 57, which was considered as the input for the proposed methodology. Then, HGAPSO + SVM is applied on the features obtained by the attribute profile (and named as HGAPSO + SVM_{AP}) and compared with 1) the result of SVM performed on the RGB data (named as SVM_{RGB}) and 2) the result of SVM performed on the features produced by the attribute profile (named as SVM_{AP}).

The data sets have been classified with SVM using a Gaussian kernel. Fivefold cross validation is taken into account to select the hyperplane parameters when SVM is used for the last step (for the classification of informative bands).

C. First Scenario

The result of classification with different techniques is listed in Table I. These results have been obtained when conventional feature selection techniques, including divergence, transformed divergence, and Bhattacharyya distance, cannot work due to the singularity of the covariance matrix. The main reasons behind this shortcoming are that the conventional feature selectors cannot eliminate the corrupted bands automatically, and this step should be done manually, which is time consuming. In addition, when there is not a balance between the number of bands and the number of training samples, the aforementioned conventional feature selection techniques will not perform well. Furthermore, almost all of the conventional feature selection methods are computationally time demanding. For those approaches, to select a subset of m features out of a total of n features, $n!/(n-m)!m!$ alternatives must be calculated, which is a laborious task and demands a lot of computational memory. In other words, the feature selection techniques are only feasible in relatively low-dimensional cases. Another shortcoming of most of the conventional methods (particularly divergence, transformed divergence, and Bhattacharyya distance) is that the number of desired features must be initialized *a priori*. In contrast, since evolutionary-based feature selection techniques (e.g., PSO + SVM, GA + SVM, and HGAPSO + SVM) are not based on the calculation of the second-order statistics, the singularity of the covariance matrix is not a problem. In addition, when an evolutionary technique is taken into consideration, there is no need to calculate all different alternatives to find the most informative bands, and, therefore, these methods are usually faster than the conventional ones. Furthermore, in the proposed method, there is no need to initialize the number of desired features, and the approach can find the most informative bands with respect to the OA of SVM over the validation samples.

Some algorithms, such as the originally proposed DBFE, demand the use of second-order statistics (i.e., the covariance matrix) to characterize the distribution of training samples with respect to the mean. In that case, if the number of available training samples is not sufficient, a good estimation of the covariance matrix might be impossible. For this purpose, the use of a sample covariance or a common covariance [1] may not be successful. As an example, either when the sample covariance or the common covariance is taken into account to estimate the statistics for each available class for DBFE, if the number of pixels in the classes is not, at least one, greater than the total number of features being used, the DBFE stops working. To handle this issue, the leave-one-out covariance [1] estimator can be used as an alternative to estimate the covariance matrix. However, this is not a problem for evolutionary-based feature selectors since they are nonparametric and do not need to estimate class conditional densities. In addition, they can efficiently handle high-dimensional data with a very limited number of training samples due to the generalization of the SVM, which has been considered as the fitness function. As can be seen from Table I, the proposed method outperforms NWFE and DBFE in terms of classification accuracies and improves the OA of DBFE and NWFE by almost 12 and 8 percent, respectively.

As can be seen from Table I, HGAPSO + SVM outperforms the other evolutionary-based feature selection techniques (e.g., GA + SVM and PSO + SVM) in terms of classification accuracy. On the other hand, PSO + SVM has the highest CPU processing time among other evolutionary-based feature selectors. The main reason of this shortcoming is that although PSO is a fast optimization method, it converged after a higher number of iterations. On the contrary, although the convergence of GA + SVM is faster than that of PSO + SVM and HGAPSO + SVM, it has the worst classification accuracies due to the premature convergence of the chromosomes.

Since all the evolutionary-based optimization methods are based on a random process, the selected features are different in different trials. In the experiments, the proposed approach selected 73–94 features in ten different trials. It should be noted that the proposed approach allows for the detection of the best distinctive features without requiring the number to be set *a priori* by the user.

D. Second Scenario

The obtained RMSE for $\mathrm{SVM_{RGB}}$, $\mathrm{SVM_{AP}}$, and $\mathrm{HGAPSO + SVM_{AP}}$ are 0.7669, 0.6461, and 0.6049, respectively. $\mathrm{HGAPSO + SVM_{AP}}$ provides the smallest RMSE among all techniques, which confirms the capability of the proposed method to detect the classes of interest. The main reason that the proposed approach outperforms $\mathrm{SVM_{AP}}$ is that although attribute profiles are a powerful technique to model spatial information of an image, they produces redundant features that can reduce the classification accuracies. However, by using the proposed technique, the most informative features can be selected leading to higher classification accuracies. Fig. 2 shows the input data, the manually produced reference map, and the maps obtained by $\mathrm{SVM_{RGB}}$, $\mathrm{SVM_{AP}}$, and $\mathrm{HGAPSO + SVM_{AP}}$, respectively. As can be seen, the proposed method detects more details from the road network as compared with the other approaches and outperforms $\mathrm{SVM_{RGB}}$ and $\mathrm{SVM_{AP}}$.

IV. Conclusion

In this letter, a new feature selection technique, which does not need to set the number of desired features *a priori*, has been introduced, based on the integration of GA and PSO. According to the experiments, the following can be concluded.

1) The proposed method can find informative bands in terms of classification accuracies in an acceptable CPU time.
2) The proposed method can be used for road detection.
3) In the novel feature selection approach, there is no need to set the number of output features since the proposed approach can automatically select the most useful features in terms of classification accuracies.
4) The proposed method is data set distribution independent, and for it, there is no need to initialize any parameters.
5) Since the proposed algorithm is based on evolutionary techniques, it is much faster than other well-known feature selection techniques that require an exhaustive process to select the most informative bands. Therefore, the new approach can work appropriately in a situation in which other feature selection techniques are not applicable.
6) Since SVM is considered as a fitness function in the proposed method, it can handle high-dimensional data with a limited number of training samples, when other feature selection techniques cannot proceed due to singularity problems of covariance matrices.

References

[1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ, USA: Wiley, 2003.
[2] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
[3] Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3374–3385, Nov. 2006.
[4] A. Daamouche, F. Melgani, N. Alajlan, and N. Conci, "Swarm optimization of structuring elements for VHR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1334–1338, Nov. 2013.
[5] F. Bergh and A. P. Engelbrecht, "A cooperative approach to particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 225–239, Jun. 2004.
[6] K. Premalatha and A. M. Natarajan, "Hybrid PSO and GA for global maximization," *Int. J. Open Probl. Comput. Math.*, vol. 2, no. 4, pp. 597–608, Dec. 2009.
[7] R. C. Eberhart and Y. Shi, "Comparison between genetic algorithms and particle swarm optimization," in *Evolutionary Programming VII*, V. W. Porto, Ed. Berlin, Germany: Springer-Verlag, 1998.
[8] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[9] D. Beasley, D. Bull, and R. Martin, "An overview of genetic algorithms," *Univ. Comput.*, vol. 15, no. 2, pp. 58–69, 1993.
[10] J. Kennedy and R. Eberhart, *Swarm Intelligence*. San Francisco, CA, USA: Morgan Kaufmann, 2001.
[11] P. Ghamisi, F. Sepehrband, J. Choupan, and M. Mortazavi, "Binary hybrid GA-PSO based algorithm for compression of hyperspectral data," presented at the 5th ICSPCS, Dec. 2011, pp. 1–8.
[12] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral–spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, May 2014.
[13] V. Mnih, "Machine learning for aerial image labeling," Ph.D. dissertation, Univ. Toronto, Toronto, ON, Canada, 2013.
[14] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, Sep. 2014.
[15] P. Ghamisi, J. A. Benediktsson, G. Cavallaro, and A. Plaza, "Automatic framework for spectral–spatial classification based on supervised feature extraction and morphological attribute profiles," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2147–2160, Jun. 2014.
[16] M. Mokhtarzade and M. V. Zoej, "Road detection from high-resolution satellite images using artificial neural networks," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 1, pp. 32–40, Feb. 2007.