Fully polarimetric synthetic aperture radar data classification using probabilistic and non-probabilistic kernel methods

Iman Khosravi*, Yury Razoumnyb, Javad Hatami Afkouieh** and Seyed Kazem Alavipanahc

*Department of Remote Sensing & GIS, Faculty of Geography, University of Tehran, Tehran, Iran; **Department of Mechanics and Mechatronics, Academy of Engineering, Peoples’ Friendship University of Russia (RUDN University), 117198 Moscow, Russia Federation

ABSTRACT
The data classification of fully polarimetric synthetic aperture radar (PolSAR) is one of the favourite topics in the remote sensing community. To date, a wide variety of algorithms have been utilized for PolSAR data classification, and among them kernel methods are the most attractive algorithms for this purpose. The most famous kernel method, i.e., the support vector machine (SVM) has been widely used for PolSAR data classification. However, until now, no studies to classify PolSAR data have been carried out using certain extended SVM versions, such as the least squares support vector machine (LSSVM), relevance vector machine (RVM) and import vector machine (IVM). Therefore, this work has employed and compared these four kernel methods for the classification of three PolSAR data sets. These methods were compared in two groups: the SVM and LSSVM as non-probabilistic kernel methods vs. the RVM and IVM as probabilistic kernel methods. In general, the results demonstrated that the SVM was marginally better, more accurate and more stable than the probabilistic kernels. Furthermore, the LSSVM performed much faster than the probabilistic kernel methods and its associated version, the SVM, with comparable accuracy.

Introduction
Fully polarimetric synthetic aperture radar (PolSAR) data are the characteristic earth observation in remote sensing applications, thanks to the weather – independence and day-night imaging potentiality. The PolSAR data classification is always one of the favourite topics for remote sensing researchers. In the literature of the last decade, a wide variety of algorithms with different architectures and procedures for PolSAR data classification exist, such as Gaussian maximum likelihood (Dargahi et al., 2013; Safaee & Sahebi, 2019; Shokrollahi & Ebadi, 2016), wishart maximum likelihood (Atwood et al., 2012; Dargahi et al., 2013; Jafari et al., 2015; Ma et al., 2014; Safaee & Sahebi, 2019; Salehi et al., 2014; Shokrollahi & Ebadi, 2016), decision tree (McNairn et al., 2014; Qi et al., 2015), random forest and rotation forest (Du et al., 2015; Khosravi et al., 2018, 2017; Uhlmann & Kiranyaz, 2014), neural network family such as multi-layer perceptron or deep learning (Alberga et al., 2008; Bruzzone et al., 2004; Safaee & Sahebi, 2019; Salehi et al., 2014; Zhang et al., 2019) and support vector machine (SVM) (Aghababaei et al., 2013; Ma et al., 2014; Niu & Ban, 2014; Safaee & Sahebi, 2019; Salehi et al., 2014; Uhlmann & Kiranyaz, 2014).

Among the algorithms mentioned above, there has been a strong inclination to use the SVM thanks to its high efficiency, generalizability and attractiveness in the classification process. However, this method has been characterized by certain drawbacks, such as a time-consuming training phase and non-probabilistic outputs, which have been addressed by researchers. This is perhaps due to the fact that a great deal of effort has been made in these last few decades to extend the SVM’s original formulation, in order to speed up its processing, produce probabilistic outputs or provide sparser models (Mounttrakis et al., 2011). For example, the least squares SVM (LSSVM) was recommended with the aim of expediting the training phase of the SVM (Wang & Hu, 2005). In another example, the relevance vector machine (RVM) was proposed as a Bayesian sparser kernel alternative to the SVM (Tipping, 2001). In addition to the RVM, the import vector machine (IVM) was another probabilistic extension of the SVM (Zhu & Hastie, 2005). Although, these three extended algorithms have been used widely for optical remote sensing data classification (Braun et al., 2011, 2012; Tipping, 2001; Zhu & Hastie, 2005), to date, there are no studies using these three methods for PolSAR data classification.
Hence, this work aims at using and comparing these four kernel methods for the classification of three PolSAR data sets, in terms of accuracy and speed. These methods have been divided into two groups: the SVM and LSSVM as non-probabilistic kernel methods vs. the RVM and IVM as probabilistic kernel methods. Section 2 presents the procedures of these algorithms for classification. The data sets are documented in Section 3 and the results of the implementation of the methods are explained in detail for the three data sets in Section 4. The conclusion of the paper is presented in Section 5.

Methodology

Suppose that, $X$ is the training data with $n$ samples $x_i$ and labels $y_i$, $i \in [1, 2, \ldots, n]$. The kernel methods aim to solve classification problems to obtain $y_i = w^T \phi(x_i) + b$, where $w$ is a weight vector, $\phi(x_i)$ is a non-linear mapping function and $b$ is a bias. A kernel function can be accordingly defined as $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. The most well-known kernel in remote sensing applications is the Gaussian radial basis function (RBF), defined by Pal et al. (2013):

$$K_{\text{gauss}}(x_i, x_j) = \exp \left( - \frac{||x_i - x_j||^2}{\sigma^2} \right)$$  \hspace{1cm} (1)

The Gaussian width parameter ($\sigma$) is the main parameter of the RBF kernel. The SVM, LSSVM, RVM and IVM are briefly described in the following section.

SVM

The SVM, as a binary supervised learning algorithm, is based on statistical learning theory (Hasani et al., 2017). This method maps data into a high dimensional space, in which the data have a simpler representation. This mapping is done implicitly using a kernel function. After mapping, the SVM estimates a separating hyperplane between two classes in the kernel space, ensuring it is at a maximum distance from the nearest samples of each class, i.e., support vectors (SVs). The optimization problem can be mathematically expressed as the following quadratic programming problem:

$$\min_{w, b, \xi} J(w, \xi) = \frac{||w||^2}{2} + C \sum_{i=1}^{n} \xi_i$$  \hspace{1cm} (2)

$$\text{subjectto} \quad y_i(w^T x_i + b) \geq 1 - \xi_i$$  \hspace{1cm} (3)

where, $J$ is a loss function, $C$ is a regularization parameter and $\xi_i \geq 0$ is a slack variable for falsely assigned training data in favour of generalization. The position of the samples with respect to this separating hyperplane is then used for their classification. For a multi-class task, the SVM uses either one-against-one or one-against-all strategies.

LSSVM

The LSSVM, with a main idea similar to the SVM, employs the least squares technique for solving the objective function problem (Wang & Hu, 2005). In mathematical terms, the loss function is changed to the quadratic loss function and the inequality constraints are replaced by equality constraints. Thus, the optimization problem is modified as follows (Suykens & Vandewalle, 1999):

$$\min_{w, b, \xi} J(w, \xi) = \frac{||w||^2}{2} + \frac{\gamma}{2} \sum_{i=1}^{n} \xi_i^2$$  \hspace{1cm} (4)

$$\text{subjectto} \quad y_i(w^T x_i + b) = 1 - \xi_i$$  \hspace{1cm} (5)

where, $\gamma$ is similar to $C$ of the SVM, which is used to control the loss function of the LSSVM.

RVM

The RVM is a Bayesian sparse kernel alternative to the SVM, which has many of the characteristics of the SVM, but without its principle limitations (Tipping, 2001). Assuming a Bernoulli distribution, a likelihood function is introduced, where a weight vector $\lambda$ is sought to maximize the conditional probability of the target $t$ depending on $x_i$ and $\lambda$:

$$P(t | \lambda) = \prod_{i=1}^{n} \sigma(y_i | x_i; \lambda)^{y_i} [1 - \sigma(y_i | x_i; \lambda)]^{1 - y_i}$$  \hspace{1cm} (6)

where, $\sigma$ is a logic function. The weights are dependent on two classes of hyperparameters $\alpha, \beta$ that are assumed to be Gaussian distributed.

These hyperparameters are iteratively optimized to maximize (6). Finally, the model only depends on the subset $RV \supseteq X$ of non-zero elements in $\lambda$, which are called relevance vectors. However, unlike the SVs, the RVs are not the points closest to the respective other class but are samples from the centre of the distribution, considered most typical for their own class (Braun et al., 2012). It is claimed that the RVM needs considerably fewer kernel basis functions, i.e., RVs (Bishop, 2006).

IVM

The IVM was proposed based on the idea of similarity in the curved shape of the loss term of the SVM, i.e., $1 - y_i f(x_i)$ and the negative logarithmic likelihood (NLL) of the binomial distribution, i.e., $\log_2(1 + \exp(y_i f(x_i)))$. By replacing the loss term in the loss-penalty form of the SVM with the NLL of the binomial distribution, a new classifier can be obtained which provides probability outputs directly.
\[
\min_{w, \xi} \frac{1}{n} \sum_{i=1}^{n} \left[ \ln(1 + e^{y_i f(x_i)}) \right] + \frac{\lambda}{2} \|w\|^2
\]  

(7)

As in the case of the SVM, only a subset \( IV \supseteq X \) called import vectors are needed to solve (7) and kernel functions are introduced to work in high-dimensional feature spaces. This algorithm needs considerably less kernel basis functions (Braun et al., 2011).

**Experiment data sets**

Three data sets, i.e., Flevoland, Foulum and Winnipeg, acquired by the L-band AIRSAR, L-band EMISAR and L-band UAVSAR sensors, respectively, were examined in this paper for comparative evaluation of the kernel methods (see Figure 1). These data can be freely accessed from the European Space Agency (ESA) website and NASA JPL website.

The first and the third data sets covered a large agricultural area, the latter in the Southeast of Winnipeg, Manitoba, Canada on 29 June 2012 and the former in Flevoland, Netherlands, in 1989. The second data set covered forest, agricultural and urban areas in a village in Denmark on 17 April 1989. The classes and the number of samples for these data sets are provided in (Tables 1, 2 and 3).

Three data sets were pre-processed using multi-looking and speckle filtering. A \( 2 \times 3 \) multi-looking mask was applied to the third data set and two other data sets were not multi-looked. In addition, a Boxcar filter, \( 5 \times 5 \) in size, was applied to the third data set, and a Lee filter, \( 7 \times 7 \) in size, was applied to the two other data sets to alleviate the speckle effect. The spatial resolution of the data sets after pre-processing was 10 m, 5 m and 12 m, respectively.

The polarimetric features, extracted for the classification in this paper, were similar to those outlined in

![Figure 1](image-url)
Table 1. Classes and the number of samples per class for Flevoland data set.

| # | Class     | Samples | # | Class     | Samples | # | Class     | Samples |
|---|-----------|---------|---|-----------|---------|---|-----------|---------|
| 1 | Bare Soil | 7948    | 5 | Lucerne   | 5219    | 9 | Stem beans| 4326    |
| 2 | Beet      | 3019    | 6 | Peas      | 5492    | 10| Water     | 2463    |
| 3 | Forest    | 8114    | 7 | Potatoes  | 5791    | 11| Wheat     | 14,704  |
| 4 | Grass     | 4942    | 8 | Rapeseed  | 11,387  |   |           |         |

Table 2. Classes and the number of samples per class for Foulum data set.

| # | Class    | Samples | # | Class    | Samples | # | Class | Samples |
|---|----------|---------|---|----------|---------|---|-------|---------|
| 1 | Forest   | 63,529  | 3 | Rye      | 33,678  | 5 | Wheat | 15,847  |
| 2 | Oats     | 35,034  | 4 | Urban    | 20,520  |   |       |         |

Table 3. Classes and the number of samples per class for Winnipeg data set.

| # | Class | Samples | # | Class | Samples | # | Class | Samples |
|---|-------|---------|---|-------|---------|---|-------|---------|
| 1 | Broadleaf | 3655    | 4 | Forage | 2047    | 7 | Wheat | 30,192  |
| 2 | Canola | 30,813  | 5 | Oats   | 15,091  |   |       |         |
| 3 | Corn  | 15,425  | 6 | Soybeans | 26,116  |   |       |         |

Table 4. Polarimetric features extracted from each PolSAR data set.

| Polarimetric features | Backscatter intensities | Backscatter ratios | Ratio values | Correlation coefficients | Cloud & Pottier parameters | Pauli & Krogager parameters | Freeman-Durden and Yamaguchi parameters |
|-----------------------|-------------------------|-------------------|--------------|--------------------------|-----------------------------|-----------------------------|---------------------------------------|
|                       | σh,σv,σhh,σvv           | σh/σv,σh/σhh,σv/σvv | σh,σv,σh/σv,σh/σh | ρh/v,ρh/h,ρh/v,ρh/h           | λ1,λ2,λ3,λ4,λ5,λ6,λ7,λ8 | [ρh]2, [γ]2, | p1, p2, p3, p4, p5, p6 |

a paper by Khosravi et al. (2017). These features and their symbols were highlighted in (Table 4).

Generally, each of above features may be useful in distinguishing between certain land-cover classes and may be less useful in distinguishing between other classes. Thus, all these features were stacked together in this work to obtain high efficiency and accuracy in distinguishing between all land covers. A total of 48 polarimetric features were extracted from each data set.

Results and discussion

Assessment methods

For accuracy assessment, the kappa coefficient and overall accuracy (OA) were applied. They can be computed as follows (Congalton & Green, 2019):

\[
OA = \frac{\sum_{i=1}^{k} n_{ii}}{N},
\]

\[
kappa = \frac{\sum_{i=1}^{k} n_{ii} - \frac{\sum_{i=1}^{k} n_{ij} \cdot n_{ji}}{n}}{\frac{\sum_{i=1}^{k} n_{ij} + \sum_{i=1}^{k} n_{ji}}{n} - \frac{\sum_{i=1}^{k} n_{ij} \cdot n_{ji}}{n} - (k - 1)},
\]

where, \( n_{ij} \) is the diagonal element of the confusion matrix, related to \( i \)th class, \( i \in [1, 2, 3, \ldots, k] \), \( k \) is the number of classes, \( n_{ii} \) is the number of samples correctly classified as class \( i \) and \( n_{ij} \) is the number of samples of class \( j \) in existing reference data.

Kappa is computed to determine whether the values in a confusion matrix are significantly better than the values in a random assignment. OA is another general metric for accuracy assessment. However, the kappa coefficient is preferred in relation to the OA metric (Negri et al., 2016). To evaluate the significance of a difference between the method results, a bilateral test, i.e., the McNemar’s test was employed. This test was calculated as follows:

\[
z = \frac{c_{12} - c_{21}}{\sqrt{c_{12} + c_{21}}}
\]

where, \( c_{12} \) were the pixels correctly classified in the first method but misclassified in the second method and vice versa for \( c_{21} \). Mathematically proven, if \( |z| \geq 1.96 \), the difference can be considered statistically significant (Braun et al., 2011; Kavzoglu et al., 2015).

A radial basis function kernel was used for all the methods. Before classification, its parameters were optimized using a grid search algorithm, based on a five-fold cross validation. The range for optimizing \( C \) was set to \([10^{-2}:10^{3}]\) and for optimizing gamma parameter (\( \sigma \)) was set to \([2^{-2}:2^{-2}]\).

For the experiment, we considered four sizes, i.e., 5%, 10%, 50% and 90% of all samples for the training samples; the remainder were used for quality control. For each size, the selection of training samples was performed randomly during 10 runs from all samples.
Table 5. The average of the kappa coefficient and the STD of OA (inside the parenthesis) of the kernel methods over 10 runs. Last column indicates the average of the kappa and the STD of OA across training sizes.

| Data set   | Methods | 5%     | 10%    | 50%    | 90%    | Mean (STD) |
|------------|---------|--------|--------|--------|--------|------------|
| Flevoland  | SVM     | 0.68 (2.45) | 0.71 (2.83) | 0.78 (3.17) | 0.87 (3.25) | 0.76 (0.08) |
|            | LSSVM   | 0.67 (1.69) | 0.72 (1.98) | 0.79 (2.16) | 0.88 (2.47) | 0.77 (0.09) |
|            | SVM     | 0.68 (4.78) | 0.72 (4.30) | 0.81 (4.24) | 0.89 (3.84) | 0.78 (0.09) |
|            | IVM     | 0.68 (5.23) | 0.73 (4.57) | 0.82 (3.56) | 0.91 (3.43) | 0.79 (0.10) |
| Foulum     | SVM     | 0.65 (4.11) | 0.69 (3.59) | 0.83 (2.55) | 0.94 (1.99) | 0.78 (0.13) |
|            | LSSVM   | 0.63 (4.34) | 0.67 (3.91) | 0.79 (3.20) | 0.87 (2.79) | 0.74 (0.11) |
|            | SVM     | 0.67 (5.65) | 0.67 (5.59) | 0.77 (4.25) | 0.86 (3.87) | 0.74 (0.09) |
|            | IVM     | 0.68 (5.98) | 0.66 (5.55) | 0.82 (3.83) | 0.89 (3.79) | 0.76 (0.11) |
| Winnipeg   | SVM     | 0.57 (3.03) | 0.62 (2.78) | 0.74 (2.17) | 0.83 (1.54) | 0.69 (0.12) |
|            | LSSVM   | 0.55 (3.09) | 0.59 (2.89) | 0.71 (2.49) | 0.80 (2.04) | 0.66 (0.11) |
|            | SVM     | 0.55 (3.45) | 0.61 (3.98) | 0.71 (4.27) | 0.81 (4.85) | 0.67 (0.11) |
|            | IVM     | 0.57 (3.79) | 0.59 (5.40) | 0.70 (5.55) | 0.84 (5.44) | 0.68 (0.12) |

Table 6. $|z|$ value of the McNemar’s test of the kernel methods in relation to SVM.

| Data set   | Methods | 5% | 10% | 50% | 90% |
|------------|---------|----|-----|-----|-----|
| Flevoland  | SVM     | 0  | 0   | 0   | 0   |
|            | LSSVM   | 3.96 | 12.01 | 3.87 | 7.87 |
|            | SVM     | 5.23 | 4.39 | 7.87 | 3.12 |
| Foulum     | SVM     | 6.48 | 14.26 | 11.26 | 9.11 |
|            | LSSVM   | 0  | 0   | 0   | 0   |
|            | SVM     | 11.56 | 14.13 | 12.31 | 11.78 |
|            | LSSVM   | 13.68 | 7.69 | 4.11 | 7.31 |
|            | SVM     | 10.01 | 3.19 | 4.87 | 2.65 |
| Winnipeg   | SVM     | 0  | 0   | 0   | 0   |
|            | LSSVM   | 2.87 | 6.57 | 6.13 | 8.39 |
|            | SVM     | 11.31 | 7.36 | 3.23 | 4.17 |
|            | IVM     | 2.19 | 3.17 | 2.39 | 9.87 |

The training and test samples were normalized for the implementation of the methods. The average of the kappa coefficients and the standard deviation (STD) of OA values over 10 runs were then calculated for each training size. In addition, the average of kappa and the STD of the OA across training sizes were computed (see Table 5). Moreover, the $|z|$ values of the McNemar’s test for comparing the SVM results with the results of other methods, are provided in (Table 6).

Analysis and discussion

In (Table 5), kappa values of the IVM and RVM methods were higher than those of the SVM and LSSVM methods for the Flevoland data set. In addition, their kappa values were higher at different training sizes. However, the STD values of the LSSVM and SVM over 10 runs for each training size, were lower than those of the competitors. Moreover, their STD values across different training sizes were fewer. These issues implied a higher stability of the non-probabilistic kernel methods, compared to the probabilistic versions in classifying the Flevoland data set. In fact, the non-probabilistic kernel methods were less sensitive to the training size for this data set. Of course, as the number of training samples increased, the stability of the SVM and LSSVM decreased and vice versa for the RVM and IVM methods.

The kappa values of the RVM and IVM were higher than those of the SVM and LSSVM at fewer training sizes in the Foulum data set. However, as training sizes increased, the kappa values of the SVM were higher than those of its competitors. In addition, the mean kappa value of the SVM across training sizes was higher than that of the other three methods. Although the kappa value of the LSSVM was less than the IVM, it was approximately equal to or even better than the RVM in some cases for the second data set. The stability of the SVM and LSSVM was higher than two probabilistic methods over 10 runs for each training size. However, the stability of the RVM and IVM across training sizes was higher than the SVM for the Foulum data set.

In the case of the Winnipeg data set, the SVM had higher kappa values than the RVM and IVM methods in almost all training sizes. In addition, the mean kappa value of the SVM across training sizes was higher than that of the RVM and IVM methods. The stability of all the methods across training sizes was almost the same for classification of this data set. This
issue indicated that all methods had an equal sensitivity to the training size in the case of the Winnipeg data set. At each training size, the stability of the SVM and the LSSVM over 10 runs was again higher than the competitors. Unlike the first data set, the SVM and LSSVM were more stable, as the training size was bigger and vice versa for the probabilistic kernel methods.

(Table 6) confirmed that all the differences between the SVM results and other methods were significant for all three PolSAR data sets. Moreover, (Figure 2) illustrates the total time taken for training and testing all the methods (on the Log10 base) with different training sizes. This figure clearly indicated that the LSSVM performed much faster than the other methods, even the closely linked SVM, in classifying all three data sets. This results from the use of the least squares idea in the LSSVM algorithm, which leads to an acceleration in solving the optimization problem.

In the case of three data sets, with a training size of 5%, the speed of the RVM was more than the SVM. However, as the training size increased, the speed of the RVM decreased. For instance, at a size of 90%, it was around 15 times slower than the SVM. The main reason for this was due to the training and testing phases of these two kernel methods. The RVM usually had a longer training process than the SVM, but a shorter testing process, due to its sparser kernels. In the case of a smaller training size, i.e., a bigger test size, the total time of the RVM, consequently, related more to the test time. Thus, it ran much faster. However, when the training size was bigger, i.e., the test size was smaller, its total time related more to the training time and then, it ran slower than the SVM. In contrast to the RVM, the IVM took less time than the RVM and the SVM across almost all training sizes for the classification of the three PolSAR data sets.

**Conclusion**

In this paper, four kernel-based methods, namely two non-probabilistic versions, i.e., the SVM and LSSVM, and two probabilistic versions, i.e., the RVM and IVM were compared in relation to the classification of L-band PolSAR data. Three PolSAR data sets with cropland and non-cropland classes were examined for this analysis. In addition, several different sizes of training samples were used for the experiment.

This paper substantiated that the SVM, as a non-probabilistic kernel method, was more efficient, more accurate and more stable than its probabilistic competitors, i.e., the RVM and IVM in classifying two of the three data sets. Moreover, the LSSVM had a comparable efficiency and accuracy with the probabilistic kernel methods. This conclusion was in contrast to the results recorded in the papers of Braun et al. (2011) and Braun et al. (2012). These two papers had indicated that the RVM and IVM were more efficient than the SVM for hyperspectral data classification.

Another conclusion which can be drawn from this paper concerned the remarkable speed of the LSSVM, compared to the other kernel methods. The LSSVM could obtain almost comparable accuracy with a speed 12 times faster than the associated SVM and around 15 times faster than the probabilistic kernel methods, i.e., the RVM and IVM. Therefore, the LSSVM can be considered as a qualified successor for the SVM in the classification of PolSAR data, with suitable accuracy and greater speed.

In the end, we should take into account that the presented results were valid only for the L-band PolSAR data sets. Therefore, examining these kernel methods for the classification of PolSAR data sets, working in other frequency bands, can be suggested in future work to make a generalized conclusion.
Acknowledgement

The authors would like to present their acknowledgments to the JPL NASA, the MacDonald, Dettwiler and Associates Ltd, the German DLR, and the SMAPVEX 2012 team, the Agriculture and Agri-Food Canada, and Dr. Saed Homayouni, from the Dept. of Geography, Environment, and Geomatics of the University of Ottawa, Canada, for providing the PolSAR and the fields survey data used in this research.

This paper has been supported by the RUDN University Strategic Academic Leadership Program (recipient Yury Razoumny, simulation model development).

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Aghababaei, 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. https://doi.org/10.5721/EurRS20134618

Alberga, V., Satalino, G., & Staykova, D. K. (2008). Comparision of polarimetric SAR observables in terms of classification performance. *International Journal of Remote Sensing*, 29(14), 4129–4150. https://doi.org/10.1080/01431160701840182

Atwood, D. K., Small, D., & Gens, R. (2012). Improving PolSAR land cover classification with radiometric correction of the coherency matrix. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 5(3), 848–856. https://doi.org/10.1109/JSTARS.2012.2186791

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Braun, A. C., Weidner, U., & Hinz, S. (2011). Support vector machines, import vector machines and relevance vector machines for hyperspectral classification—A comparison. In 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), pp. 1–4. IEEE, 2011.

Braun, A. C., Weidner, U., & Hinz, S. (2012). Classification in high-dimensional feature spaces—assessment using SVM, IVM and RVM with focus on simulated EnMAP data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 5(2), 436–443. https://doi.org/10.1109/JSTARS.2012.2190266

Bruzzzone, L., Marconcini, M., Wegmüller, U., & Wiesmann, A. (2004). An advanced system for the automatic classification of multitemporal SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6), 1321–1334. https://doi.org/10.1109/TGRS.2004.826821

Congalton, R. G., & Green, K. (2019). Assessing the accuracy of remotely sensed data: Principles and practices. CRC press.

Dargahi, A., Maghsoudi, Y., & Abkar, A. A. (2013). Supervised classification of polarimetric SAR imagery using temporal and contextual information. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-1/W3, 2013, SMPR 2013, 5 – 8 October 2013, Tehran, Iran, pp. 107-110. doi:10.5194/isprsarchives-XL-1-W3-107-2013

Du, P., Samat, A., Waske, B., Liu, 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. https://doi.org/10.1016/j.isprsjprs.2015.03.002

Hasani, H., Samadzadegan, F., & Reinartz, P. (2017). A metaheuristic feature-level fusion strategy in classification of urban area using hyperspectral imagery and LiDAR data. *European Journal of Remote Sensing*, 50(1), 222–236. Http://earth.esa.int/PolSARpro/datasets. html. Http://uavsar.jpl.nasa.gov/cgi-bin/data.pl. https://doi.org/10.1002/22797254.2017.1314179

Jafari, M., Maghsoudi, Y., & Valadan Zoej, M. J. (2015). A new method for land cover characterization and classification of polarimetric SAR data using polarimetric signatures. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 8(7), 3595–3607. https://doi.org/10.1109/JSTARS.2014.2387374

Kavzoglu, T., Colkesen, I., & Yomralioglu, T. (2015). Object-based classification with rotation forest ensemble learning algorithm using very high resolution worldview-2 image. *Remote Sensing Letters*, 6(11), 834–843. https://doi.org/10.1080/2150704X.2015.1084550

Khosravi, I., Safari, A., & Homayouni, S. (2018). Multiple classifier systems for classification of multifrequency PolSAR images with limited training samples. *International Journal of Remote Sensing*, 39(21), 7547–7567. https://doi.org/10.1080/01431161.2018.1471543

Khosravi, I., Safari, A., Homayouni, S., & McNairn, H. (2017). Enhanced decision tree ensembles for land-cover mapping from fully polarimetric SAR data. *International Journal of Remote Sensing*, 38(23), 7138–7160. https://doi.org/10.1080/01431161.2017.1372863

Ma, X., Shen, H., Yang, J., Zhang, L., & Li, P. (2014). Polarimetric-spatial classification of SAR images based on the fusion of multiple classifiers. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 7(3), 961–971. https://doi.org/10.1109/JSTARS.2013.2265331

McNairn, H., Kross, A., Lapen, D., Caves, R., & Shang, J. (2014). Early season monitoring of corn and soybeans with terraSAR-X and RADARSAT-2. *International Journal of Applied Earth Observation and Geoinformation*, 28, 252–259. https://doi.org/10.1016/j.jag.2013.12.015

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Negri, R. G., Dutra, L. V., & Sant’Anna, S. J. S. (2016). Comparing support vector machine contextual approaches for urban area classification. *Remote Sensing Letters*, 7(5), 485–494. https://doi.org/10.1016/j.rsl.2015.11.007

Niu, X., & Ban, Y. (2014). A novel contextual classification algorithm for multitemporal polarimetric SAR data. *IEEE Geoscience and Remote Sensing Letters*, 11 (3), 681–685. https://doi.org/10.1109/LGRS.2013.2274815

Pal, M., Maxwell, A. E., & Warner, T. A. (2013). Kernel-based extreme learning machine for remote-sensing image classification. *Remote Sensing Letters*, 4(9), 853–862. https://doi.org/10.1080/2150704X.2013.805279

Qi, Z., Yeh, A. G.-O., Li, X., & Zhang, X. (2015). A three-component method for timely detection of
land cover changes using polarimetric SAR images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 107, 3–21. https://doi.org/10.1016/j.isprsjprs.2015.02.004

Safaee, B., & Sahebi, M. R. (2019). A class-based approach to classify PolSAR imagery using optimum classifier. *European Journal of Remote Sensing*, 52(1), 294–307. https://doi.org/10.1080/22797254.2019.1602000

Salehi, M., Sahebi, M. R., & Maghsoudi, Y. (2014). Improving the accuracy of urban land cover classification using radarsat-2 PolSAR data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 7(4), 1394–1401. doi: 10.1109/JSTARS.2013.2273074

Shokrollahi, M., & Ebadi, H. (2016). Improving the accuracy of land cover classification using fusion of polarimetric SAR and hyperspectral images. *Journal of the Indian Society of Remote Sensing*, 44(6), 1017–1024. https://doi.org/10.1007/s12524-016-0559-4

Suykens, J. A. K., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3), 293–300. https://doi.org/10.1023/A:1018628609742

Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1, 211–244 https://www.jmlr.org/papers/volume1/tipping01a/tipping01a.pdf

Uhlmann, S., & Kiranyaz, S. (2014). Integrating color features in polarimetric SAR image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(4), 2197–2206. https://doi.org/10.1109/TGRS.2013.2258675

Wang, H., & Hu, D. 2005. “Comparison of SVM and LS–SVM for regression,” In 2005 International Conference on Neural Networks and Brain, Beijing, China, 13-15 Oct. 2005, (Vol. 1, pp. 279-283). IEEE. doi: 10.1109/ICNNB.2005.1614615

Zhang, A., Yang, X., Jia, L., Ai, L., J., & Dong, Z. (2019). SAR image classification using adaptive neighborhood-based convolutional neural network. *European Journal of Remote Sensing*, 52(1), 178–193. https://doi.org/10.1080/22797254.2019.1579616

Zhu, J., & Hastie, T. (2005). Kernel logistic regression and the import vector machine. *Journal of Computational and Graphical Statistics*, 14(1), 185–205. https://doi.org/10.1198/106186005X25619