Chapter 3

An unsupervised technique for optimal feature selection in attribute profiles for spectral-spatial classification of hyperspectral images

3.1 Introduction

Hyperspectral remote sensing images are characterized by huge spectral information, which makes them more powerful than panchromatic and multi-spectral images for discrimination among different land-cover objects [189]. While classification of pixels of an HSI, the surrounding pixels can provide useful information because of the homogeneous nature of objects. Thus, fusion of spectral and spatial information improves the classification results of HSI [55, 62, 110, 171, 172, 220, 247]. Morphological profiles (MPs) and attribute profiles (APs) have proved their ability to fuse the spectral and spatial information [14, 55, 58, 63, 82, 159]. Their effectiveness in analysis of remote sensing images makes them deserving candidates for consideration in classification of HSI [95].

MP for a gray-scale image is built by repeated use of geodesic opening and geodesic closing on the image with a structuring element (SE) of an increasing size [13, 14, 206, 227]. Thus, MP is a collection of filtered images along with the original image. In order to construct mathematical morphology based spectral-spatial

profile for HSI, MP is created for each of its component image in reduced dimension leading to an extended morphological profile (EMP) [14, 82]. Ample number of works in the literature have used EMP based spectral-spatial information [11, 13, 181, 188, 227]. Although, MPs are powerful tool to represent spectral-spatial information but are not free from limitations because of the fixed shape of SE, imability to filter based on gray-level properties of connected components and computational burden. These limitations were overcome in [56], where in place of conventional SE based MP, APs are built using attribute filters (AFs) [21].

AP for a gray-scale image is obtained by a sequence of AF operations performed on it. The following characteristics of AFs overcome the limitations of MPs. First, they allowed variable length neighborhood to consider spatial information. Second, they create a max-tree only once on which multiple filter operations can be performed by pruning according to filter parameters. This makes the construction of APs computationally efficient. And finally, since the connected components of the image are represented as nodes of constructed max-tree, the connected components can be easily filtered based on any gray-level property. In case of HSI, analogous to EMP, extended attribute profile (EAP) is constructed by concatenating the APs constructed for each image in reduced dimension of the original HSI [55]. APs allow the flexibility of using any measure as an attribute, which can be computed on gray-values of a connected component in an image [197]. Some examples of attribute are *area, standard deviation, moment of inertia, diagonal of bounding box, etc.* [197]. One can construct multiple EAPs for an HSI considering different attributes and if we stack the constructed EAPs together, it will lead to extended multi attribute profile (EMAP) [55, 58, 63, 79, 210]. The construction of AP, EAP and EMAP are described in Section 1.2.3.

Although, EMAP has proved itself to be effective in analysis of HSI, selecting filter parameters for constructing profiles is difficult. One solution to this is to construct the profile using all the filter parameters sampled in very small intervals. Such profiles are termed as entire extended multi attribute profile (EEMAP) [183], which has its own limitations in terms of very large dimensionality and high redundancy. As the dimensionality increases, relatively we need to increase the number of training samples required for training a specific classifier [88]. This rapid increase in the number of training samples required for density estimation introduces *curse of dimensionality* [12], which leads to the *Hughes phenomenon* in classification [115]. Furthermore, it also affects the computational complexity of the model. A challenging yet effective way to deal with these, is to reduce the dimension of EEMAP [103, 112]. The dimensionality reduction of EEMAP can

be done by adopting a feature selection technique that select optimal features in APs. To this end, a supervised approach is presented in [183]. This method uses GAs to select an optimal subset of feature from the EEMAP by maximizing the classification accuracy. In our knowledge no unsupervised attempt is made for this purpose.

This chapter presents an unsupervised feature selection technique for spectral-spatial classification of HSIs. In order to consider spectral-spatial feature for HSI classification, the proposed technique first constructs an EEMAP by taking into account variety of spatial information along with spectral features. For this, separate EAPs are constructed considering three different attributes namely, *area*, *standard deviation* and *moment of inertia*. The filter parameters corresponding to each attribute are selected from a wide range in reasonably small intervals. Finally, by concatenating these EAPs an EEMAP is constructed. Although, the EEMAP constructed by the proposed method contain variety of spatial information, it has very high dimensionality with ample redundancy that increases the computational cost and also may introduce curse of dimensionality problem. To mitigate these problems, in this chapter an unsupervised feature selection technique is presented that selects an optimal subset of feature from the constructed EEMAP for the classification of HSI. The proposed technique first generates a dissimilarity matrix to measure the dissimilarity between each pair of filtered images (features) in EEMAP using mutual information (MI). Then, to select an optimal subset of feature from EEMAP, GAs are exploited by defining a novel objective function based on the generated dissimilarity matrix. Note that the generation of dissimilarity matrix by the proposed technique is unsupervised in nature that makes the whole feature selection process unsupervised. The subset of the filtered images selected by the proposed technique is finally used for classification of HSI.

To demonstrate the effectiveness of proposed technique, experiments are conducted on four real hyperspectral data sets described in Appendix A using SVM classifier. The experimental results show the robustness of the proposed technique in terms of classification accuracy and computational time.

The remaining part of the chapter is organized as follows. The Proposed method is presented in Section 3.2. Section 3.3 presents experimental results obtained on the considered data sets. Finally, Section 3.4 concludes the chapter with some future scopes.

3.2 Proposed method

The proposed technique aims at improving the classification results of HSI by incorporating spatial information in the form of attribute profile. The overall architecture of proposed method is shown in Fig. 3-1. It has three main steps. In first step, an EEMAP is constructed which provides rich spectral-spatial information but at the same time increases the dimension substantially. In second step, a dissimilarity matrix is constructed by computing the distance among all the pairs of filtered images using normalized mutual information. In the third step, an optimal subset of features is selected from the constructed profile in an unsupervised way by using GAs with a novel fitness function designed based on the dissimilarity matrix. Finally, the selected features in the EEMAP are considered for classification of HSI. The detail of each step is given in the following subsections.

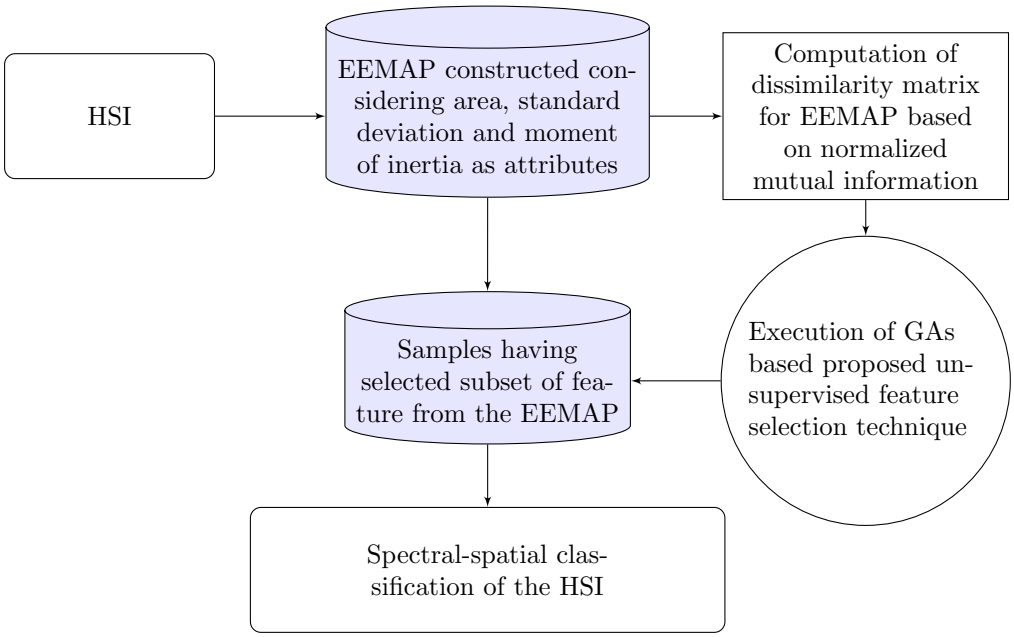


Figure 3-1: Block diagram of the proposed architecture.

3.2.1 Construction of entire extended multi attribute profile (EEMAP)

In order to incorporate variety of spatial information along with spectral information an EEMAP is constructed corresponding to an HSI. Construction of the EEMAP starts from dimensionality reduction of the original HSI. In this work the dimension of the original HSI is reduced by selecting l PCs with highest vari-

ance. For each selected PC, an AP is constructed using Eq. 1.25 considering filter parameters from a wide range, sampled in reasonably small interval. These constructed APs are concatenated to form one EAP. Fig. 1-8 shows the construction of an EAP. Considering multiple attributes namely *area*, *standard deviation* and *moment of inertia*, corresponding EAPs are constructed using Eq. (1.26). Basics on construction of EAPs is given in Section 1.2.3. The details of the constructed EAPs for this work are given below.

Area profile (EAP_a) is constructed for an HSI by considering *area* as an attribute while applying AFs on PCs. Area is measured in terms of number of pixels in the connected component of the image. In *area* filtering the connected components having greater area than the specified threshold value are preserved and the connected components with smaller area are merged with their local background component. By varying the values of the threshold in a wide range, a large number of filtered images are constructed and stacked to form an AP. AP constructed for each PC are concatenated to form an EAP_a.

Standard deviation profile (EAP_s) is constructed by considering *standard deviation* of gray-values in a connected component of the image. The *standard deviation* represents the homogeneity of the connected component. The connected components with higher standard deviation are preserved and those with lower standard deviation are merged with local background. A sequence of standard deviation thresholds are selected from a wide range to construct AP for each PC leading to another large set of filtered images representing EAP_s.

Moment of inertia profile (EAP_i) is another spectral-spatial profile constructed by considering the attribute *moment of inertia*. Again for this attribute, the threshold values are varied in small interval between 0 and 1 for each PC leading to large profile representing EAP_i.

Finally, the constructed EAPs are concatenated to form an EMAP using Eq. 1.27. Fig. 1-9 shows the construction of an EMAP. Since during the construction of each EAP, the filter parameter values are sampled densely from a wide range to incorporate maximum spatial information. Hence, the constructed EMAP is called EEMAP.

Although the constructed EEMAP has sufficient spectral-spatial information but because of its large dimensionality and existence of many redundant features, the curse of dimensionality and computation burden become a major issue for classification of HSI using the EEMAP. An effective method to handle these problems is to develop a feature selection technique that selects an optimal subset of filtered image from the constructed EEMAP for classification purpose.

In this work an unsupervised feature selection technique using GAs is developed. The proposed technique first generates a dissimilarity matrix that measures the dissimilarity between all the pairs of filtered images in EEMAP. Then to select an optimal subset of feature in EEMAP, GAs are exploited by defining a novel objective function based on the generated dissimilarity matrix.

3.2.2 Generation of dissimilarity matrix

The dissimilarity matrix is generated to measure the dissimilarity between all the pairs of filtered images in EEMAP. In the proposed work, normalized mutual information is used to measure the independence between two filtered images in EEMAP. The concept is based on information theory [52]. The information content of one filtered image I_i in EEMAP can be measured in terms of entropy as follows:

$$H_e(I_i) = -\sum_{g \in \mathcal{G}v} P(g) log(P(g)) \tag{3.1}$$

where $\mathcal{G}v$ is the set of distinct gray-values in the image I_i and $P(g)$ represents mass probability. Mutual information (MI) for two images I_i and I_j represents the correlation between them and can be computed as:

$$MI(I_i, I_j) = \sum_{g_x \in \mathcal{G}v_i} \sum_{g_y \in \mathcal{G}v_j} P(g_x, g_y) log\left(\frac{P(g_x, g_y)}{P(g_x)P(g_y)}\right) \tag{3.2}$$

where $\mathcal{G}v_i$ and $\mathcal{G}v_j$ are the set of distinct gray-levels in image I_i and I_j, respectively. $P(g_x, g_y)$ represents joint probability mass function. Mutual information is a non-negative value. $MI(I_i, I_j) = 0$ indicates two images I_i and I_j are totally independent and don't have any common information. The two images I_i and I_j share common information for higher values of $MI(I_i, I_j)$. As the value of MI increases, the correlation between two images increases. MI can increase till minimum of $\{H_e(I_i), H_e(I_j)\}$.

If the information content (i.e. entropy) of the two images are small, in that case, even if they are highly correlated, the obtained MI will remain small. Thus, for a fair comparison in terms of similarity, MI should be normalized. Normalized mutual information (NMI) is obtained as follows [160].

$$NMI(I_i, I_j) = \frac{2 \times MI(I_i, I_j)}{H_e(I_i) + H_e(I_j)} \tag{3.3}$$

NMI computed in Eq. (3.3) represents similarity between a pair of filtered images.

The dissimilarity between the pair of filtered images can be computed using Eq. 3.4 [160]:

$$D_{NMI}(I_i, I_j) = \left(1 - \sqrt{NMI(I_i, I_j)}\right)^2 \tag{3.4}$$

In this work, a dissimilarity matrix is generated by considering all the pairs of filtered images in EEMAP. If d is the number of features (filtered images) in EEMAP, the size of the dissimilarity matrix will be $d \times d$. Next, based on the generated dissimilarity matrix, GAs are exploited to select informative features in the EEMAP for HSI classification.

3.2.3 Feature selection using GAs

GAs are optimization tool, motivated from natural evolutionary process [99, 179]. In [209], GAs are successfully used for feature selection. Feature selection problem can be considered as an optimization problem where the task of the optimizer is to select an optimal subset of feature by optimizing a criterion. In order to take into account sufficient spatial information, the EEMAP constructed by the proposed technique contain huge number of features (filtered images) and many of them are redundant. To reduce the dimensionality of the constructed EEMAP, GAs are used to select a subset of feature in EEMAP that are independent and equally informative as the original features. The details of the proposed GAs based technique is given below:

Chromosome representation: A chromosome is a binary vector representing the index of a subset of filtered images in EEMAP. If one index is represented by b bits then size of the chromosome representing m filtered images is $b \times m$.

Population initialization: GAs generate multiple chromosomes randomly and are kept together in a group called population. The size of population is the number of chromosomes in the population.

Fitness evaluation: The computation of fitness value of the chromosomes in population is one of the most important component of the GAs. The function that is used to compute the fitness value of a chromosome is called fitness function. In this work, a novel fitness function is proposed with help of the generated dissimilarity matrix that provides the goodness of a set of filtered images represented by a chromosome. In greater details, let the EEMAP constructed by the proposed technique contains filtered images $\{I_1, I_2, ..., I_d\}$ and a chromosome in the popu-

lation represents the m filtered images in EEMAP as $(o_1, o_2, ..., o_m)$, where o_1, o_2, ..., o_m are the index of the filtered images in the EEMAP. To compute the fitness value of the chromosome, clustering of the filtered images in EEMAP is performed by considering I_{o_1}, I_{o_2}, ..., I_{o_m} as cluster representatives. An image in EEMAP is assigned to a cluster if the cluster's representative and the image itself have lowest dissimilarity value as compared to the other cluster representatives. Let Cl_{o_1}, Cl_{o_2}, ..., Cl_{o_m} be the clusters formed by considering I_{o_1}, I_{o_2}, ..., I_{o_m} as cluster representatives, respectively. The fitness function of a chromosome is defined as:

$$f(I_{o_1}, I_{o_2}, ..., I_{o_m}) = \left(\frac{\frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{|Cl_{o_i}|} \sum_{I_j \in Cl_{o_i}} D_{NMI}(I_{o_i}, I_j) \right)}{\frac{1}{m-1} \sum_{i=1}^{m} \left(\min_{\substack{j \neq i \\ j \in 1,2,...,m}} \{D_{NMI}(I_{o_i}, I_{o_j})\} \right)} \right) \tag{3.5}$$

where $|Cl_{o_i}|$ represents the number of images belong to the cluster Cl_{o_i}. The fitness function defined in Eq. (3.5) can be explained as:

$$f(I_{o_1}, I_{o_2}, ..., I_{o_m}) = \left(\frac{average\ of\ average\ intra\text{-}cluster\ distance}{average\ of\ minimum\ inter\text{-}cluster\ distance} \right) \tag{3.6}$$

The average intra-cluster distance signifies how compact one cluster is. In numerator, the average of average intra-cluster distance represents the compactness of all the clusters formed on EEMAP considering the filtered images of the chromosome as cluster representatives. Minimizing this value one can select compact clusters. On the other hand, inter-cluster distance is the distance between two nearest cluster representatives. In denominator, the average of minimum inter-cluster distance is computed by taking average of the minimum inter-cluster distance computed for each cluster representative. One can select diverse cluster representatives by maximizing this value. Thus, minimization of fitness function in Eq. (3.5) helps us to select an optimal subset of filtered images in the EEMAP.

Selection: After the fitness evaluation of each chromosome, the fittest chromosomes are selected and kept for next generation. The *stochastic uniform* selection strategy is adopted here.

Crossover: Similar to natural evolutionary process, new generation of child chro-

mosomes are generated based on the information exchange between parent chromosomes selected as fittest in previous iteration. The crossover is performed with a probability represented by crossover factor.

Mutation: Bits of a chromosome are flipped with certain probability and it is called mutation. This operation is performed on all the chromosomes present in the current generation.

Termination: The above steps are repeated until they converge to a stable situation.

Once the process terminates, the subset of filtered images represented by the fittest chromosome in the population is considered as optimal subset of filtered images in the EEMAP. Note that the fitness function defined in Eq. (3.5) is completely based on the dissimilarity matrix which is generated in an unsupervised way. Thus, the proposed spectral-spatial feature selection technique is unsupervised in nature. Moreover, since the fitness value of a chromosome is computed based on the dissimilarity matrix, once the dissimilarity matrix is generated, GAs takes less time to converge.

3.3 Experimental results

3.3.1 Design of experiments

The experiments are conducted on four real hyperspectral data sets described in Appendix A. In order to generate spectral-spatial features for HSI classification, first the EAPs considering the attributes *area*, *moment of inertia* and *standard deviation* are constructed. The EAPs are constructed by considering only the first five PCs of the hyperspectral data set. In the experiments, for all the considered data sets, the filter parameter values for the three attributes are varied as follows:

1. *area*: [1 9 25 49 81 121 169 225 289]
2. *moment of inertia*: [0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9]
3. *standard deviation*: [2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0 22.5]

From the above configuration one can see that the construction of EAP based on the *area* attribute (denoted as EAP_a) uses 9 threshold values. Since AP

for each PC is constructed by applying 9 thinning and 9 thickening operations leading to generate 19 filtered images (including PC itself) for single PC. Thus, EAP_a constructed with 5 PCs has 95 filtered images to reflect spectral-spatial information. Similarly, the constructed EAPs based on the *moment of inertia* attribute (denoted as EAP_i) and the *standard deviation* attribute (denoted as EAP_s) contain 95 filtered images each. After generating the EAPs, an EEMAP with above mentioned three attributes is constructed by considering all the filtered images in EAP_a, EAP_i, and EAP_s with no repetition of PCs. Thus, the constructed EEMAP contain 275 (i.e. $95 \times 3 - 10$) filtered images to represents spectral-spatial information of a hyperspectral data set. The profiles are generated using the tool available in [55].

The proposed technique exploits GAs to select optimal subset of filtered images from the constructed spectral-spatial profiles. In the present experiment for all the data sets the population size of GAs is taken as 40. Stochastic selection strategy is used to select fittest chromosomes from the mating pool. The crossover and mutation probability is set as 0.8 and 0.01, respectively.

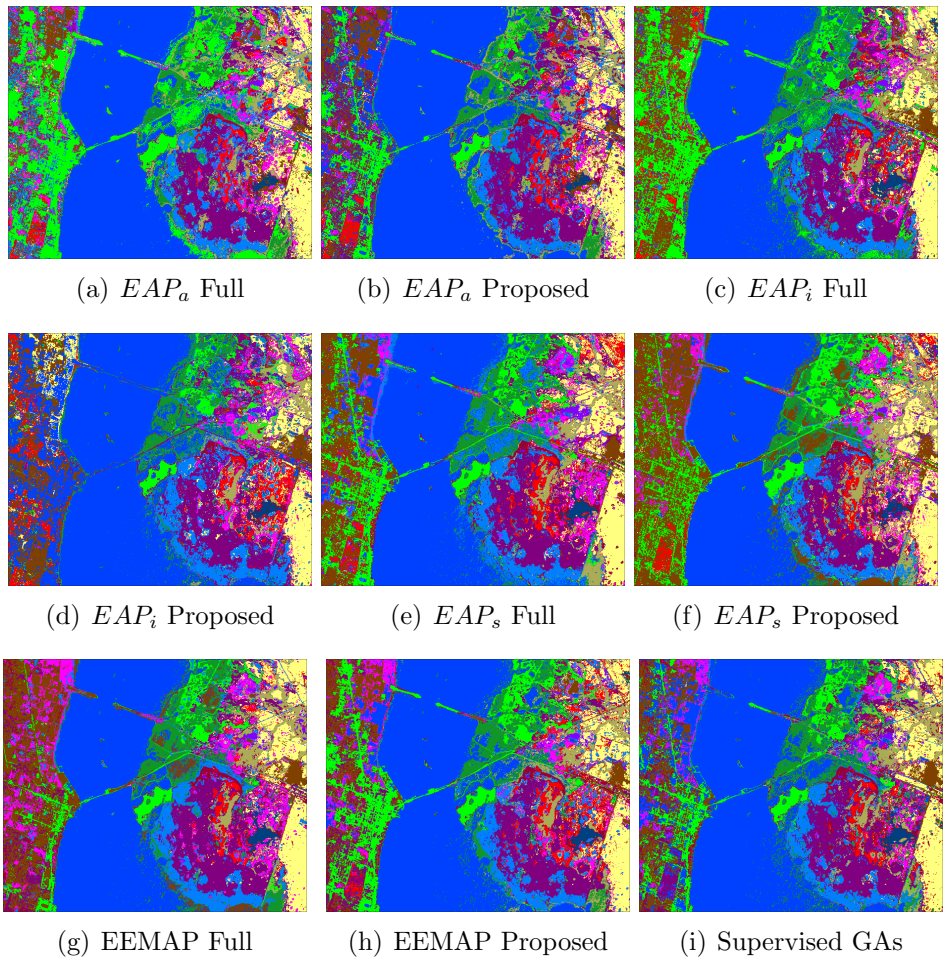
In order to show the effectiveness of the proposed technique the classification results obtained by the selected subset of feature are compared with the results obtained using full feature space. Moreover, it is also compared with GAs based supervised feature selection technique (referred as Supervised-GAs) presented in [183]. In contrast to the proposed unsupervised technique, the Supervised-GAs technique selects an optimal subset of feature in EEMAP by defining the fitness function of GAs in a supervised manner.

In this work one-against-all SVM classifier with radial basis function (RBF) is used for performing classification task. The SVM parameters $\{\sigma, C\}$ (i.e., the spread of RBF kernel and the regularization parameter) are obtained by applying grid search with 5-fold cross-validation. In this experiment, for all the considered data sets 30% of the available labeled samples from each class are randomly selected to train the SVM classifier and the rest 70% are used for testing purpose.

The subset of feature selected by the different techniques may vary from one run to another. To reduce this random effect for all the considered data sets the average classification accuracies obtained by 10 runs are used for comparison. The proposed technique is implemented in Maltab (R2015a). The LIBSVM library [38] is used to implement SVM classifier.

3.3.2 Results: KSC data set

In order to assess the effectiveness of the proposed technique, first experiment is carried out using the hyperspectral KSC data set (described in Appendix A.1). Table 3.1 reports the average class-wise accuracy, average overall accuracy ($\overline{OA}$), its standard deviation (std) and average kappa accuracy (kappa) obtained for ten runs. To show the importance of the spectral-spatial features generated by the different attribute profiles of the proposed technique, the classification results obtained by using only spectral features (i.e., 176 features) and the classification results obtained using the spectral-spatial features generated by EAP_a, EAP_i, EAP_s and EEMAP are reported in Table 3.1. From the table one can see that the spectral-spatial features generated by the different APs provided better results in every respect than using only spectral features for the KSC data set. Considering 176 spectral bands of the KSC data set the $\overline{OA}$ is obtained as 84.70%. Whereas, the $\overline{OA}$ obtained using the spectral-spatial features generated by EAP_a, EAP_i, EAP_s and EEMAP is 96.36%, 97.65%, 98.21% and 98.70%, respectively. From the table one can also observed that among the three constructed EAPs, the EAP_s provides higher $\overline{OA}$. Thus for the KSC data set, the *standard deviation* attribute provides better spatial information than the *area* and the *moment of inertia* attributes. Since the EEMAP is constructed by considering all the three attributes, it provides best classification result. Furthermore, to asses the effectiveness of the proposed feature selection technique the dimensionality of the constructed EEMAP is reduced from 275 to 60 and the dimensionality of all the three constructed EAPs are reduced from 95 to 40 by using the proposed GA based feature selection technique. The detailed classification results obtained for this reduced feature space are reported in Table 3.1. From this table one can see that the $\overline{OA}$ obtained by considering all the features in EAP_a, EAP_i, EAP_s and EEMAP is 96.36%, 97.65%, 98.21% and 98.70%, respectively. Whereas, the $\overline{OA}$ obtained for a subset of feature in EAP_a, EAP_i, EAP_s and EEMAP selected by the proposed technique is 95.08%, 97.23%, 96.97% and 98.07%, respectively. The classification results obtained in the reduced feature space is similar to the results obtained in full feature space. Finally, the performance of the proposed technique is compared with the existing Supervised-GAs technique. The results reported in Table 3.1 show that the subset of feature in EEMAP selected by the proposed unsupervised technique have similar discriminative power to the subset of feature in EEMAP selected by the Supervised-GAs technique. This shows the potential of the proposed feature selection technique. For visual analysis, Fig. 3-2 shows the classification maps generated for each of the spectral-spatial profiles in its full feature-space and in the subspace selected by the proposed method. The



(a) EAP_a Full (b) EAP_a Proposed (c) EAP_i Full

(d) EAP_i Proposed (e) EAP_s Full (f) EAP_s Proposed

(g) EEMAP Full (h) EEMAP Proposed (i) Supervised GAs

Figure 3-2: Classification maps of the best results obtained for KSC data set considering full feature space and the reduced feature space.

classification maps obtained in reduced subspace is not much different than those obtained in the full feature space. This confirms the effectiveness of the proposed method.

3.3.3 Results: University of Pavia data set

Second experiment is carried out using the hyperspectral University of Pavia data set (described in Appendix A.2). Table 3.2 reports the average class-wise accuracy, average overall accuracy ($\overline{OA}$), its standard deviation (std) and average kappa accuracy (kappa) obtained for ten runs. From this table one can see that the classification results obtained using the spectral-spatial features generated by EAP_a, EAP_i, EAP_s and EEMAP are much better than the classification results obtained by using only spectral features. In greater detail, considering 103 spectral bands of the University of Pavia data set the $\overline{OA}$ is obtained as 95.28%. Whereas, the

Table 3.1: Average class-wise accuracy, average overall accuracy ($\overline{OA}$), its standard deviation (std) and average kappa accuracy (kappa) obtained on ten runs by considering spectral and different spectral-spatial features (KSC data set)

Class No.	Full Spectral features (176)	EAP_a		EAP_i		EAP_s		EEMAP		
		Full (95)	Proposed (40)	Full (95)	Proposed (40)	Full (95)	Proposed (40)	Full (275)	Proposed (60)	Supervised -GAs (60)
1	96.32	99.02	98.89	99.25	98.31	99.92	99.19	99.98	99.55	99.62
2	90.65	93.53	91.47	91.82	91.76	96.35	95.24	96.76	94.12	96.18
3	94.80	97.26	96.48	97.04	94.92	97.37	96.87	98.44	98.04	97.54
4	12.61	90.34	88.75	93.64	92.61	94.15	87.61	95.80	95.63	93.64
5	64.69	92.83	88.41	94.87	94.42	95.22	92.12	96.99	96.46	96.37
6	9.94	87.31	82.56	96.25	97.00	98.38	93.44	98.56	96.13	98.38
7	80.27	93.29	91.51	95.34	94.38	97.81	94.79	98.08	94.79	95.34
8	89.47	93.84	93.38	98.71	99.24	99.14	98.05	99.44	98.87	99.40
9	90.00	97.80	94.64	99.67	99.56	99.75	98.71	99.86	99.40	99.48
10	95.34	95.27	94.31	95.87	94.35	95.51	94.81	97.53	95.94	96.08
11	96.62	98.87	98.33	98.81	98.70	99.04	99.25	99.45	99.39	99.52
12	81.85	95.28	93.30	96.39	96.05	96.79	95.28	97.16	96.93	97.27
13	99.12	99.80	99.72	99.54	99.49	99.40	99.46	99.43	99.52	99.12
$\overline{OA}$	84.70	96.36	95.08	97.65	97.23	98.21	96.97	98.70	98.07	98.20
kappa	0.8289	0.9595	0.9451	0.9739	0.9691	0.9800	0.9662	0.9855	0.9785	0.9799
std	0.6445	0.3359	0.3519	0.2179	0.4642	0.2690	0.2853	0.2065	0.3026	0.2814

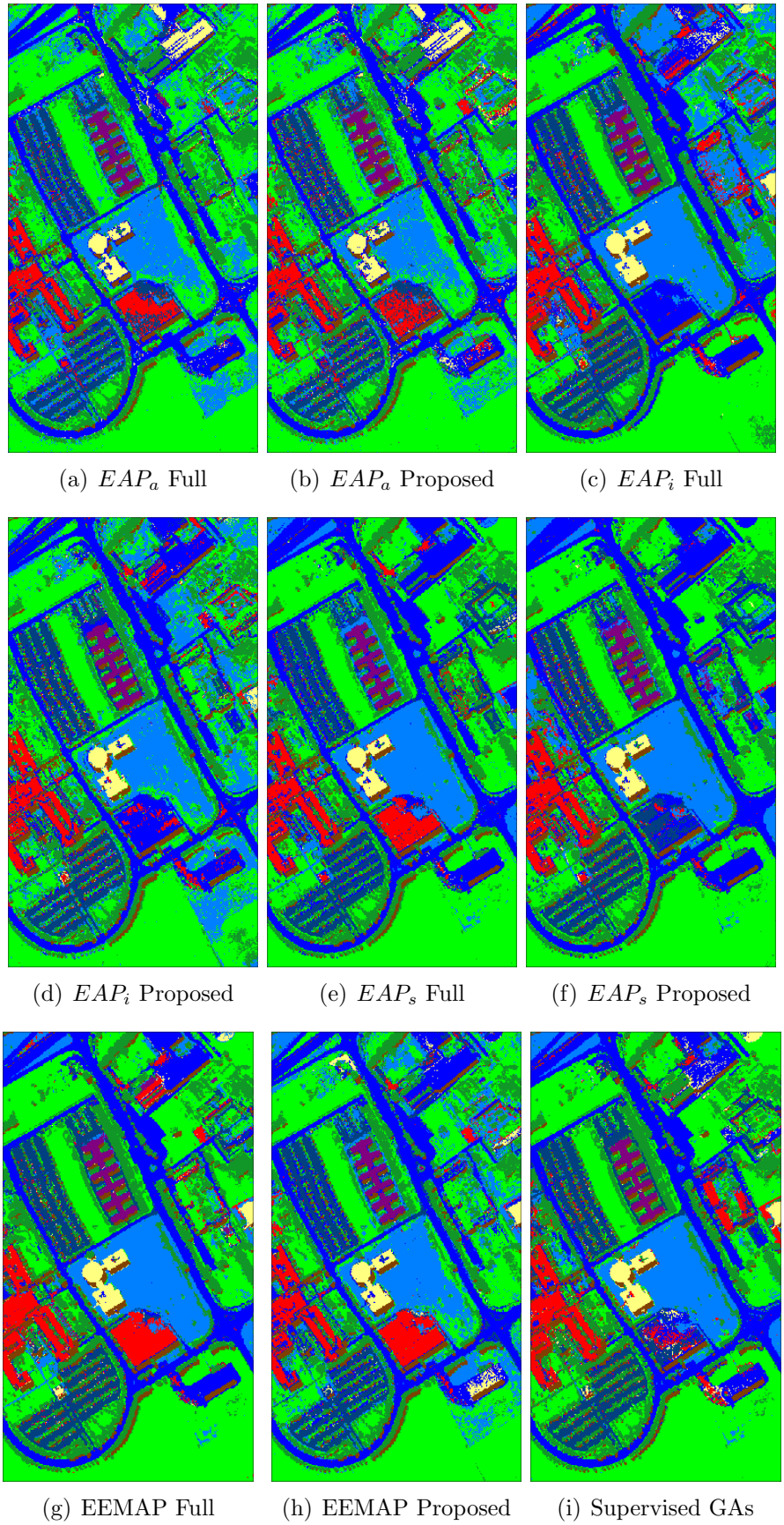
$\overline{OA}$ obtained using the spectral-spatial features generated by EAP_a, EAP_i, EAP_s and EEMAP is 97.63%, 99.69%, 99.69% and 99.91%, respectively. This shows the importance of the spatial-spectral features generated in the proposed technique. From the table one can also observe that among the three constructed EAPs, the EAP_i and the EAP_s provides better $\overline{OA}$ than EAP_a. Thus, for University of Pavia data set, the standard deviation and the moment of inertia are better attributes for taking into account the spatial information. Since the EEMAP is constructed by considering all the three attributes, it provides best classification result. To assess the effectiveness of the proposed GA based feature selection technique the classification results obtained using the subset of feature selected by the proposed technique are also reported in Table 3.2. From the results one can see that the $\overline{OA}$ obtained by considering all the features in EAP_a, EAP_i, EAP_s and EEMAP is 97.63%, 99.69%, 99.69% and 99.91%, respectively. Whereas, the $\overline{OA}$ obtained for a subset of feature in EAP_a, EAP_i, EAP_s and EEMAP selected by the proposed technique is 96.36%, 98.95%, 99.69% and 99.52%, respectively. Analogous to first experiment, in this experiment, the classification results obtained in the reduced feature space is also similar to the results obtained in full feature space. Finally, the performance of the proposed technique is compared with the existing Supervised-GAs technique. The results reported in Table 3.2 show that the subset of feature in EEMAP selected by the proposed unsupervised technique have similar discriminative power in comparison to the subset of feature in EEMAP selected by the Supervised-GAs technique. This again shows the effectiveness of the proposed feature selection technique for the University of Pavia data set. For visual analysis the classification maps obtained for all the EAPs in full feature space and reduced feature space is shown in Fig. 3-3. The classification maps obtained in reduced feature space are quite similar to those obtained in full feature space as well as to those obtained by the state-of-the-art Supervised-GAs. This again confirms the effectiveness of the proposed method.

3.3.4 Results: Indian Pines data set

For validating the effectiveness of the proposed technique, the third experiment is carried out using hyperspectral Indian Pines data set (described in Appendix A.3). Table 3.3 reports the classification result obtained for ten runs. To show the importance of spectral-spatial features generated by the different attribute profiles used in the proposed technique, the classification results obtained using only spectral features (i.e., 185 features) and the classification results obtained using the spectral-spatial feature generated by EAP_a, EAP_i, EAP_s and EEMAP are

Table 3.2: Average class-wise accuracy, average overall accuracy ($\overline{OA}$), its standard deviation (std) and average kappa accuracy (kappa) obtained on ten runs by considering spectral and different spectral-spatial features (University of Pavia data set).

Class No.	Full Spectral features (103)	EAP_a		EAP_i		EAP_s		EEMAP		
		Full (95)	Proposed (40)	Full (95)	Proposed (40)	Full (95)	Proposed (40)	Full (275)	Proposed (60)	Supervised -GAs (60)
1	95.07	97.72	96.19	99.72	99.58	99.53	99.59	99.90	99.79	99.80
2	98.26	98.89	98.37	99.88	99.32	99.92	99.88	99.98	99.73	99.94
3	83.26	94.57	92.23	99.14	99.00	99.15	99.38	99.74	99.24	99.65
4	95.90	97.67	97.10	98.69	97.80	98.91	98.88	99.73	99.26	99.51
5	99.17	99.80	99.77	99.76	99.69	99.79	99.67	99.86	99.76	99.76
6	92.95	94.95	91.39	99.70	96.71	99.88	99.79	99.98	98.93	99.91
7	88.42	94.15	92.28	99.90	99.52	99.33	99.75	99.83	99.27	99.79
8	90.04	96.46	94.54	99.59	99.15	99.48	99.56	99.74	99.08	99.73
9	99.62	99.56	99.20	100	99.85	99.89	99.79	99.97	99.85	100
$\overline{OA}$	95.28	97.63	96.36	99.69	98.95	99.69	99.69	99.91	99.52	99.84
kappa	0.9374	0.9686	0.9517	0.9959	0.9860	0.9958	0.9959	0.9988	0.9937	0.9979
std	0.0937	0.0907	0.1027	0.0352	0.0713	0.0255	0.0235	0.0132	0.0355	0.0304



(a) EAP_a Full (b) EAP_a Proposed (c) EAP_i Full

(d) EAP_i Proposed (e) EAP_s Full (f) EAP_s Proposed

(g) EEMAP Full (h) EEMAP Proposed (i) Supervised GAs

Figure 3-3: Classification maps of the best results obtained for University of Pavia data set considering full feature space and the reduced feature space.

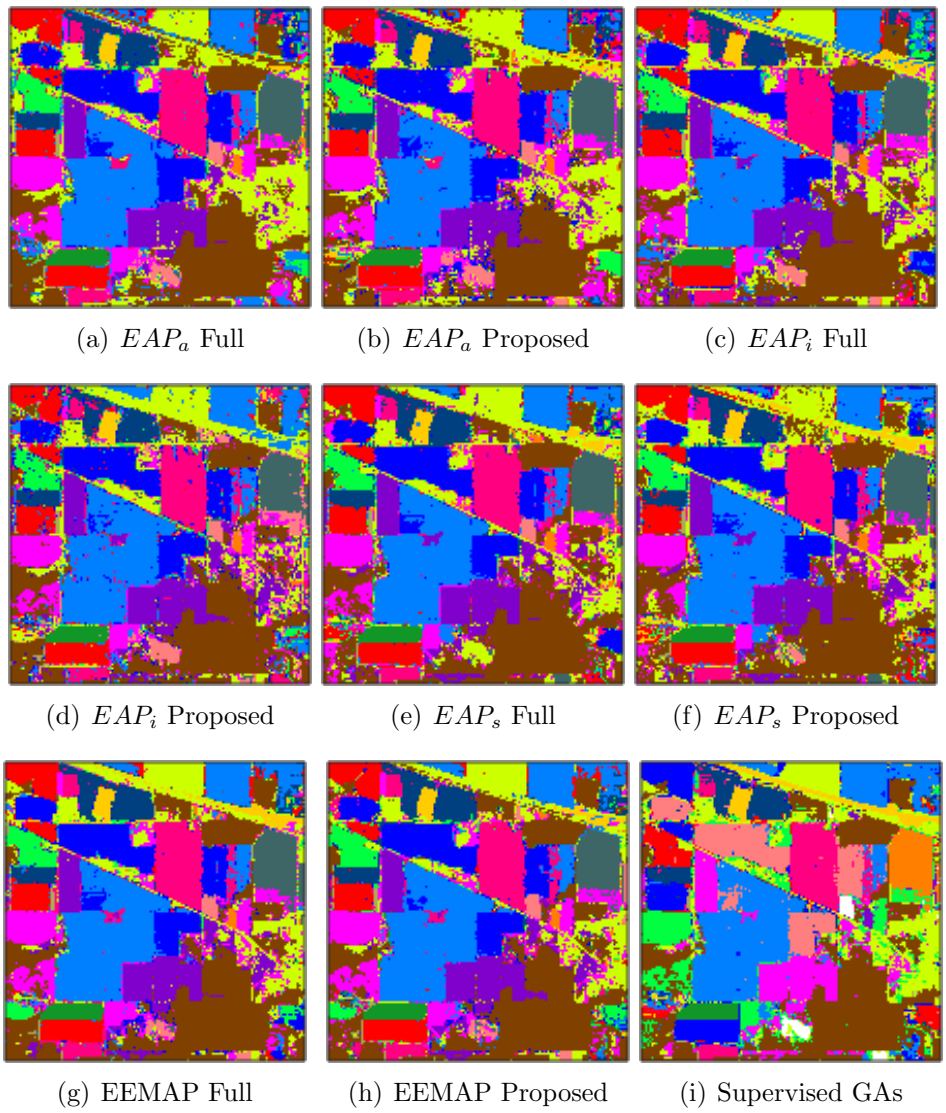
reported in Table 3.3. From this table one can see that the $\overline{OA}$ obtained for Indian Pines data set with 185 spectral bands is 91.03%. Whereas, the $\overline{OA}$ obtained for the spectral-spatial features generated by EAP_a, EAP_i, EAP_s and EEMAP is 96.64%, 96.42%, 97.18% and 97.93%, respectively. Thus, approximately 5% of $\overline{OA}$ is increased by using the spectral-spatial feature generated by the different APs instead of using only spectral feature. From the table one can also see that for Indian Pines data set, among the three attributes used to construct EAPs, the *standard deviation* attribute incorporated better spatial information than the *area* and the *moment of inertia* attributes. Like other two data sets for this data set also, the EEMAP which is constructed by considering all the three attributes provided best classification result. Finally to assess the effectiveness of the proposed GA based feature selection technique, the classification results obtained in the reduced feature space is also reported in Table 3.3. From the table one can see that the $\overline{OA}$ obtained in full feature space for EAP_a, EAP_i, EAP_s and EEMAP is 96.64%, 96.42%, 97.18% and 97.93%, respectively. Whereas, the $\overline{OA}$ obtained in reduced feature space (using proposed technique) of EAP_a, EAP_i, EAP_s and EEMAP is 95.71%, 93.20%, 96.10% and 97.51%, respectively. Maintaining the consistency from earlier experiments, the classification results obtained in the reduced feature space is similar to the results obtained in full feature space. Moreover, the performance of the proposed technique is compared with the existing Supervised-GAs technique. The results reported in Table 3.3 show that the subset of feature in EEMAP selected by the proposed unsupervised technique have similar discriminative power to the subset of feature in EEMAP selected by the Supervised-GAs technique. Furthermore, The classification maps are shown in Fig 3-4 obtained for all the EAPs in full and reduced feature space. The maps obtained in reduced feature space and the full feature space are similar. This established the robustness of proposed feature selection technique.

3.3.5 Results: University of Houston data set

For validating the effectiveness of the proposed technique, the fourth experiment is carried out using University of Houston data set (described in Appendix A.4). The classification results obtained using only spectral features (i.e., 144 features) and the classification results obtained using the spectral-spatial feature generated by EAP_a, EAP_i, EAP_s and EEMAP are reported in Table 3.4 after 10 runs of the experiments. The results demonstrate that for this data set also the proposed method successfully reduced the dimension maintaining the information content. For the EEMAP with 275 features, the $\overline{OA}$ is 98.79%, whereas the $\overline{OA}$ obtained for

Table 3.3: Average class-wise accuracy, average overall accuracy ($\overline{OA}$), its standard deviation (std) and average kappa accuracy (kappa) obtained on ten runs by considering spectral and different spectral-spatial features (Indian Pines data set).

Class No.	Full Spectral features (185)	EAP_a		EAP_i		EAP_s		EEMAP		
		Full (95)	Proposed (40)	Full (95)	Proposed (40)	Full (95)	Proposed (40)	Full (275)	Proposed (60)	Supervised -GAs (60)
1	90.94	96.56	96.25	93.13	92.19	93.75	91.88	97.19	96.25	96.25
2	89.28	94.97	93.71	93.95	89.54	95.79	94.45	96.64	94.98	94.93
3	86.21	97.57	97.07	96.51	90.05	98.07	97.47	99.12	98.24	98.38
4	85.12	95.60	93.61	91.81	86.39	93.37	93.31	95.48	95.12	96.08
5	95.27	96.12	95.12	96.01	93.99	95.95	95.98	97.46	96.75	95.62
6	98.36	99.55	99.59	99.61	98.90	99.94	99.80	99.96	99.71	99.69
7	87.50	95.50	97.00	91.50	94.50	97.50	97.00	96.50	95.50	96.50
8	99.73	99.97	100	99.85	99.73	100	99.79	100	99.94	100
9	77.14	90.71	93.57	85.00	86.43	95.00	92.86	90.71	96.43	90.00
10	86.88	95.35	93.79	93.63	90.26	95.26	95.66	95.65	95.54	95.31
11	90.73	96.04	95.38	96.96	94.30	97.30	96.67	97.86	98.07	97.39
12	91.47	92.67	90.19	92.31	88.43	94.24	92.27	95.69	95.49	94.43
13	99.44	99.09	99.09	99.23	98.88	99.44	99.09	99.23	99.16	99.16
14	95.55	99.15	98.56	99.58	97.23	99.29	97.39	99.79	99.60	99.55
15	76.59	95.96	93.00	97.48	89.56	96.44	87.48	98.96	97.52	97.59
16	84.77	98.46	98.77	97.54	94.15	97.69	96.46	98.77	98.46	99.38
$\overline{OA}$	91.03	96.64	95.71	96.42	93.20	97.18	96.10	97.93	97.51	97.24
kappa	0.8977	0.9617	0.9511	0.9591	0.9225	0.9679	0.9555	0.9764	0.9716	0.9685
std	0.2790	0.2364	0.2883	0.2914	0.3079	0.2954	0.2945	0.2535	0.2655	0.3452



(a) EAP_a Full (b) EAP_a Proposed (c) EAP_i Full

(d) EAP_i Proposed (e) EAP_s Full (f) EAP_s Proposed

(g) EEMAP Full (h) EEMAP Proposed (i) Supervised GAs

Figure 3-4: Classification maps of the best results obtained for the Indian Pines data set considering full feature space and the reduced feature space.

profile after feature selection using the proposed method is 98.07%. Similarly, for other spectral-spatial profiles also, the classification results obtained in the reduced feature space is similar to the results obtained in full feature space. Moreover, the $\overline{OA}$ obtained for proposed method and the Supervised-GAs technique are also similar. This shows that the proposed unsupervised method is as effective as the Supervised-GAs. Furthermore, The classification maps are shown in Fig. 3-5 obtained for all the EAPs in full and reduced feature space. The maps obtained in reduced feature space and the full feature space are similar. This established the robustness of proposed feature selection technique.

Table 3.4: Average class-wise accuracy, average overall accuracy ($\overline{OA}$), its standard deviation (std) and average kappa accuracy (kappa) obtained on ten runs by considering spectral and different spectral-spatial features (University of Houston data set)

Class No.	Full Spectral features (144)	EAP_a		EAP_i		EAP_s		EEMAP		
		Full (95)	Proposed (40)	Full (95)	Proposed (40)	Full (95)	Proposed (40)	Full (275)	Proposed (60)	Supervised -GAs (60)
1	99.01	98.36	97.49	98.78	97.99	98.13	97.57	99.08	98.88	98.58
2	98.77	97.49	98.50	98.95	97.68	98.80	98.83	98.55	98.20	99.01
3	100.00	99.92	99.61	100.00	100.00	100.00	99.96	100.00	100.00	99.94
4	99.45	98.79	98.05	99.55	99.36	98.68	98.67	99.40	99.07	99.54
5	99.41	99.57	99.40	99.36	99.26	99.94	99.71	99.91	99.70	99.97
6	97.89	98.63	98.59	99.25	99.43	98.50	98.68	99.21	99.21	98.68
7	94.72	97.38	97.57	99.03	98.56	98.38	96.06	99.23	99.01	98.81
8	95.04	95.71	93.81	97.34	95.05	96.75	93.40	98.46	97.05	97.67
9	91.66	89.82	89.27	97.71	93.87	96.34	95.43	98.28	98.12	98.15
10	97.83	94.71	94.45	98.79	96.05	98.61	97.79	98.66	98.27	99.42
11	95.90	89.81	88.97	97.44	96.18	96.49	95.63	97.94	96.91	97.27
12	95.17	89.95	88.76	97.30	91.15	96.27	91.83	98.18	95.63	98.03
13	73.45	81.01	78.38	82.96	75.82	78.57	77.96	94.30	89.73	92.16
14	99.40	99.77	99.07	99.63	98.67	99.83	96.87	99.97	99.80	99.87
15	99.33	99.87	99.81	100.00	99.85	99.98	99.31	100.00	99.85	99.76
$\overline{OA}$	96.34	95.36	94.80	98.14	96.31	97.51	96.26	98.79	98.07	98.59
kappa	0.9604	0.9499	0.9438	0.9799	0.9601	0.9730	0.9595	0.9869	0.9791	0.9847
std	0.1364	0.1560	0.2487	0.2650	0.2062	0.2033	0.2430	0.1161	0.1282	0.1940

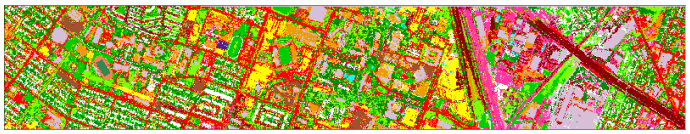
(a) EAP_a Full

(b) EAP_a Proposed

(c) EAP_i Full

(d) EAP_i Proposed

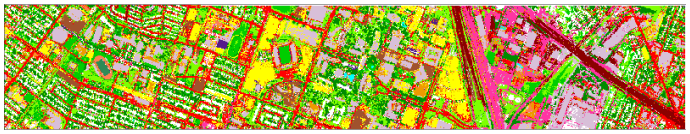
(e) EAP_s Full



(f) EAP_s Proposed

(g) EEMAP Full

(h) EEMAP Proposed



(i) Supervised GAs

Figure 3-5: Classification maps of the best results obtained for the University of Houston data set considering full feature space and the reduced feature space.

3.3.6 Results: computational time

The fifth experiment deals with the computational time required by the proposed technique using the same experimental setting as described in Section 3.3.1. All the experiments are carried out on a PC (Intel(R) Core(TM) i5-6500 CPU @ 3.20 GHz with 4 GB RAM). In this experiment, the computational time taken by the proposed technique is divided into two parts DM and FS. DM represents the average computational time taken by the proposed technique to generate the dissimilarity matrix of the constructed EEMAP. FS represents the average computational time taken by the GAs to select optimal subset of feature in EEMAP using the generated dissimilarity matrix. Table 3.5 reported the values of DM and FS for the four considered data sets. From the table one can see that the values of DM are quite higher than the values of FS. The values of DM can further be reduced by generating the dissimilarity matrix using less number of samples. Table 3.5 shows the average overall accuracy ($\overline{OA}$) and the computational time (in seconds) taken to generate the dissimilarity matrix using 100%, 20% and 10% of the samples. From the table one can see that the $\overline{OA}$ obtained by the proposed technique does not change significantly when the dissimilarity matrix is generated using 100%, 20% and 10% of the samples but significantly reduced the values of DM. Thus, the computational time required to generate the dissimilarity matrix can be significantly reduced by using a subset of samples instead of considering all the samples. Since the proposed technique exploits dissimilarity matrix to select optimal subset of feature, the value of FS does not depend on the number of samples. From the experimental results in Table 3.5 one can see that the proposed technique takes few minutes to select an optimal subset of feature in EEMAP.

According to our knowledge, there is no unsupervised feature selection technique existing in remote sensing literature that select optimal spectral-spatial features from the generated EEMAP. To further assess the effectiveness of the proposed technique, the computational time taken by it is compared with the time taken by the existing Supervised-GAs technique. Table 3.5 also reports the computational time taken by the Supervised-GAs technique. From the table one can see that the computation time taken by the Supervised-GAs is much higher than the proposed technique. For example the time taken by Supervised-GAs technique for the KSC, the University of Pavia, the Indian Pines and the University of Houston data set is 1.6 hours, 58 hours, 16 hours and 10 hours, respectively. Whereas, the proposed technique takes few minutes for the same. Since, in Supervised-GAs the fitness function of GAs are defined by considering the classification accuracy of the classifier. The classification task need to be performed at every iteration

of GAs to compute the fitness value of a chromosome in the population, which is highly computational demanding. In contrast, the proposed technique is not only unsupervised in nature it also reduced the computational time significantly by defining the fitness function of GAs using dissimilarity matrix.

Table 3.5: Details of average computational time (in seconds) required by the proposed and the Supervised-GAs techniques for different hyperspectral data sets

Data sets	Proposed				Supervised -GAs
	% of samples	$\overline{OA}$	Execution Time		Execution Time(Sec)
			DM	FS	
KSC	100	98.07	5304.5	82.0	5846
	20	98.03	1922.6		
	10	97.93	1352.7		
University of Pavia	100	99.52	3787.3	68.0	211235
	20	99.52	1605.3		
	10	99.49	1269.3		
Indian Pines	100	97.51	1121.6	80.3	59102
	20	97.37	1039.3		
	10	97.14	798.8		
University of Houston	100	98.02	13539	84.3	37187
	20	97.89	3135		
	10	97.74	1914		

3.3.7 Results: curse of dimensionality

The last experiment is carried out to see the importance of the feature selection technique in order to mitigate the curse of dimensionality problem. As stated in section 3.1, for constant number of training samples the density estimation of patterns suffer as the dimensionality increases. This phenomenon is referred as the curse of dimensionality [12], which leads to the *Hughes phenomenon* in classification [115]. A challenging yet effective strategy to deal with curse of dimensionality is to reduce the dimension of the patterns [103, 112]. The curse of dimensionality problem is not visible in the results shown in the first, second, third and fourth experiments. This could have happened because of the adequate number of samples being used to train the classifier or the dimensionality of the generated samples are not high enough. In order to see the curse of dimensionality effect, in this experiment for the Indian Pines data set, an EEMAP is constructed considering the three attributes namely, area, moment of inertia and standard deviation on the first 10 PCs of the original HSI. The filter parameter values are sampled in much smaller intervals from a wide range for each attribute. The filter parameter values

for the *area* attribute are 1^2, 2^2,..., 17^2 which lead to 17 thinning and 17 thickening operation. For the *standard deviation* attribute the filter parameter values are sampled in the range [2.5 25] with step-size of 1, leading to 22 thinning and 22 thickening operations. For the *moment of inertia* attribute the filter parameter values are sampled in the range [1 9] with step-size of 0.5, leading to 17 thinning and 17 thickening operations. With this configuration on 10 PCs of Indian Pines data set, the constructed EEMAP has a dimension of 1150. To train the SVM classifier, 5 labeled samples of each class are randomly selected from the EEMAP. Fig. 3-6 shows the average classification accuracy versus the number of features selected in EEMAP using the proposed unsupervised feature selection method. From the figure, one can see that first the classification accuracy improves by increasing the number of features till 200 and then decreases consistently. Similar behavior is also observed for other data sets. Thus, when dealing with high dimensional data with limited available labeled samples feature selection plays an important role to mitigate the curse of dimensionality problem.

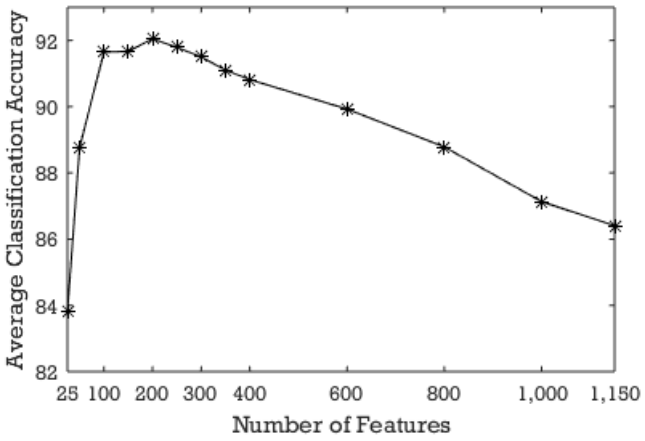


Figure 3-6: Average classification accuracy obtained using different number of features selected in the EEMAP for the Indian Pines data set.

3.4 Discussion and conclusion

In this chapter, we have proposed a method for improving the classification results of HSI by incorporating spatial information using attribute profiles in its optimal feature space. For fusing spectral and spatial information a large EEMAP is constructed for the HSI considering filter parameter values sampled in very small interval from a wide range. Although, the EEMAP constructed by the proposed method contain variety of spatial information, it has very high dimensionality with ample redundancy that increases the computational cost and also may introduce

curse of dimensionality problem. To mitigate these problems here an unsupervised feature selection technique is presented that selects an optimal subset of feature from the constructed EEMAP for spectral-spatial classification of HSI. The proposed technique first generates a dissimilarity matrix to measure the dissimilarity between pair of filtered images in EEMAP using mutual information. Then, to select an optimal subset of feature from the EEMAP, GAs are exploited by defining a novel objective function based on the generated dissimilarity matrix. The selected subset of features are considered for classification of the HSI. An experiment demonstrates the importance of feature selection from a large EEMAP to avoid *curse of dimensionality* problem.

Experimental results obtained for the four real hyperspectral remote sensing data sets show the robustness of the proposed method in terms of computational time and classification accuracies. In the experiments, the proposed unsupervised feature selection technique is applied on each of the generated spectral-spatial profiles and the results demonstrate that the selected subset of filtered images provided almost similar classification results to those obtained by considering full profile. Moreover, the proposed technique is compared with the existing Supervised-GAs technique. This comparison not only shows the effectiveness of the proposed technique in term of classification accuracy, it also shows that the proposed technique takes significantly less amount of computational time than the existing one.