A class-based approach to classify PolSAR imagery using optimum classifier

Bahram Safaee and Mahmod Reza Sahebi

Department of Photogrammetry and Remote Sensing, Faculty of Geodesy & Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran

ABSTRACT
Classification of Polarimetric SAR (PolSAR) imagery is still one of the challenges in active remote sensing applications. Although a large number of features and different classifiers have been proposed, no unique approach has been found to satisfy all of the images and classes yet. In this paper, given the extracted features from PolSAR data, a new class-based feature selection (CBFS) algorithm is proposed to find the most optimum features for each class. Maximizing the discrimination of each class from the others is the main contribution of the CBFS which yields distinctive features. The selected features are then employed by classifiers to generate different classification results. Finally, a new approach is developed to combine these classification results to produce the final land cover map. Five different classifiers of Wishart Maximum Likelihood, Gaussian Maximum Likelihood, Support Vector Machine, Multi Layer Perceptron and Fuzzy Inference System are also used for classification. Given the CBFS results, two different Radarsat-2 and AirSAR PolSAR data were classified. Selected features led to improvement of about 5% in producer accuracies in comparison with two well-known Genetic Algorithm Feature Selection (GAFS) and Prototype Space Feature Selection (PSFS) methods. Moreover, Comparison results demonstrate that the fuzzy classifiers could improve the accuracies about 3% if they are suitably constructed and well designed. The achieved higher overall accuracy for the final classified map shows the effectiveness of the proposed approach over the other compared classification procedures.

Introduction
Full Polarimetric Synthetic Aperture Radar (PolSAR) images have been successfully used in different applications such as forestry, agriculture, urban, land cover/land use classification and etc (Lee & Pottier, 2009). Most of these applications require classified PolSAR data. Generally, producing an accurate classified data requires discriminative features as well as a proper classifier (Haddadi, Sahebi, & Mansourian, 2011; Maghsoudi, Collins, & Leckie, 2012).

Full PolSAR sensors collect complex values of backscattering information resulted from the interaction of microwave pulses with ground objects in HH, HV, VH and VV polarizations (Lee & Pottier, 2009). Considering the backscattering values, Covariance and Coherency matrices as original features of PolSAR images, lots of other features have been developed and employed to classify PolSAR data in a supervised and unsupervised manner. These features mainly include SAR discriminators, coherent and incoherent target decomposition parameters which most of them try to classify pixels into different physical backscattering mechanisms of single bounce, double bounce, volume and helix scattering. Detailed explanation about these parameters could be found in (Lee & Pottier, 2009).

The possibility of generating a large number of features and the limitation of simultaneous use of them in classification are considered as the main challenge in PolSAR data classification (Maghsoudi et al., 2012). Increase of feature space dimension makes the ill-posed condition in the estimation of classifier parameters (especially in the parametric classifiers) (Theodoridis & Koutroumbas, 2008). Therefore, simultaneous use of all features in the classification procedure is not recommended and using a method for dimensionality reduction is known as an essential solution in this area. So, an effective strategy to classify the PolSAR data has been generally focused on the simultaneous identification of optimum features and a proper classifier.

There have been a number of studies aimed to PolSAR data classification based on the above mentioned strategy. For example, using a convex optimization method to find the optimal features among the 61 PolSAR features (Chen, Yang, Liu, & Sun, 2010), and applying the binary genetic algorithm to find the optimal features among 57 and 105 features (Haddadi et al., 2011; Salehi, Sahebi, & Maghsoudi, 2014) are two deterministic and non-deterministic approaches which have been used in the field of PolSAR feature selection. Furthermore, the Fisher Discrimination Analysis
(FDA) has also been used as a supervised dimensionality reduction procedure in (Maghsoudi et al., 2012). Exploitation of a non-parametric procedure to generate between class and within class matrices are the main contributions of their research. Sparse support vector machine feature selection algorithm has also been proposed by (Bai, Peng, Yang, Chen, & Yang, 2014) to find the optimum features among 123 features for the classification of the PolSAR data.

As mentioned before, finding a proper classifier is also another challenge in pattern recognition problems. In this regard, there have been a number of researches for PolSAR data classification. Different classifiers such as MLP Neural Networks (Haddadi et al., 2011; Salehi et al., 2014), SVM (Aghababaei, Amini, & Tzeng, 2013; Bai et al., 2014; Chen et al., 2010; Maghsoudi et al., 2012), Fuzzy Inference systems (FIS) (Hellmann & Jäger, 2002), Decision Tree (Qi, Yeh, Li, & Lin, 2012), Random Forest (Du, Samat, Waske, Lie, & Li, 2015), Wishart (Lee, Grunes, & Kwok, 1994) and maximum likelihood (ML) (Lee & Pottier, 2001) have been used as classifier.

So far, all developed feature selection algorithms have been designed to find the features that are guaranteed to act as the discriminative features simultaneously for all of the classes. Meanwhile, if a feature has similar behaviour for all classes except one class, in this feature selection mechanism will have low ranking scores and its capabilities of extracting a special class from others are simply neglected. The potential of these features that are capable to discriminate one class from others has not been yet investigated. In this paper, we tried to benefit from the potential of such features in a new class-based feature selection (CBFS) method and use them in an innovative classification procedure.

The proposed CBFS has been inspired from a well-known and innovative feature selection method called Prototype Space Feature Selection (PSFS), which has been developed for hyperspectral image processing (Mojaradi, Abrishami-Moghaddam, Zoej, & Duin, 2009). The proposed CBFS method not only employs high capabilities of features for a single class, but it also uses a wider domain of available features in the classification process. In this method, an independent set of features is used to extract each class from others and finally the classification results of each class are overlaid to build the final classified map in an innovative manner. The rest of the paper is organized as follows. Section 2 introduces the feature selection used in this work, classification methods and proposed algorithm to generate final classified map. Dataset and pre-processing are provided in Section 3. Section 4 is dedicated to experimental results and discussions and finally, the conclusion is given in Section 5.

Methodology

The Framework of this study is shown in Figure 1. The main contributions of this paper are also starred and highlighted in this figure.

The right part of Figure 1 is dedicated to performance evaluation purposes and the left part shows the proposed method in this paper. As mentioned in introduction, the common methods of PolSAR data classification include a feature selection algorithm and a classifier. In this paper (right part of Figure 1), eight combinations of feature selection and classification were used to evaluate the performance of the proposed method. Two of these combinations have also been introduced as the state of the art methods in PolSAR data classification. These two combinations include using GA algorithm as the feature selection method and the SVM as a classifier proposed by (Salehi et al., 2014) and using the GA algorithm with MLP classifier proposed by (Haddadi et al., 2011). In this paper, the same neural network configuration (number of layers and number of neurons) was used for MLP. Moreover, Wishart classifier (middle part of Figure 1) was also employed as an independent method to assess the efficiency of the proposed procedure. This classifier has been widely used as a baseline for performance evaluation of the developed algorithms for PolSAR data classification in many researches (Haddadi et al., 2011; Maghsoudi et al., 2012; Salehi et al., 2014; Yang, Gao, Xu, & Yang, 2015).

The left part of Figure 1 demonstrates the proposed method. In this method, two innovative aspects in PolSAR data classification have been proposed. These aspects include class-based features selection and a strategy to combine class-based classification maps.

The main idea of this approach is to find \( K \) set of features \( (\text{CBFS}_i, i = 1, 2, ..., K \) where \( K \) is the number of classes) that each set emphasizes on the separability of one class from others. Subsequently, each set of features are used to produce a classification map for all classes. Obviously, by using \( i^{th} \) set of features, it is expected that the \( i^{th} \) class is detected more precisely than others. So, the second step of our approach is dedicated to combine all of the classification results. In combination procedure, having removed all labels except the \( i^{th} \) label for \( i^{th} \) classification map, the binary classification maps are obtained. Then, a pixel is labelled as a specified class, if and only if it appears in just one of these binary classification maps, otherwise, that pixel is labelled as unclassified. Finally, for all of the unclassified pixels, the most-frequent label between the classification maps will decide the final label of that pixel. A complete explanation of the proposed approach is presented as follows.
PolSAR feature generation

Complex scattering matrix $[S]$ with quad-polarizations is measured by fully polarimetric radar sensors.

$$[S] = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{HV} & S_{VV} \end{bmatrix}$$

(1)

Where, $S_{HH}$, $S_{HV} = S_{HV}$ and $S_{VV}$ are complex back-scattered values for horizontal $(H)$ and vertical $(V)$ transmitting and receiving signals.

The second-order descriptors of the $3 \times 3$ average polarimetric covariance $[C]$ and coherency $[T]$ matrices can be derived from $[S]$. In this regards, two target vectors $h_i$ and $k_i$ for a sample $i$ can be formed using the Lexicographic and Pauli bases (Lee & Pottier, 2009) respectively.

$$h_i = \begin{bmatrix} S_{HH} \sqrt{2}S_{HV} & S_{VV} \end{bmatrix}^T$$

(2)

$$k_i = \frac{1}{\sqrt{2}} [S_{HH} + S_{VV} S_{HH} - S_{VV} 2S_{HV}]^T$$

(3)

The multi-look polarimetric covariance and coherency matrices could be calculated using the following formulas.

$$\langle [C] \rangle = \frac{1}{p} \sum_{i=1}^{p} h_i h_i^T$$

(4)

$$\langle [T] \rangle = \frac{1}{p} \sum_{i=1}^{p} k_i k_i^T$$

(5)

The superscripts "*" and "T" denote the complex conjugate and the matrix transpose, respectively, and $p$ denotes the number of samples.

Three types of polarimetric features could be generated and employed for classification including (i) the features obtained directly from original data, (ii) the features derived using decomposition algorithms and (iii) the SAR discriminators (Maghsoudi et al., 2012). All decomposition and SAR discriminator features are directly or indirectly estimated from these three matrices of $[S]$, $[C]$ and $[T]$. The objective of decomposition methods is to somehow classify pixels into different physical backscattering mechanisms of single bounce, double bounce, volume and helix scattering. The SAR discriminators are also estimated and used as indicators to separate surface types. Considering these three types of features, in this paper, a total number of 138 features were generated using POLSARPro Software (version 5) developed by ESA (Pottier & Ferro-Famil, 2012). Table 1 shows all of the generated features used in this study.

Class-based features (CBFs) selection

Since we encounter a wide variety of features, the optimum feature selection algorithm is necessary. The proposed algorithm to select optimum features includes an innovative class-based feature selection.
Table 1. Extracted features from PolSAR images used in this study.

| Feature Description | Symbol | Number | Feature Description | Symbol | Number |
|---------------------|--------|--------|---------------------|--------|--------|
| **Original Features** | | | | | |
| Scattering matrix | [S] | 3 | Covariance matrix | [C] | 6 |
| Coherency matrix | [T] | 8 |
| **Decomposition Features** | | | | | |
| Krogager | [Krog] | 3 | Yamaguchi | [Yama] | 4 |
| (Krogager, Dall, & Madsen, 1995) | | | (Yamaguchi, Moriyama, Ishido, & Yamada, 2005) | | |
| Huynen (Huynen, 1970) | [T_Huyn] | 9 | Touzi (Touzi, 2007) | [Touzi] | 4 |
| Barnes2 | [T_barn2] | 9 | Neumann | [Neumann] | 4 |
| (Barnes, 1988) | | | (Neumann, 2009) | | |
| Cloude (Cloude, 1985) | [T_Cloud] | 9 | An Yang | [An_Yang] | 4 |
| Holm2 | [T_Holm2] | 9 | Arii3_NNED (Arii3_NNED) | | 3 |
| (Holm & Barnes, 1988) | | | (J. V. Zyl, Ari, & Kim, 2011) | | |
| Vanzyli3 | [Vanzyl3] | 3 | Zhang | [Zhang] | 5 |
| (Zyl, 1989) | | | (Singh, Yamaguchi, & Park, 2013) | | |
| Freeman Durden (Freeman & Durden, 1998) | [Free] | 3 | Freeman_Durden (Freeman & Durden, 1998) | | |
| Anisotropy_Complex_kozlov | [A_kozlov] | 21 | HAAlpha_Diversity | | 8 |
| | | | Index | | |
| Entropy Freeman | [H_Free] | | | | |
| Entropy Vanzyli | [H_Vanzyl] | | | | |
| Cloude-Pottier | | | | | |
| H, A, Alpha, beta, A12, gamma, lambda, A, luen, (1-H)(1-A), HA, (1-H)(A, RVI, H_Shannon, H_Shannon, H_ShannonP, Pedestal, DERD, SERD, Asym | | | | |
| (Cloude & Pottier, 1997) | | | | |
| **SAR Discriminators** | | | | | |
| SPAN | [SPAN] | | | | |
| Polarization Fraction | [polFrac] | | | | |
| Depolarization index | [Depol_ind] | | | | |
| Conformity Coefficient | [Conformity] | | | | |
| Scattering Predominance | [Scatt_Predominance] | | | | |

Plotting features’ mean values of training samples for each class revealed that some features explicitly have high potential to discriminate special classes. This fact led us to the proposed class-based feature selection described in the following:

Considering \( \mathbf{[\mu]}_{1 \times n} = [\mu_1, \mu_2, \ldots, \mu_n] \) and \( \mathbf{[s]}_{1 \times n} = [s_1, s_2, \ldots, s_n] \), \( j = 1, 2, \ldots, K \) as respectively the mean and standard deviation vectors of training samples for class \( j \), here \( n = 138 \) is the number of the generated features and \( K \) is the number of classes, the fitness of \( f^j_i \) \( (K-1) \times 1 \) is defined as follow:

\[
\bar{f}^j_i = \left[ \frac{d^j_i}{\text{std mean}} \right] \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, K \tag{6}
\]

Where \( d^j_i \) \( (K-1) \times 1 \) and \( \text{std mean}^j \) \( (K-1) \times 1 \) are dimension vectors (Equations (7) and (8))

\[
\bar{d}^j_i (m) = |\mu^j_i - \mu^m| \quad m = 1, 2, \ldots, K \quad j \neq m \tag{7}
\]

\[
\text{std mean}^j (m) = \left( \bar{s}_i + \bar{s}_m \right)/2 \quad m = 1, 2, \ldots, K \quad j \neq m \tag{8}
\]

The minimum value of \( f^j_i \) will represent the fitness of \( i^{th} \) feature to separate \( j^{th} \) class (Equation (9)).

\[
\text{Fitness}_i^j = \arg\min \left( \bar{f}^j_i \right) \tag{9}
\]

To select optimum features for \( j^{th} \) class, the following procedure is employed: In the first step, the feature with the maximum value of \( \text{Fitness}_i^j \) is considered as the best optimum feature \( C^0_1 \) to separate class \( j \) from others (Equation (10)).

\[
C^0_1 = \arg\max_i \left( \text{Fitness}_i^j \right) \tag{10}
\]

Another criterion must be considered to select next features to avoid selecting redundant features. The redundant features do not provide any mutual separabilities in feature space for classification process. In this algorithm, an Euclidean distance measure is used to add features to optimum feature set that could maximize mutual separabilities at the same time. Equation (11) shows the relation used to select the second and the next features considering this criterion.

\[
C^j_i = \arg\max_i \left( \text{Fitness}_i^j \times \sqrt{\prod_{m=1}^{j-1} D^m_{ij}} \right) \tag{11}
\]

Here, \( C^j_i \) is the \( i^{th} \) selected feature for \( j^{th} \) class and \( D^j_{ij} \) is the Euclidean distance between \( i^{th} \) feature and \( s^{th} \) previously selected feature (Equation (12)).
This process is repeated for each class until all of the desired features are found. The number of features to be selected and used in classification process was also another challenge of this study. To overcome this problem, we used the overall accuracy of classification as a criterion for the number of optimum features for each class. In the CBFs Selection algorithm, feature selection for each class is repeated until no further overall accuracy improvement is observed by increasing the number of features.

Combining the classification maps

$K$ set of selected optimum features are used to classify PolSAR data and $K$ different classified maps are generated. So, considering the availability of $K$ different classified maps, a proper procedure must be used to combine these classified maps. It is expected that each classification map has potentially detected one class more precisely. This potential should be considered in producing final classification map. So, in this approach, best class priority of each classification map has the major role in generation of final output.

Considering $Map_j$ as a classification map obtained from $C$ features, the proposed combination method used to produce final classified map as the following pseudo code. It should be noted that the values registered in each $Map_j$ ranges from 1 to $K$ corresponding to a class ID.

Final classification map resulted from the above mentioned pseudo code is quantified to $(K + 1)$ values. Pixels with zero value correspond to unclassified regions and pixels with numeric values of 1 to $K$ show specific semantic classes.

Datasets and pre-processing

In this study, two full polarimetric dataset from the same area were used for performance evaluation of the proposed method. The first one is a C-band (5.5 cm of wavelength) Radarsat-2 image of San Francisca acquired on 9 April 2008. The used subset in this study includes 1270 by 850 pixels and covers an area of about 50.63 km$^2$. The incidence angle ranges from 28.54° to 29.08° and ground range and azimuth pixel spacings are around 9.8 m and 4.8 m respectively. The study area contains four main classes of vegetation, ocean, urban and road. High-resolution optical satellite images were used to select ground truth data. Figure 2 shows the Pauli representation of the image with selected ground truth data. Pauli representation assigns [HH-VV], [HV] and [HH+VV] to the Red, Green and Blue channels. The ground truth data information is also provided in Table 2.

The second PolSAR data is the NASA/Jet Propulsion Laboratory Airborne SAR (AIRSAR) L-band data (24 cm of wavelength) of San Francisco. This fully Polarimetric SAR dataset has been acquired in August 1989, with average incidence angle of about 20 degrees. The used subset in this study includes 570 by 800 pixels and the ground range and azimuth pixel spacings are around 10 m. Although the spatial coverage of this dataset is almost as same as the previous dataset, having different incidence angles, different imaging band and different spatial resolution, makes it another convenient dataset to evaluate the efficiency of the proposed method. Figure 3 shows the Pauli representation of this image with selected ground truth data. The ground truth data information is also provided in Table 3.

The first important step in pre-processing is dealing with inherent speckle noise of SAR images. So far, different speckle filtering algorithms have been developed to decrease its side effects keeping the original information of PolSAR images. Speckle reduction filters must reach the best possible compromise between noise reduction within homogeneous extended targets and spatial details preservation (Foucher & Lopez-Martinez, 2009). In this paper, to extract incoherent decomposition parameters, first, the 5 × 5 Refined Lee speckle filter (Lee, Grunes, & Grandi, 1999) for Radarsat-2 dataset and the 3 × 3 Refined Lee speckle filter for AirSAR dataset was applied on coherency matrix and then correspondent features (Table 1) were calculated. Subsequently, to preserve the phase information, the coherent decomposition parameters were generated from original coherency matrix and then filtered to reduce speckle. Finally, in order to scale data into the same dynamic range, all of the generated features were normalized to $[-1 \sim 1]$ and used in the feature selection and classification steps.
Results and discussions

This section is dedicated to present obtained results and their performance evaluation. The proposed procedure was evaluated using different methods. Since common procedures for PolSAR data classification have been developed by making a variation in feature selection and classifier schemes, different feature selection algorithms and different classifiers were used to evaluate the proposed method. In other words, in this paper, the performance of the proposed method is evaluated in two stages of feature selection and classifier efficiency. To assess the feature selection performance, two well-known feature selection methods (Genetic Algorithm (GA) Features Selection and Prototype Space (PS) Feature Selection) were used to find the optimum features.

In the GA feature selection method, first, a binary encoding chromosome consisting $n$-genes (where $n = 138$ is the number of features) is used as an individual solution. Each gene represents one feature and could take a binary value. Genes with true value will be selected and the others will be omitted. Simultaneously reaching to the highest overall accuracy of the classification results as well as the minimum number of used features are considered as the fitness criteria of each chromosome (Salehi et al., 2014). The GA algorithm is run as an optimization procedure that could finally find the best solution considering the performance of chromosomes.

On the other hand, the PS is a $K$-dimensional Cartesian space which $K$ is the number of classes. Each feature becomes a point in this space. The coordinate of each feature is defined using the mean value of training samples in all classes. So, the features that fall close to the diagonal axis of prototype space show the same response for all classes and could not be used to discriminate classes and must be omitted from the feature set. Moreover, the features that fall near each other show the same separability behaviours and one of them is sufficient to keep. (Mojaradi et al., 2009).

In PS feature selection method, the selection was continued until no overall accuracy improvement was observed and for the GA, all of the 138 features were used as input and the final optimum set of features

| Class   | # Training | # Test |
|---------|------------|--------|
| Vegetation | 3142       | 2707   |
| Ocean   | 2205       | 2332   |
| Urban   | 2621       | 4396   |
| Road    | 283        | 401    |
| Total   | 8251       | 9836   |

Figure 2. Pauli representation of Radarsat-2 image of San Francisco with its ground truth data.

Table 2. Number of training and test pixels of Radarsat-2 image for each class.

Figure 3. Pauli representation of AirSAR image of San Francisco with its ground truth data.
with simultaneously the highest overall accuracy and minimum number of features was found.

To assess the classifier performance of the proposed approach, four types of classifiers i.e. Fuzzy, SVM, MLP and ML were used.

The Wishart classifier is also another classifier that has been widely used as a baseline to classify PolSAR images. For homogeneous regions, it has been proven that the covariance matrix has a complex multivariate Wishart distribution. So, applying maximum likelihood classifier with Wishart distribution (instead of Gaussian distribution) on the nine elements of covariance matrix will result in a classified map. More detailed information on Wishart classifier could be found in (Lee & Pottier, 2009). Wishart classifier was also selected as another evaluator of the proposed method (Figure 1).

Results of the Radarsat-2 dataset processing

Table 4 shows the classification results (overall accuracy (OA) and producer accuracies (PA) of each class) for test data of Radarsat-2 image based on different previously developed feature selection and classification methods.

In Table 4, two combinations of using GA-SVM and GA-MLP for feature selection and classification have been proposed in (Salehi et al., 2014) and (Haddadi et al., 2011) respectively.

As could be seen from Table 4, the overall accuracies are fairly high but the producer accuracies for some classes (especially for road class) are low. This means that although, these methods could result in high overall accuracies, some classes are not well extracted from the image.

Almost all of the implemented methods presented in Table 4 resulted in poor producer accuracies for the road class. Considering the spatial resolution of the dataset (9.8 m) and also regarding the height of adjacent buildings of roads, physical radar backscattering mechanism is very complicated for the roads surrounded by buildings. On the other hand, because of the proximity of urban and road pixels, finding pure pixels of road have higher limitations compared to other classes. So, higher uncertainty in the road training pixels and higher diversity of training samples of this class could be considered as the main reasons of low classification accuracy of the road class.

Another main reason of poor producer accuracies could be the use of the same features to simultaneously identify all classes. So, it is reasonable to conclude that using different set of features to extract each class will result in high accuracies of each class as well as high overall accuracies. Therefore, the proposed class-based features selection was developed.

By applying the proposed class-based features selection method, the optimum features to extract each class were found (CBFs, $i = 1, 2, \ldots, K$). As mentioned before, the CBFs selection process was repeated until no further overall accuracy improvement is observed. So, the number of optimum features for each class will be different. Identified optimum features for each class is summarized in

Table 3. Number of training and test pixels of AirSAR image for each class.

| Class   | # Training | # Test   |
|---------|------------|----------|
| Vegetation | 712        | 1862     |
| Ocean   | 1850       | 2053     |
| Urban   | 464        | 733      |
| Road    | 199        | 253      |
| Total   | 3225       | 4901     |

Table 4. Classification accuracies for test data of Radarsat-2 image based on different previously developed feature selection and classification methods.

| FS method | Classifier     | Number of Features | OA   | PA (Vegetation) | PA (Ocean) | PA (Urban) | PA (Road) |
|-----------|----------------|--------------------|------|-----------------|------------|------------|-----------|
| P         | Fuzzy          | 9                  | 94.58| 90.19           | 99.06      | 97.50      | 75.70     |
|           | SVM            | 8                  | 93.56| 87.82           | 99.42      | 98.12      | 58.26     |
|           | MLP            | 9                  | 92.81| 85.13           | 98.12      | 98.09      | 79.11     |
|           | ML             | 4                  | 92.92| 87.55           | 96.99      | 97.88      | 77.57     |
| GA        | Fuzzy          | 12                 | 94.10| 93.21           | 89.19      | 97.69      | 69.20     |
|           | SVM            | 19                 | 94.27| 92.19           | 98.82      | 91.26      | 82.75     |
|           | MLP            | 18                 | 91.57| 90.90           | 93.51      | 90.55      | 69.91     |
|           | ML             | 6                  | 91.31| 90.47           | 91.02      | 88.32      | 81.20     |
|           | Wishart        | 9                  | 90.06| 80.59           | 99.19      | 94.63      | 85.05     |

Table 5. Optimum selected features for each class using proposed CBFS algorithm for Radarsat-2 image.

| Class            | Features                                                                 | Number of Features |
|------------------|--------------------------------------------------------------------------|--------------------|
| Vegetation       | H(1-A), CCC, Entropy_P, Praks_Colin, A, (1-H)A, T22_mod_Barn2, DERO, Entropy_Shanon_P | 8                  |
| Ocean            | Alpha, Vanzyl3_Vol, Asymmetry, S_hv | 4                  |
| Urban            | Entropy_Shanon, Krogager_Kd, Entropy_Shannon_I, T11_mod, Vanzyl3_Vol, Entropy_Shannon_P | 12                 |
| Road             | T22_mod_Barn2, T2_mod_Barn2, S_hv, T22_mod_Holm2, Vanzyl3_Vol, T33_mod_Barn2, T22_mod | 8                  |
| Total number of Unique features: 25
Table 5. The mean and standard deviation values for optimum features of each class are also displayed in Figure 4 as well.

Figure 4 shows that the proposed feature selection algorithm has succeeded to select features that have high separability for each class from others with trying to make the standard deviations as low as possible. Different patterns of separability between selected features have also been guaranteed.

After finding optimum features, using each set of them, the classification process by different classifiers was performed to obtain four classified maps and then, the confusion matrices for test data were generated. Table 6 summarizes these results.

As can be seen from Table 6, classification results using each set of optimum features have the highest producer accuracy for the corresponding class. But in comparison with Table 4, this procedure does not necessarily guarantee to reach higher overall accuracies. For example, results of classification based on optimum features of ocean show that overall accuracies are much lower than the results reported in Table 4. On the other hand, producer accuracy of vegetation for optimum feature set of vegetation (e.g. 95.47% for Fuzzy classifier) has the highest value between all optimum feature sets (80.07, 91.52 and 92.04%). For ocean, we also have 99.24 per cent compared to 94.38, 94.92 and 97.46 per cent for fuzzy classifiers. Urban areas classification accuracy was acceptable and Road pixels have also been extracted more precisely for the optimum feature set of road areas.

Figure 5 separately shows the extracted classes using different classifiers based on the proposed class-based feature selection algorithm.

So far, for each classifier, we have four classified maps and we need to combine them to build the final classified map and calculate the classification accuracies. Using the proposed algorithm, classified maps were combined and the final classified map for each classifier was generated. The producer and overall accuracies of the proposed approach are summarized in Table 7.

The obtained results of Table 7 could be compared with the results provided in Table 4. However, the high overall accuracy of 98.09% (for fuzzy classifier) as well as high producer accuracies of above 93% for the test selected samples show the effectiveness of the proposed method. These high values of accuracies are the results of class-based feature selection,

### Table 5

| Class  | Features of Class | Classifier | OA (Vegetation) | PA (Ocean) | PA (Urban) | PA (Road) |
|--------|-------------------|------------|-----------------|------------|------------|-----------|
| Vegetation | Radarsat-2 image | Fuzzy      | 92.91           | 95.47      | 94.38      | 88.71     | 92.21     |
|         |                   | SVM        | 91.66           | 92.15      | 96.85      | 99.42     | 46.11     |
|         |                   | MLP        | 90.43           | 93.11      | 94.19      | 94.87     | 63.32     |
|         |                   | ML         | 92.87           | 92.68      | 93.61      | 97.44     | 50.78     |
| Ocean   | Radarsat-2 image | Fuzzy      | 75.92           | 80.07      | 99.24      | 49.89     | 91.11     |
|         |                   | SVM        | 61.82           | 15.13      | 98.38      | 92.11     | 29.28     |
|         |                   | MLP        | 81.15           | 73.13      | 95.76      | 79.05     | 49.52     |
|         |                   | ML         | 60.47           | 14.79      | 94.38      | 85.50     | 80.69     |
| Urban   | Radarsat-2 image | Fuzzy      | 93.40           | 91.52      | 94.92      | 98.21     | 91.90     |
|         |                   | SVM        | 93.19           | 88.30      | 98.65      | 96.89     | 54.21     |
|         |                   | MLP        | 91.54           | 87.65      | 95.09      | 97.18     | 69.03     |
|         |                   | ML         | 86.30           | 71.81      | 95.14      | 92.94     | 74.45     |
| Road    | Radarsat-2 image | Fuzzy      | 93.57           | 92.04      | 97.46      | 92.63     | 93.77     |
|         |                   | SVM        | 93.90           | 87.76      | 97.91      | 94.74     | 79.42     |
|         |                   | MLP        | 94.16           | 88.67      | 96.99      | 93.42     | 92.61     |
|         |                   | ML         | 90.53           | 79.86      | 97.53      | 93.66     | 88.64     |

Figure 4. mean and standard deviation of training classes of Radarsat-2 image for selected optimum features of vegetation (a), Ocean (b), Urban (c) and Road (d).
Nevertheless, this high value of overall accuracy could also be related to the selected test areas, and the overall accuracy for other test samples might be different.

Almost all of the classifiers have succeeded in discriminating desired classes. However, some differences could be seen in the results. Referring to Table 7, fuzzy classifier, because of its inherent characteristics to handle uncertainties, has reached higher accuracies. The uncertainties in PolSAR images might be originated from different sources. Speckle noise, target geometry, dielectric constant and surface roughness cause different backscattering patterns for the same class in

| Classifier | OA (Vegetation) | PA (Ocean) | PA (Urban) | PA (Road) |
|------------|-----------------|------------|------------|-----------|
| Fuzzy      | 98.09           | 97.90      | 99.50      | 97.44     | 93.64     |
| SVM        | 97.26           | 93.42      | 98.33      | 96.82     | 80.39     |
| MLP        | 94.65           | 93.85      | 95.41      | 95.55     | 91.07     |
| ML         | 93.78           | 92.92      | 94.05      | 93.81     | 89.24     |

Figure 5. Identified areas as special classes by their CBFs for Radarsat-2 image (from left: Vegetation, Ocean, Urban and Road).
a PolSAR image. Moreover, since the same objects are located in different parts of the image, they experience different incidence angles and most probably different orientation with respect to radar viewing direction which are other sources of different backscattering patterns for the same class. So, it seems that the fuzzy classifiers with capabilities to handle uncertainties related to different parameters could be a convenient candidate in PolSAR imagery classification. Based on these results, the fuzzy classifier has been nominated as the best classifier in this paper.

In the Fuzzy classifier of this work, the initial membership functions for each feature were built using statistical parameters (mean and standard deviation) of training areas for each class. So, there will be K membership functions for each feature. All training samples were used to produce all possible rules and redundant rules were eliminated. Subsequently, this Fuzzy Inference System (FIS) was used as initial FIS in Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang, 1993). The resulted optimized FIS from ANFIS was used as final classifier. Final classified map generated from the proposed approach for the best classifier (Fuzzy Classifier) is provided in Figure 6.

**Results of AirSAR dataset processing**

All of the processing applied on Radarsat-2 image was repeated for the AirSAR dataset to verify the performance of the proposed method. Table 8 shows the classification accuracies in each class for test data of AirSAR image based on different previously developed feature selection and classification methods.

Comparing Table 8 with Table 4 shows that the overall accuracy is again acceptable but these values are about seven percent lower than those of Radarsat-2 image due to different imaging wavelength and spatial resolution. Again, the producer accuracies for the road class are low and are expected to be improved by using the proposed class-based features selection and classification scheme.

By applying proposed CBFS selection method, optimum features for each class were found for AirSAR image of San Fransisco and are summarized in Table 9. The mean and standard deviation values for optimum features of each class are also displayed in Figure 7 as well.

Comparing Tables 9 and 5 shows that some features to extract a special class are identical for the both of Radarsat-2 and AirSAR images. However, this does not mean that always same features could be identified as

![Figure 6. Final classified map for Radarsat-2 image generated from the proposed approach based on Fuzzy classifier.](image)

| FS method | Classifier | Number of Features | OA | PA (Vegetation) | PA (Ocean) | PA (Urban) | PA (Road) |
|-----------|------------|--------------------|----|----------------|------------|------------|-----------|
| PSFS      | Fuzzy      | 8                  | 86.40 | 98.74 | 97.73 | 81.68 | 53.27 |
| SVM       | 16         | 84.26              | 85.25 | 95.50 | 82.97 | 38.65 |
| MLP       | 12         | 82.78              | 83.46 | 97.23 | 75.62 | 40.72 |
| ML        | 7          | 82.62              | 89.19 | 96.47 | 70.69 | 59.30 |
| GA        | Fuzzy      | 12                 | 88.09 | 92.70 | 98.09 | 87.72 | 51.75 |
| SVM       | 11         | 85.44              | 98.74 | 97.24 | 79.53 | 36.13 |
| MLP       | 9          | 84.32              | 89.25 | 97.50 | 80.37 | 45.12 |
| ML        | 6          | 82.99              | 79.78 | 99.50 | 78.02 | 50.95 |
| –         | Wishart    | 9                  | 82.15 | 75.70 | 98.50 | 76.51 | 70.87 |

| Class      | Features                                      | Number of Features |
|------------|----------------------------------------------|--------------------|
| Vegetation | CCC, Polarisation Fraction, Combination_HA, Combination_1mHA, VanZyl3_Vol, RO12 | 6                  |
| Ocean      | VanZyl3_Vol, Serd, Yamaguchi4_Vol, SVM_alpha, Svv, Svv | 5                  |
| Urban      | Krogager_ks, C13_mod, VanZyl3_Vol, Svv, T12_mod_Barn2, T22_mod | 10                 |
| Road       | Entropy_Shanon, Svv, Svv, T12_mod_Holm2 | 7                  |

Total number of Unique features: 24
optimum features for each class. Therefore, this optimum feature selection step is always necessary but by finding the same features, the execution time to find the optimum features could be decreased. More statistical analyses to prove this claim should be performed in the next researches.

The classification process was followed by classifying the AirSAR image using CBFs to obtain four classified maps. The results are provided in Table 10.

| Using Optimum Features of Class | Classifier | OA (Vegetation) | PA (Ocean) | PA (Urban) | PA (Road) |
|---------------------------------|------------|-----------------|------------|------------|------------|
| Vegetation                      | Fuzzy      | 84.93           | 95.19      | 95.47      | 79.74      |
|                                 | SVM        | 83.69           | 93.15      | 97.73      | 80.39      |
|                                 | MLP        | 82.09           | 88.20      | 93.82      | 78.27      |
|                                 | ML         | 86.51           | 90.26      | 96.47      | 81.03      |
| Ocean                            | Fuzzy      | 72.40           | 61.38      | 98.24      | 78.45      |
|                                 | SVM        | 74.66           | 98.60      | 97.48      | 40.35      |
|                                 | MLP        | 73.89           | 81.27      | 96.98      | 75.12      |
|                                 | ML         | 71.56           | 49.72      | 96.32      | 77.80      |
| Urban                            | Fuzzy      | 78.50           | 68.54      | 96.47      | 87.09      |
|                                 | SVM        | 84.31           | 98.88      | 97.73      | 86.21      |
|                                 | MLP        | 80.72           | 90.46      | 95.12      | 82.25      |
|                                 | ML         | 85.10           | 80.76      | 97.73      | 86.21      |
| Road                             | Fuzzy      | 83.08           | 85.25      | 96.47      | 66.59      |
|                                 | SVM        | 82.64           | 97.33      | 96.22      | 73.28      |
|                                 | MLP        | 78.27           | 90.30      | 94.56      | 65.12      |
|                                 | ML         | 79.18           | 70.65      | 98.24      | 70.69      |

Table 10. Classification accuracies for test data of AirSAR image based on optimum selected features for each class using proposed CBFS algorithm.

As expected, higher producer accuracies have been obtained for each class using its optimum set of features and all of the classifiers have resulted in almost the same overall accuracies.

Finally, each class generated using its CBFs was extracted and then combined to generate the final classified map. Figure 8 separately shows the extracted classes using different classifiers based on the proposed class-based feature selection algorithm.

Given the proposed algorithm, classified maps were combined and the final classified map was generated. The producer and overall accuracies of the proposed approach are summarized in Table 11.

As can be seen in Table 11, the Fuzzy classifier again has reached to the highest accuracy and could be considered as the optimum classifier. Final classified map generated from the proposed approach for the best classifier (Fuzzy Classifier) is also provided in Figure 9.

One important issue that should be considered in the evaluation procedures is the computation costs and the time of execution. The well-known Wishart classifier is based on just covariance matrix parameters and as soon as these parameters are calculated, the final classified map could be generated in a few seconds. However, the proposed method includes two main parts of feature selection and classification which most of the time of executing the algorithm is consumed in feature selection. In general, optimum feature selection methods are divided into two types of search-based and deterministic ones. The search algorithms like Genetic Algorithm require large number of repetitions and accordingly, they need higher calculation time. The proposed class-based feature selection method is more deterministic in comparison with the GA search algorithms. So, it requires less time than search-based methods. However, since the proposed method selects different sets of optimum features for each class, the number of required calculations is higher than the deterministic methods.
of algorithm repetitions is based on the number of classes. Thus, by increasing the number of classes, the calculation time of algorithm will be linearly increased.

All of the methods discussed in this paper were implemented through MATLAB 2009a installed on a computer with Intel Core i5-2.4GHz CPU and 4GB of RAM. Table 12 indicates the execution time for all of the implemented methods. It should be noted that, the time values presented in this table do not include the time required to generate features (Table 1). Depending on the number of required features, feature generation itself may take a few minutes to be performed. While the Wishart classifier with only nine required features has the minimum feature generation time, other implemented methods need

| Classifier | OA   | PA (Vegetation) | PA (Ocean) | PA (Urban) | PA (Road) |
|------------|------|-----------------|------------|------------|-----------|
| Fuzzy      | 92.53| 95.89           | 98.70      | 87.59      | 88.15     |
| SVM        | 90.57| 93.12           | 97.82      | 86.28      | 70.73     |
| MLP        | 89.82| 89.47           | 96.67      | 83.15      | 79.36     |
| ML         | 90.26| 91.66           | 96.35      | 87.46      | 86.05     |

Figure 8. identified areas as special classes by their CBFs for AirSAR image (from left: Vegetation, Ocean, Urban and Road).

Figure 9. Final classified map for AirSAR image generated from the proposed approach based on Fuzzy classifier.
higher number of features and consequently higher identical feature generation time.

In Table 12, CBFS denotes the class-based features Selection and PSFS denotes the Prototype Space Feature Selection method. As mentioned and expected before, the GA-based methods had high execution time and the Wishart classifier had the minimum. The proposed method also falls in between from the execution time point of view. It seems that using the parallel processing techniques could significantly decrease the processing time of the proposed algorithm.

**Conclusion and perspectives**

In this paper, initially, the PS feature selection and the GA feature selection algorithms were used to find the optimum features that could provide high discrimination for all classes and then the classification was performed. Results demonstrated that although high values of overall accuracies could be achieved, low producer accuracies for some classes may also occur. Therefore, a new class-based feature selection algorithm was proposed to achieve higher accuracies for all of the classes. The proposed algorithm improved the producer accuracies by around 5%. This proves that optimum features for each class could result in higher accuracy for that class.

Moreover, a method to combine results of classification for each set of optimum features was suggested. Using this algorithm, the final classified map was generated so that the high percent of overall and producer accuracies were achieved.

Furthermore, the Fuzzy classifier showed better results for classification. It seems that these achievements have been caused by the characteristics of designed Fuzzy inference system (FIS). Results showed that FIS as a flexible expert system could be adopted with challenges inherently existed in PolSAR imagery classification. Recently, another type of fuzzy systems named fuzzy type-2 has been used by scientific society for different applications due to their abilities to handle higher degrees of uncertainty. It is also worth to assess this type of FIS to investigate their abilities to improve classification accuracies in the future investigations.

Results of Radarsat-2 and AirSAR images processing show that the proposed method could be successfully used to obtain simultaneously higher overall and producer accuracies. However, each developed method has some limitations and is considered as a new tool along with the other developed techniques. Since the proposed algorithm is based on statistics, it is regarded as a statistical pattern recognition approach and practically, diverse features could be used as the input. So, the knowledge of the physical mechanisms of radar waves backscattering in the coverage of the image does not have any influence on the proposed algorithm and from this point of view, the proposed algorithm could be preferred to the classifiers based on physical backscattering mechanisms. Meanwhile, it is evident that since the backscattering mechanism is not known, it is necessary to inject diverse informative features to the proposed algorithm. Therefore, producing large number of features that most of them might not be selected as efficient features imposes a high volume of processing to the algorithm. By the way, all of the feature selection algorithms hold this limitation.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**References**

Aghababae, H., Amini, J., & Tzeng, Y.-C. (2013). Contextual PolSAR image classification using fractal dimension and support vector machines. *European Journal of Remote Sensing, 46*(1), 317–332. doi:10.5721/EuJRS20134618

An, W., Cui, Y., & Yang, J. (2010). Three-component model-based decomposition for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing, 48*(6), 2732–2739. doi:10.1109/TGRS.2010.2041242

An, W., Cui, Y., Yang, J., & Zhang, H. (2010). Fast alternatives to H/a for polarimetric SAR. *IEEE Geoscience and Remote Sensing Letters, 7*(2), 343–347. doi:10.1109/LGARS.2009.2035135

Bai, Y., Peng, D., Yang, X., Chen, L., & Yang, W. (2014). Supervised feature selection for polarimetric SAR...
classification. 12th International Conference on Signal Processing (ICSP).

Barnes, R.M. (1988). Roll invariant decompositions for the polarization covariance matrix. Lincoln Laboratory, MIT, Lexington, MA, USA, 02171-0073, Internal Report.

Chen, L., Yang, W., Liu, Y., & Sun, H. (2010). Feature evaluation and selection for polarimetric SAR image classification. IEEE 10th International Conference on Signal Processing (ICSP).

Claude, S., & Pottier, E. (1997). An entropy based classification scheme for land applications of polarimetric SAR. IEEE transactions on geoscience and remote sensing, 35, 68–78. doi:10.1109/83.551935

Claude, S.R. (1985). Radar target decomposition theorems. Inst. Elect. Eng. Electron, L**,ett., 21(1), 22–24.

Du, P., Samat, A., Waske, B., Lie, S., & Li, Z. (2015). Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. ISPRS Journal of Photogrammetry and Remote Sensing, 105, 38–53. doi:10.1016/j.isprsjprs.2015.03.002

Foucher, S., & Lopez-Martinez, C. (2009). An evaluation of PolSAR speckle filters. Proc. IEEE int. Geoscience and Remote Sensing Symp., IGARSS 2009, Cape Town, South Africa, 845–848.

Freeman, A., & Durden, S.L. (1998). A three-component scattering model for polarimetric SAR data. IEEE Transactions on Geoscience and Remote Sensing, 36(3), 963–973. doi:10.1109/69.673687

Haddadi, A., Sahebi, M.R., & Mansourian, A. (2011). Polarimetric SAR feature selection using a genetic algorithm. Canadian Journal of Remote Sensing, 37(1), 27–36. doi:10.5589/m11-013

Hellmann, M., & Jäger, G. (2002). Fuzzy rule based classification of polarimetric SAR data. Aerospace Science and Technology, 6, 217–232. doi:10.1016/S1270-9638(02)01152-5

Holm, W., & Barnes, R. (1988). On radar polarization mixed target state decomposition techniques. Proc. 1988 National Radar Conf, 248–254.

Huynen, J. (1970). Phenomenological theory of radar targets. Ph.D. dissertation, Delft University, Delft, The Netherlands.

Jang, J.-S.R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. IEEE Systems, Man, and Cybernetics Society, 23(3), 665–685. doi:10.1109/21.256541

Krogager, E., Dall, J., & Madsen, S. (1995). Properties of sphere, diplane and helix decomposition. Proc. 3rd Int. Workshop on Radar Polarimetry.

Lee, J.S., Grunes, M.R., & Grandi, G.D. (1999). Polarimetric SAR speckle filtering and its implication for classification. IEEE Transactions on Geoscience and remote sensing, 37(5), 2363–2373. doi:10.1109/36.789635

Lee, J.S., Grunes, M.R., & Kwok, R. (1994). Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution. International Journal of Remote Sensing, 15(11), 2299–2311. doi:10.1080/01431169408954244

Lee, J.-S., & Pottier, E. (2009). Polarimetric radar imaging from basics to applications. Boca Raton, FL: CRC Press.

Lee, J.-S., & Pottier, M.K.G.E. (2001). Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR. IEEE Geoscience and Remote Sensing Society, 3901–3905. doi:10.1109/36.960370

Mghouchi, Y., Collins, M., & Leckie, D.G. (2012). Polarimetric classification of Boreal forest using non-parametric feature selection and multiple classifiers. International Journal of Applied Earth Observation and Geoinformation, 19, 139–150. doi:10.1016/j.jag.2012.04.015

Mojarradi, B., Abrishami-Moghaddam, H., Zogei, M.J.V., & Duin, R.P.W. (2009). Dimensionality reduction of hyperspectral data via spectral feature extraction. IEEE Transactions on Geoscience and Remote Sensing, 47, 7.

Neumann, M. (2009). Remote sensing of vegetation using multi-baseline polarimetric SAR interferometry: Theoretical modeling and parameter retrieval. Ph.D. Thesis, University of Rennes-I, Rennes, France.

Pottier, E., & Ferro-Famil, L. (2012). PolSARPro V5.0: An ESA educational toolbox used for self-education in the field of POLSAR and POL-INSAR data analysis. Geoscience and Remote Sensing Symposium (IGARSS).

Praks, J., Koeniguer, E.C., & Hallikainen, M.T. (2009). Alternatives to target entropy and alpha angle in SAR polarimetry. IEEE Transactions on Geoscience and Remote Sensing, 47(7), 2262–2274. doi:10.1109/TGRS.2009.2013459

Qi, Z., Yeh, A.G.-O., Li, X., & Lin, Z. (2012). A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. Remote Sensing of Environment, 118, 21–39. doi:10.1016/j.rse.2011.11.001

Salehi, M., Sahebi, M.R., & Mghouchi, Y. (2014). Improving the accuracy of urban land cover classification using Radarsat-2 PolSAR data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 4.

Singh, G., Yamaguchi, Y., & Park, S.E. (2013). General four-component scattering power decomposition with unitary transformation of coherency matrix. IEEE Transactions on Geoscience and Remote Sensing, 51(5), 3014–3022. doi:10.1109/TGRS.2012.2212446

Theodoridis, S., & Koutroumbas, K. (2008). Pattern Recognition. USA: Academic Press.

Touzi, R. (2007). Target scattering decomposition in terms of roll-invariant target parameters. IEEE Transactions on Geoscience and Remote Sensing, 45(1), 73–84. doi:10.1109/TGRS.2006.886176

Yamaguchi, Y., Moriyama, T., Ishido, M., & Yamada, H. (2005). Four-component scattering model for polarimetric SAR image decomposition. IEEE Transactions on Geoscience and Remote Sensing, 43(8), 1699–1706. doi:10.1109/TGRS.2005.852084

Yang, F., Gao, W., Xu, B., & Yang, J. (2015). Multi-frequency polarimetric SAR classification based on Riemannian manifold and simultaneous sparse representation. Remote Sensing, 7, 8469–8488. doi:10.3390/rs70708469

Zhang, L., Zou, V., Cai, H., & Zhang, Y. (2008). Multiple-component scattering model for polarimetric SAR image decomposition. IEEE Geoscience and Remote Sensing Letters, 5(4), 2008. doi:10.1109/LGRS.2008.2000795

Zyl, J.J.C. (1989). Unsupervised classification of scattering behavior using radar polarimetry data. IEEE Transactions on Geoscience and Remote Sensing, 27(1), 36–45. doi:10.1109/36.20273

Zyl, J.J.V., Arii, M., & Kim, Y. (2011). Model-based decomposition of polarimetric SAR covariance matrices constrained for nonnegative eigenvalues. IEEE Transactions on Geoscience and Remote Sensing, 49(9), 3452–3459. doi:10.1109/TGRS.2011.2128325