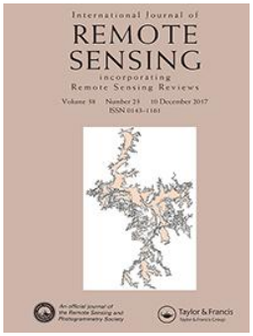

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

Enhanced decision tree ensembles for land-cover mapping from fully polarimetric SAR data

Iman Khosravi, Abdolreza Safari, Saeid Homayouni & Heather McNairn

To cite this article: Iman Khosravi, Abdolreza Safari, Saeid Homayouni & Heather McNairn (2017) Enhanced decision tree ensembles for land-cover mapping from fully polarimetric SAR data, International Journal of Remote Sensing, 38:23, 7138-7160, DOI: 10.1080/01431161.2017.1372863

To link to this article: https://doi.org/10.1080/01431161.2017.1372863

Published online: 31 Aug 2017.

Submit your article to this journal

Article views: 65

View related articles

View Crossmark data

Citing articles: 3 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=tres20

INTERNATIONAL JOURNAL OF REMOTE SENSING, 2017
VOL. 38, NO. 23, 7138–7160
https://doi.org/10.1080/01431161.2017.1372863

Taylor & Francis
Taylor & Francis Group

Check for updates

Enhanced decision tree ensembles for land-cover mapping from fully polarimetric SAR data

Iman Khosravi[a], Abdolreza Safari[a], Saeid Homayouni[b] and Heather McNairn[c]

[a]School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran; [b]Department of Geography, Environment, and Geomatics, University of Ottawa, Ottawa, Canada; [c]Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Canada

ABSTRACT

Fully polarimetric synthetic aperture radar (PolSAR) Earth Observations showed great potential for mapping and monitoring agro-environmental systems. Numerous polarimetric features can be extracted from these complex observations which may lead to improve accuracy of land-cover classification and object characterization. This article employed two well-known decision tree ensembles, i.e. bagged tree (BT) and random forest (RF), for land-cover mapping from PolSAR imagery. Moreover, two fast modified decision tree ensembles were proposed in this article, namely balanced filter-based forest (BFF) and cost-sensitive filter-based forest (CFF). These algorithms, designed based on the idea of RF, use a fast filter feature selection algorithms and two extended majority voting. They are also able to embed some solutions of imbalanced data problem into their structures. Three different PolSAR datasets, with imbalanced data, were used for evaluating efficiency of the proposed algorithms. The results indicated that all the tree ensembles have higher efficiency and reliability than the individual DT. Moreover, both proposed tree ensembles obtained higher mean overall accuracy (0.5–14% higher), producer's accuracy (0.5–10% higher), and user's accuracy (0.5–9% higher) than the classical tree ensembles, i.e. BT and RF. They were also much faster (e.g. 2–10 times) and more stable than their competitors for classification of these three datasets. In addition, unlike BT and RF, which obtained higher accuracy in large ensembles (i.e. the high number of DT), BFF and CFF can also be more efficient and reliable in smaller ensembles. Furthermore, the extended majority voting techniques could outperform the classical majority voting for decision fusion.

ARTICLE HISTORY
Received 19 April 2017
Accepted 22 August 2017

1. Introduction

Land-cover mapping using polarimetric synthetic aperture radar (PolSAR) observations images has recently received considerable attention in the remote sensing community for agro-environmental mapping and monitoring. PolSAR data can be acquired day or night, and are virtually unaffected by weather and atmospheric conditions unlike the

CONTACT Abdolreza Safari asafari@ut.ac.ir School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

© 2017 Informa UK Limited, trading as Taylor & Francis Group

optical remotely-sensed data (Qi et al. 2015; Du et al. 2014). In addition to this oft-quoted vantage, various polarimetric features can be extracted from PolSAR data. These features are scattering components, coherency and covariance matrices, target decompositions, and polarimetric discriminators, to name a few. They provide useful information about physical, geometrical, and phenological properties of ground objects and can lead to increase overall land-cover classification (Alberga 2007; Alberga et al. 2008; Lardeux et al. 2009; Lonnqvist et al. 2010; Du et al. 2014; Du et al. 2015; Tamiminia et al. 2017). However, there may be a relatively high correlation among some polarimetric features (Salehi, Sahebi, and Maghsoudi 2014).

In the past two decades, several well-known classification algorithms such as maximum likelihood (Dargahi, Maghsoudi, and Abkar 2013; Shokrollahi and Ebadi 2016), Wishart maximum likelihood (Lee, Grunes, and Kwok 1994; Ferro-Famil, Pottier, and Lee 2001; Souissi, Doulgeris, and Eltoft 2007; Gao and Ban 2008; Ainsworth, Kelly, and Lee 2009; Lardeux et al. 2009; Lonnqvist et al. 2010; Sánchez-Lladó, Pajares, and López-Martínez 2011; Atwood, Small, and Gens 2012; Qi et al. 2012; Dargahi, Maghsoudi, and Abkar 2013; Salehi, Sahebi, and Maghsoudi 2014; Ma et al. 2014; Jafari, Maghsoudi, and Valadan Zoej 2015; Qi et al. 2015; Shokrollahi and Ebadi 2016), decision tree (DT) (Mishra, Singh, and Yamaguchi 2011; Qi et al. 2012; McNairn et al. 2014; Qi et al. 2015), neural network (Hara et al. 1994; Song, Yun, and Hui 2004; Bruzzone et al. 2004; Alberga et al. 2008; Salehi, Sahebi, and Maghsoudi 2014; Xu, Li, and Brenning 2014), Support Vector Machine (SVM), and kernel methods (Fukuda and Hirosawa 2001; Waske and Van Der Linden 2008; Lardeux et al. 2009; Shah Hosseini et al. 2011; Maghsoudi, Collins, and Leckie 2012; Xiangwei et al. 2013; Souissi, Ouarzeddine, and Belhadj-Aissa 2014; Bai et al. 2014; Niu and Ban 2014; Salehi, Sahebi, and Maghsoudi 2014; Du et al. 2014; Ma et al. 2014; Xu, Li, and Brenning 2014; Uhlmann and Kiranyaz 2014; Qi et al. 2015; Habibi et al. 2016; Shokrollahi and Ebadi 2016) have widely been used for land-cover mapping from PolSAR data. The high correlation among the polarimetric features can cause several difficulties, such as increasing complexity and computational load, in land-cover classification by the above-mentioned methods, except for SVM and kernel methods. SVM method can be effective in this condition by employing kernel functions. However, it may face some other challenges as follows: SVM's efficiency is highly affected by its kernel function's parameters. Moreover, tuning these parameters is a challenging task in data classification. Furthermore, SVM is sensitive to noise or outliers in data (Khosravi and Mohammad-Beigi 2014). Some studies have used feature selection algorithms for improving overall accuracy of land-cover mapping using PolSAR data (Alberga et al. 2008; Lardeux et al. 2009; Haddadi, Reza Sahebi, and Mansourian 2011; Maghsoudi, Collins, and Leckie 2012; Qi et al. 2012; Xiangwei et al. 2013; Salehi, Sahebi, and Maghsoudi 2014; Bai et al. 2014; Shokrollahi and Ebadi 2016). Feature selection algorithms can, in general, reduce data dimensionality for classification, however, some of them cannot singly often lead to significant increase in accuracy.

In the recent years, decision tree ensembles, especially bagged tree (BT) and random forest (RF), as the state-of-art classification methods, have also been used in a few studies for land-cover mapping using PolSAR imagery (e.g. Waske and Van Der Linden 2008; Waske and Braun 2009; Shang et al. 2011; Deschamps et al. 2012; Uhlmann and Kiranyaz 2014; Du et al. 2015). These approaches have several advantages over SVM; they are less sensitive to outliers in data (Breiman 2001). Moreover, RF, as a combination of BT and a feature

selection algorithm, typically performs better and is much less sensitive to the imprecise selection or estimation of its parameters. The feature selection algorithm, used in RF, is the random subset feature selection (RSFS) method. RSFS is a wrapper algorithm, which employs *k*-nearest neighbour (*k*-NN) classifier to evaluate the selected feature subsets (Räsänen and Pohjalainen 2013). Wrapper algorithms although usually lead to produce high accuracy, they are really time-consuming. By contrast, filter algorithms are generally much faster than wrapper algorithms (Zeng et al. 2015). Fast-correlation based filter (FCBF) method, as a filter algorithm, is proved to have an excellent efficiency in high-dimensional feature space (Senliol et al. 2008). Consequently, RSFS can be replaced with FCBF method. Moreover, RF employs a majority voting (MV) technique for fusing decisions. In MV method, a winning class, which has received the highest number of votes by individual DTs, is assigned for a given sample (Mangai et al. 2010). However, two ambiguities may occur during MV procedure. First, two or several classes may have the same highest number of votes for a given sample. Second, no class has the highest number of votes. Accordingly, MV may not explicitly perform in these two conditions. Therefore, it is necessary to resolve these two ambiguities by extending MV technique.

Another challenging issue we may face with in a classification task is the imbalanced data problem. This issue occurs when one or several classes, i.e. minority classes, have much lower number of data compared to other classes, i.e. majority ones (He and Garcia 2009). This problem can be also a discussible issue in land-cover classification using PolSAR data, yet has received virtually no attention in previous PolSAR classification studies. All the classical classification algorithms, even BT and RF are naturally designed to minimize the overall error rate rather than addressing minority classes and thus do not usually perform very well in classification of imbalanced data. To address this problem, we aim to propose two fast enhanced decision tree ensembles called balanced filter-based forest (BFF) and cost-sensitive filter-based Forest (CFF). These two proposed tree ensembles benefit from the main idea of generating RF with several enhancements as follows: (i) using a filter feature selection algorithm, i.e. FCBF method; (ii) embedding two strategies into their structures to resolve the imbalanced data problem; and (iii) proposing two extended majority voting for the proposed ensembles in decision fusion step.

The main objectives of this article can be summarized as follows: (i) comparing the tree ensembles with individual DT for classifying three imbalanced PolSAR datasets; (ii) comparing classical tree ensembles, i.e. BT and RF, with proposed tree ensembles, i.e. BFF and CFF, in terms of overall accuracy, efficiency, reliability, stability and training speed; (iii) evaluating the impact of different training sample sizes on the classification efficiency by all algorithms; and (iv) comparing extended voting techniques and classical majority voting for decision fusion.

2. Methodology

2.1. *Classical ensemble systems*

Ensemble systems are normally referred to as multiple classifier systems in which variants of the same classifier (so-called 'base classifier') are combined by manipulation of training samples or features space, or both (Du et al. 2012). The most frequently used ensemble systems are BT and RF.

2.1.1. *Bagged tree*

BT is a special case of bagging system. In bagging (the acronym of bootstrap aggregating) method, the training samples are randomly, with replacement, divided into several bags or sub samples. Each bag is then used to learn a different type of the same classifier, e.g. DT as the base classifier for BT method. The outputs of the classifiers are finally fused together by employing the MV technique (Du et al. 2012). BT has some parameters, one of which is the most important to be tuned, i.e. the number of decision trees used in the ensemble (nTrees).

2.1.2. *Random forest*

RF integrates the idea of the BT and feature selection method by manipulating both training samples and feature sets. In fact, at each bag of RF algorithm, several features (mTry) are chosen by the RSFS algorithm, from the entire feature set. RSFS uses a random, without replacement, feature search space and employs *k*-NN classifier to evaluate the feature subsets (Räsänen and Pohjalainen 2013; Pohjalainen, Räsänen, and Kadioglu 2015). Unlike BT, RF has two main important parameters to be tuned: nTrees and mTry.

2.2. *Proposed ensemble systems*

Two proposed ensemble systems, BFF and CFF, are designed based on the idea of RF. However, these proposed algorithms are different from the classical RF in three main ways (see Figure 1): feature selection step, decision fusion step, and focus on the imbalanced data. They are as follows:

- Feature selection step:

Unlike RF, which uses a wrapper algorithm (i.e. RSFS), a filter algorithm, called FCBF algorithm is used in both BFF and CFF during the feature selection step.

First proposed by Yu and Liu (2003), FCBF is a multivariate algorithm, which starts with a full set of features. It uses symmetrical uncertainty (SU) to calculate dependency of features, and finds the best subset using a backward selection technique with sequential search strategy. SU is a normalized information theory measure based on the entropy values and is defined as follows (Senliol et al. 2008):

$$\mathrm{SU}(X,Y) = 2\left(\frac{\mathrm{IG}(X,Y)}{H(X)+H(Y)}\right) = 2\left(\frac{H(X)-H(X|Y)}{H(X)+H(Y)}\right), \tag{1}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two feature vectors, $H(\boldsymbol{X})$ and $H(\boldsymbol{Y})$ are their entropy, and $\mathrm{IG}(\boldsymbol{X},\boldsymbol{Y})$, which represents information gain, is their mutual information. SU values are ranged between [0, 1]. A value of 1 indicates that using one feature can completely predict the values of another, while, zero implies that two features are totally independent. FCBF stops when there are no features left to be eliminated (Senliol et al. 2008).

FCBF is a correlation-based algorithm which runs significantly faster than other feature selection algorithms, especially wrapper algorithms such as RSFS. In addition, it

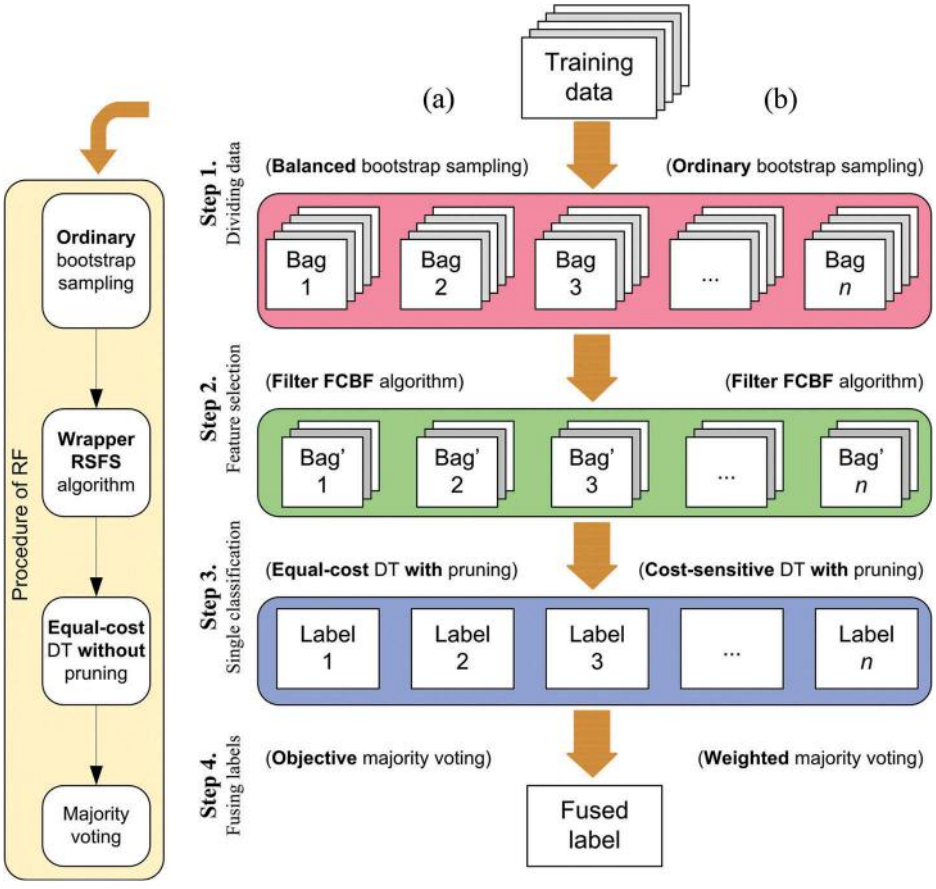

Figure 1. Schematic flowchart of proposed ensembles: (a) BFF and (b) CFF. Comparison with procedure of RF.

has shown high efficiency in detecting the most informative features in high-dimensional data (Yu and Liu 2003; Senliol et al. 2008).

- Decision fusion step:

Unlike RF, in which an MV is used, two extended majority voting techniques are proposed for the proposed ensembles in the decision fusion step.

As previously said, two ambiguities may occur during the MV procedure. First, two or several classes may have the same highest number of votes for a given sample. In other words, the mode of decisions of individual DTs is non-unique. As an example, Figure 2(a) shows two classes of w_2 and w_6, which have received three votes by the DTs for sample x. Which of w_2 or w_6 should be selected as winning class for x? Second, no class has the highest number of votes. In other words, the individual DTs have the diverse decisions about some samples, for example, for y in Figure 2(b). Which classes should be selected as winning class for y?

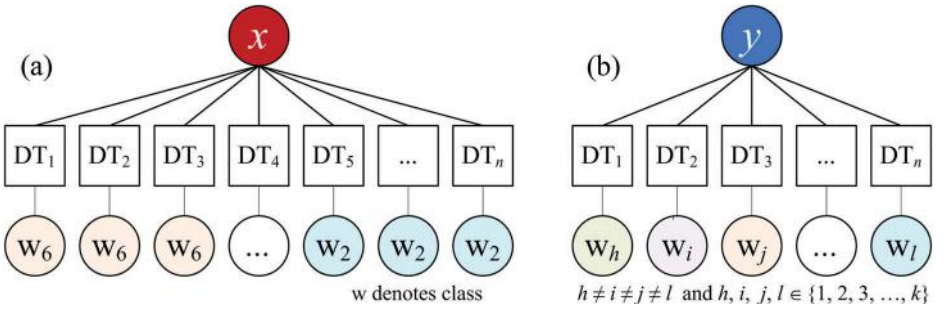

Figure 2. Two ambiguities may occur in MV procedure. (a) First ambiguity: e.g. for sample *x*, two classes of w_2 and w_6 have the same highest number of votes. (b) Second ambiguity: e.g. for sample *y*, all the classes have vote of 1.

Due to the above-mentioned ambiguities, MV accordingly cannot explicitly decide for all samples. Therefore, this article proposes an objective majority voting (OMV), as an extended version of MV, with a goal to objectively resolve these ambiguities by offering two practical strategies. As can be seen in Figure 3(a), for a test sample *p*, OMV may perform in one of the following ways:

(1) In an unequivocal case, when the number of votes for class w_i is more than or equal to $n/2 + 1$ (n is the number of the DTs) or the maximum vote, then OMV operates similar to MV and assigns w_i to *p*.
(2) When the number of votes for two or several classes (w_i, w_j, ... w_l; $i \neq j \neq l$, $i, j, l \in \{1, 2, 3, \ldots, k\}$) is equal together (first ambiguity in MV), then OMV performs based

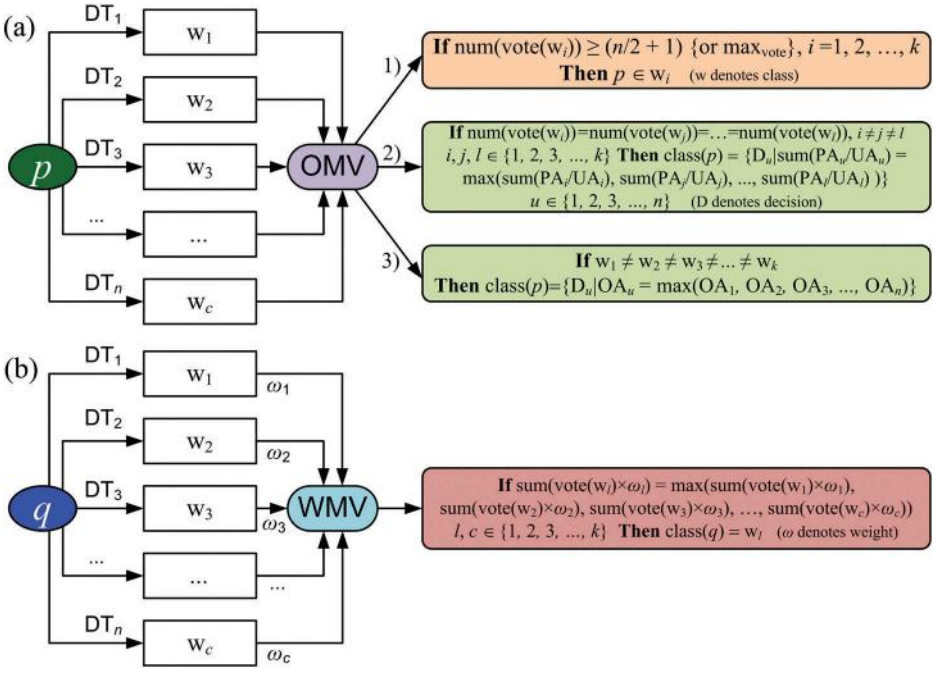

Figure 3. Schematic flowchart of proposed voting techniques: (a) OMV and (b) WMV.

on the ratio of the producer's accuracy (PA) and the user's accuracy (UA) of the DTs for respective classes as follows: p belongs to class obtained by D_u (decision of DT_u, $u \in \{1, 2, 3, \ldots, n\}$), if the summation of PA_u/UA_u is maximum among all PA/UA.

(3) When the individual DTs have the diverse decisions about sample p (second ambiguity in MV), OMV then operates based on the overall accuracy (OA) of the DTs as follows: p belongs to the class which is output by the DT having the highest OA.

Weighted majority voting (WMV) is another extended version of MV, which is more competent for decision fusion when the classifiers operate in the dissimilar conditions or the weights of data and/or classes are not equal. Here, the WMV employs a weight (ω_i, $i = 1, 2, 3, \ldots, k$) for each class. Then, a winning class, which has received the highest number of weighted votes by the individual DTs, is assigned to the test sample q (see Figure 3(b)). In this article, OMV is employed for BFF, and WMV for CFF. In the next section, we will describe how the weight of each class is determined.

- Imbalanced data problem:

Unlike RF, which is not naturally designed for solving the imbalanced data problem, several strategies are considered in the structure of the proposed ensembles.

Generally, two common strategies are considered in literature to solve the problem of imbalanced data classification (He and Garcia 2009). First strategy is to use the sampling techniques, either of under-sampling the majority class or over-sampling the minority class, to establish an artificial balance in the distribution of the classes. The most important sampling method, i.e. random under-sampling method, establishes the balance between the data by disregarding some of the majority class samples and random over-sampling method aims to establish the balance between the data by randomly replicating or reusing the minority class samples (Longadge and Dongre 2013).

Second strategy is based on the cost-sensitive learning. In contrast to sampling techniques, cost-sensitive learning considers the costs associated with misclassifying samples (He and Garcia 2009). The foundation of the cost-sensitive learning is based on the cost matrix. It can be defined by the penalty of classifying samples from one class to another. For example, in a two-class task, C(Maj, Min) is defined as the cost of misclassifying a minority samples as a majority one and C(Min, Maj) is the reverse. Certainly, there is no cost for correct classification of either class. Moreover, in an imbalanced data problem, the misclassification of cost of the minority samples is higher than the majority samples ones, i.e. C(Maj, Min) > C(Min, Maj). The cost matrix concept can be extended to a multi-class task (see Figure 4).

In this case, $C(i, j)$ represents the cost of predicting class w_i when the true class is w_j ($w_i, w_j \in \boldsymbol{Y} = \{1, 2, \ldots, k\}$), and $C(i, i) = 0$. Then, the cost-sensitive learning aims to develop a hypothesis that minimizes the overall cost on the training samples, which is usually the Bayes conditional risk. It is defined as $R(i|x) = \Sigma j\ P(j|x) \times C(i, j)$, where $P(j|x)$ indicates the probability of each class w_j for a given sample x (He and Garcia 2009).

The modified form of first strategy is employed in BFF sampling phase and the second strategy is embedded into CFF learning phase.

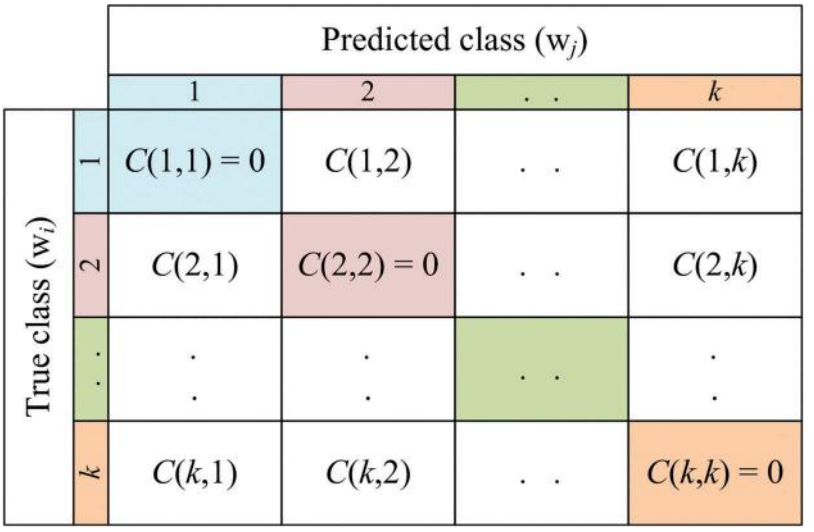

Figure 4. Cost matrix for a multi-class problem (He and Garcia 2009).

Based on the three differences above mentioned, two proposed decision tree ensemble algorithms operate as follows:

2.2.1. *Balanced filter-based forest (BFF)*

BFF has the following steps (see Figure 1(a)): similar to RF during the first step, the training samples are divided into several bags by the bootstrap re/sampling method. The bootstrap method is a uniformly random sampling method with replacement. In each bag, around 60–75% of the entire training samples are randomly drawn (Du et al. 2012). Due to the sampling scheme with replacement, some observations may be repeated in each bag. However, for a training procedure using the imbalanced data, there is a high probability that each selected bag contains few or even none of the minority classes. This can lead to building a DT with weak efficiency for the classification of these classes. To solve this problem, we do not employ two above-mentioned sampling techniques. Since there are several concerns about these methods. Under-sampling methods may lose the valuable majority samples in process of removing. The duplicated or synthetically generated minority data may be also improper representatives of the minority class and be then useless for classification in over-sampling methods (Longadge and Dongre 2013). Therefore, this article proposes a different sampling technique called as balanced bootstrap sampling. It performs as follows: for each bag, a bootstrap sample is first selected from the minority class. The same number of cases is then randomly chosen with replacement from the other classes.

After generating the bags, the FCBF algorithm is employed to select the optimum feature subsets for each bag. Unlike the RF, it is not necessary to determine the number of features that should be selected for each bag by the FCBF in the BFF. Then, an individual DT classifier is trained using each bag. In contrast to the RF, we allow the

individual DTs to be pruned after growing in the BFF. Finally, the outputs of the individual DTs are fused by the proposed OMV algorithm.

2.2.2. *Cost-sensitive filter-based forest (CFF)*

CFF has these steps (see Figure 1(b)): a simple bootstrap sampling, similar to RF, is used to generate the training bags. Afterwards, similar to the BFF, the FCBF algorithm is employed to select the optimum feature subsets for each bag. An individual DT classifier with pruning is then trained using each bag. In BFF, each individual DT was trained with equal costs for all the classes. However, in CFF, the cost-sensitive learning is employed in the individual DTs learning phase.

Since each DT tends to be biased towards the majority classes, we should consider a heavier penalty on the misclassification of the minority classes. Therefore, for each bag, a square cost matrix, with the size of $k \times k$, where k is the number of classes, is established in such a way that the minority classes receive a higher misclassification cost than the majority classes. This matrix is then injected to the learning phase of each tree. After training all the DTs, their outputs are fused by the WMV algorithm, in which, the weight of each class is the normalized sum of the respective rows of that class in the cost matrix.

3. Datasets and pre-processing

3.1. *PolSAR data*

Three PolSAR datasets acquired by different SAR sensors over different areas are used in this article for efficiency evaluation of the proposed approach (see Table 1).

First dataset was acquired by the airborne AIRSAR sensor in L-Band over the Flevoland, the Netherlands in 1989 with a nominal spatial resolution of 10 m. It covers a large agricultural area of flat topography and homogeneous soils. The second dataset was collected by the airborne EMISAR sensor in L-Band over Foulum, a village in Denmark, on 17 April 1989. The data has a 5 m spatial resolution and the region consists of forest, agricultural, and urban areas. The first and the second datasets can be downloaded at the European Space Agency (ESA) Website (http://earth.eo.esa.int/PolSARpro/datasets. html). The final dataset was acquired by the airborne UAVSAR L-Band sensor over an annual agricultural cropland area in Winnipeg, Manitoba, Canada, on 29 June 2012. This image has a 12 m spatial resolution. The third data can be downloaded at the UAVSAR NASA JPL website (http://uavsar.jpl.nasa.gov/cgi-bin/data.pl). The Pauli RGB and the reference land-cover maps for these datasets are shown in Figure 5.

As the pre-processing step, the refined Lee filter with a window size of 7×7 is applied to all three images to reduce the speckle. This filter can preserve the edges, as well as the polarimetric properties of the ground targets (Lee, Grunes, and De Grandi

Table 1. Properties of the PolSAR datasets.

Dataset	Platform	Properties	Resolution (m)	Number of looks
Flevoland, Netherlands	AIRSAR	Airborne PolSAR mode	10×10	4×4
Foulum, Denmark	EMISAR	Airborne PolSAR mode	5×5	1×1
Winnipeg, Canada	UAVSAR	Airborne PolSAR mode	12×12	2×2

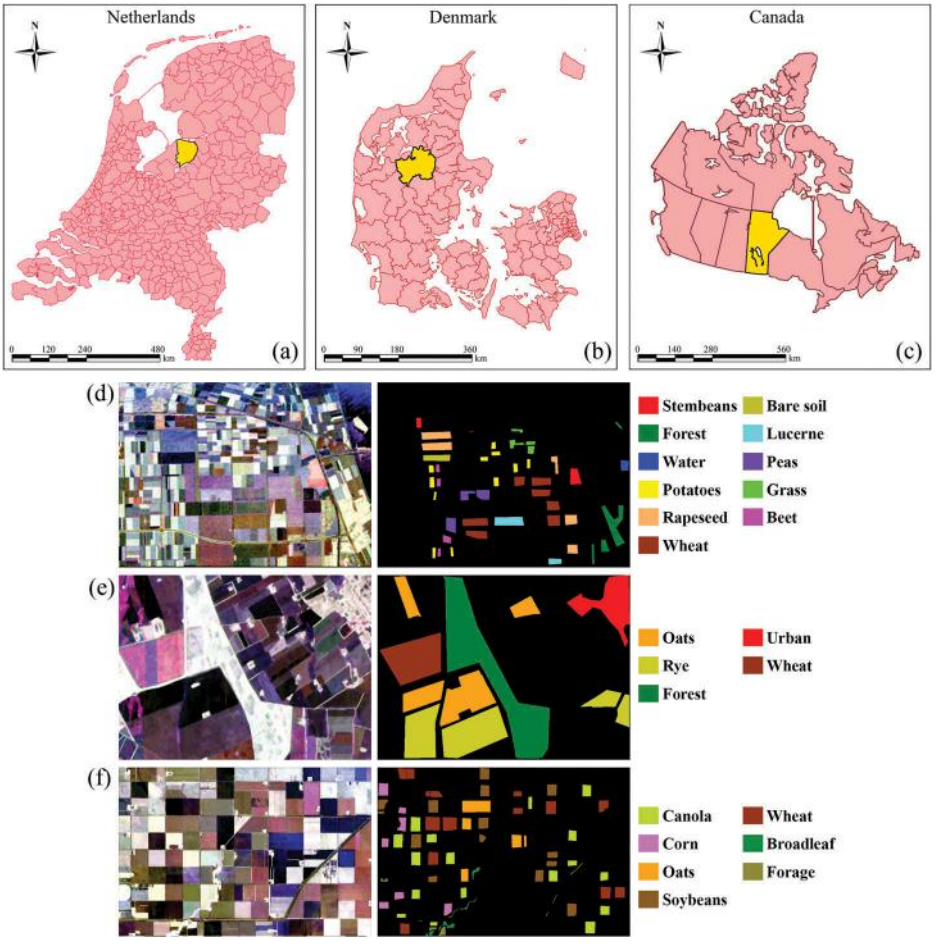

Figure 5. Three PolSAR datasets used in this paper, Flevoland, Foulum, and Winnipeg: (a), (b), and (c) show their respective geographical locations, (d), (e), and (f) Pauli RGB colour composites (blue: |HH + VV|, green: 2|HV|, red: |HH − VV|) and the reference maps of the PolSAR datasets.

1999). This filter and kernel size produce the highest overall accuracy based on our experimental tests.

3.2. *Training and test samples*

According to the reference map of these datasets, there are 11 land covers for Flevoland as follows: stembeans (St), forest (Fo), water (Wa), potatoes (Po), rapeseed (Ra), wheat (Wh), bare soil (Ba), lucerne (Lu), peas (Pe), grass (Gr), and beet (Be), 5 land covers for Foulum as follows: oats (Oa), rye (Ry), forest (Fo), urban (Ur), and wheat (Wh), and 7 crop types for Winnipeg as follows: canola (Ca), corn (Co), oats (Oa), soybeans (So), wheat (Wh), broadleaf (Br), and forage (Fo). Table 2 presents the number of samples for each class and its percent in relation to all samples. For the experiment, we choose 10% of all samples for training algorithms and the remaining are used for the accuracy assessment. The number of samples for classes of all three datasets is significantly imbalanced. In

Table 2. All samples of classes for the PolSAR datasets.

(a) Flevoland			(b) Foulum			(c) Winnipeg		
Class	Number	Ratio (%)	Class	Number	Ratio (%)	Class	Number	Ratio (%)
St	4326	5.89	Oa	35,034	20.78	Ca	30,818	24.99
Fo	8114	11.05	Ry	33,678	19.97	Co	15,425	12.51
Wa	2463	3.36	Fo	63,529	37.68	Oa	15,091	12.23
Po	5791	7.89	Ur	20,520	12.17	So	26,116	21.17
Ra	11,387	15.51	Wh	15,847	9.40	Wh	30,192	24.48
Wh	14,704	20.03				Br	3655	2.96
Ba	7948	10.83				Fo	2047	1.66
Lu	5219	7.11						
Pe	5492	7.48						
Gr	4942	6.73						
Be	3019	4.11						

Flevoland, the ratio of 1 to 7, in Foulum, the ratio of 1 to 4, and in Winnipeg, the ratio of 1 to 15 are established between the minority and majority classes. Water, beet, stem-beans, grass, lucerne, peas, and potatoes in Flevoland, wheat, and urban in Foulum, and forage, and broadleaf in Winnipeg, are the minority classes.

3.3. *Polarimetric feature extraction*

A variety of useful polarimetric features can be extracted from PolSAR data for land-cover classification. Features related to the scattering, coherency and covariance matrices and features obtained by coherent and incoherent target decomposition methods have shown potential for classification of various land covers (Tamiminia et al. 2017).

The primary features that can be extracted from the PolSAR imagery and are useful for classifying diverse land covers are the linear backscatter intensities ($\boldsymbol{\sigma}_{hh}$, $\boldsymbol{\sigma}_{hv}$, and $\boldsymbol{\sigma}_{vv}$), the ratio of linear polarizations ($\boldsymbol{R}_{hhvv}$, $\boldsymbol{R}_{hvhh}$, and $\boldsymbol{R}_{hvvv}$) and the ratio values of linear polarizations to total backscattering power, i.e. span ($\boldsymbol{R}_{hh}$, $\boldsymbol{R}_{hv}$, and $\boldsymbol{R}_{vv}$). In addition, based on the polarization synthesize theory, the circular backscatter intensities ($\boldsymbol{\sigma}_{rr}$, $\boldsymbol{\sigma}_{rl}$, and $\boldsymbol{\sigma}_{ll}$), the ratio of circular polarizations ($\boldsymbol{R}_{rrll}$, $\boldsymbol{R}_{rlrr}$, and $\boldsymbol{R}_{rlll}$) and the ratio values of circular polarizations to total power ($\boldsymbol{R}_{rr}$, $\boldsymbol{R}_{rl}$, and $\boldsymbol{R}_{ll}$) can be helpful. Furthermore, the linear ($\boldsymbol{\rho}_{hhvv}$, $\boldsymbol{\rho}_{hvhh}$, $\boldsymbol{\rho}_{hvvv}$, and $\boldsymbol{r}$) and circular ($\boldsymbol{\rho}_{rrll}$, $\boldsymbol{\rho}_{rlrr}$, $\boldsymbol{\rho}_{rlll}$, and $\boldsymbol{r}^{cir}$) correlation coefficients are important features for discriminating natural areas from urban areas (Nghiem et al. 1992; Moriyama et al. 2004; Lardeux et al. 2009; Lonnqvist et al. 2010). Here, 'h' and 'v' denote the horizontal and vertical linear polarization basis, and 'r' and 'l' denote the right-hand and left-hand circular polarization basis.

In addition to those described above, the features extracted by the coherent and incoherent target decomposition methods can provide useful information about the scattering mechanisms of land-cover classes and objects. Pauli ($|\boldsymbol{\alpha}|^2$ and $|\boldsymbol{\beta}|^2$) and Krogager ($|\boldsymbol{k}_d|^2$ and $|\boldsymbol{k}_h|^2$) decomposition parameters are the most important coherent decompositions features (Krogager 1990; Cloude and Pottier 1996). By contrast, Freeman-Durden ($\boldsymbol{P}_s$ and $\boldsymbol{P}_d$), Yamaguchi ($\boldsymbol{P}_s^Y$, $\boldsymbol{P}_d^Y$, $\boldsymbol{P}_v^Y$, and $\boldsymbol{P}_c^Y$), and Cloude and Pottier parameters including entropy ($\boldsymbol{H}$), anisotropy ($\boldsymbol{A}$) and alpha ($\bar{\boldsymbol{\alpha}}$), four combinations of $\boldsymbol{H}$ and $\boldsymbol{A}$, i.e. $\boldsymbol{HA}$, $\boldsymbol{H}(1-\boldsymbol{A})$, $(1-\boldsymbol{H})\boldsymbol{A}$, and $(1-\boldsymbol{H})(1-\boldsymbol{A})$, eigenvalues of the coherency matrix ($\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$, and $\boldsymbol{\lambda}_3$), pedestal height ($\boldsymbol{\psi}$), and radar vegetation index (RVI),

are the most important incoherent polarimetric features (Lee and Pottier 2009; Cloude and Pottier 1997; Freeman and Durden 1998; Yamaguchi et al. 2005). It is noteworthy that the third Pauli ($|\boldsymbol{y}|^2 = 2|\boldsymbol{S}_{hv}|^2$), third Freeman-Durden ($\boldsymbol{P}_v = |\boldsymbol{S}_{hv}|^2$) and the first Krogager ($|\boldsymbol{k}_s|^2 = |\boldsymbol{S}_{rl}|^2$) parameters are not employed in this paper, because they are equal to the backscatter intensities. Here, the subscripts 's', 'd', and 'v' denote the surface, double-bounce, and volume scattering powers, respectively. Moreover, the subscripts 'h' and 'c' denote helix scattering power for Krogager and Yamaguchi (Y) decompositions, respectively.

In general, although each of the above-mentioned features may be useful for classifying or discriminating a few or some land-cover classes, they may be less useful for discriminating others. In this work, we aim to stack all these features together to obtain high efficiency and accuracy for discriminating all land covers. Therefore, in total, 48 polarimetric features are extracted from each dataset (see Table 3).

4. Implementation and results

The experimental tests performed were as follows: (i) the tree ensemble systems were compared with the individual DT in terms of accuracy; (ii) the classical ensembles, the BT and the RF, were compared with two proposed ensembles, the BFF and the CFF, at nTrees of [5, 10:10:100, 200] in terms of accuracy and performance speed. For the RF,

Table 3. Polarimetric features extracted by the PolSAR datasets.

Description	Feature symbols and formulas
Backscattering intensities (dB)	$\sigma_{hh} = 10\log_{10}\lvert S_{hh}\rvert^2$, $\sigma_{hv} = 10\log_{10}\lvert S_{hv}\rvert^2$, $\sigma_{vv} = 10\log_{10}\lvert S_{vv}\rvert^2$ {h: horizontal, and v: vertical} $\sigma_{rr} = 10\log_{10}\lvert S_{rr}\rvert^2$, $\sigma_{rl} = 10\log_{10}\lvert S_{rl}\rvert^2$, $\sigma_{ll} = 10\log_{10}\lvert S_{ll}\rvert^2$ {r: right-handed, and l: left-handed}
Polarization ratio (dB)	$R_{hhvv} = 10\log_{10}(\lvert S_{hh}\rvert^2/\lvert S_{vv}\rvert^2)$, $R_{hvhh} = 10\log_{10}(\lvert S_{hv}\rvert^2/\lvert S_{hh}\rvert^2)$, $R_{hvvv} = 10\log_{10}(\lvert S_{hv}\rvert^2/\lvert S_{vv}\rvert^2)$ $R_{rrll} = 10\log_{10}(\lvert S_{rr}\rvert^2/\lvert S_{ll}\rvert^2)$, $R_{rlrr} = 10\log_{10}(\lvert S_{rl}\rvert^2/\lvert S_{rr}\rvert^2)$, $R_{rlll} = 10\log_{10}(\lvert S_{rl}\rvert^2/\lvert S_{ll}\rvert^2)$
Ratio values	$R_{hh} = \lvert S_{hh}\rvert^2/\text{span}*$, $R_{hv} = \lvert S_{hv}\rvert^2/\text{span}$, $R_{vv} = \lvert S_{vv}\rvert^2/\text{span}$ $R_{rr} = \lvert S_{rr}\rvert^2/\text{span}$, $R_{rl} = \lvert S_{rl}\rvert^2/\text{span}$, $R_{ll} = \lvert S_{ll}\rvert^2/\text{span}$
Correlation coefficients modulus	ρ_{hhvv}, ρ_{hvhh}, ρ_{hvvv}, ρ_{rrll}, ρ_{rlrr}, ρ_{rlll}, r^{cir}
Cloud and Pottier parameters	$H = -\sum_{i=1}^{3} p_i \log_3 p_i **$, $A = (\lambda_2 - \lambda_3)/(\lambda_2 + \lambda_3)$, $\bar{\alpha} = \sum_{i=1}^{3} p_i a_i$, HA, $H(1-A)$, $(1-H)A$, $(1-H)(1-A)$, $\lambda_1, \lambda_2, \lambda_3, \psi = \{\min(\lambda_1 + \lambda_2 + \lambda_3)\}/(\lambda_1 + \lambda_2 + \lambda_3)$, $\text{RVI} = 4\lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$
Pauli parameters	$\lvert a\rvert^2 = \lvert(S_{hh} + S_{vv})/\sqrt{2}\rvert^2$, $\lvert\beta\rvert^2 = \lvert(S_{hh} - S_{vv})/\sqrt{2}\rvert^2$
Krogager parameters	$\lvert k_d\rvert^2 = \min(\lvert S_{rr}\rvert^2, \lvert S_{ll}\rvert^2)$, $\lvert k_h\rvert^2 = \text{abs}(\lvert S_{rr}\rvert^2 - \lvert S_{ll}\rvert^2)$
Freeman-Durden parameters	$P_s = f_s(1 + \lvert\beta\rvert^2)$, $P_d = f_d(1 + \lvert a\rvert^2)$ {s: surface, and d: double-bounce scattering}
Yamaguchi parameters	$P_s^Y = f_s(1 + \lvert\beta\rvert^2)$, $P_d^Y = f_d(1 + \lvert a\rvert^2)$, $P_v^Y = f_v$, $P_c^Y = f_c$ {v: volume, and c: helix scattering}

$*\text{span} = \lvert S_{hh}\rvert^2 + 2\lvert S_{hv}\rvert^2 + \lvert S_{vv}\rvert^2 = \lvert S_{rr}\rvert^2 + 2\lvert S_{rl}\rvert^2 + \lvert S_{ll}\rvert^2$, $**p_i = \lambda_i / \sum_{k=1}^{3} \lambda_k$

mTry = 7, i.e. the round of the square root of all features, 48, according to the previous studies (e.g. Breiman 2001; Hastie, Tibshirani, and Friedman 2009); (iii) the impact of different sizes of the training samples on the classification accuracy was evaluated, and (iv) the comparative analysis of OMV and WMV with MV, in terms of accuracy, was also performed. For accuracy assessment, we used the common metrics namely OA, PA, and UA, extracted from the confusion matrix. OA is a generic performance metric. Moreover, a higher PA of a class indicates a better efficiency of an algorithm in detecting that class, while a high UA implies the high reliability of the results produced by the algorithm (Congalton 1991; Khoshelham et al. 2010). Table 4 presents the evaluation results of the algorithms on the PolSAR data. Figures 6 and 7 demonstrate the OA plots and training time of the tree ensembles at all the nTrees. Figures 8–10 also demonstrate the PA and UA plots averaged over nTrees for the tree ensembles and the PA and the UA of the individual DT for Flevoland, Foulum, and Winnipeg datasets, respectively.

4.1. *Tree ensembles* versus *individual tree*

From Table 4, one could see that the OA mean values (over the nTrees) of all the tree ensembles were higher than the OA of the individual DT: around 2–9% higher for Flevoland dataset (the lowest difference), 21–35% higher for Foulum dataset (the highest difference) and 10–16% for Winnipeg dataset. All the tree ensembles were also more efficient than the individual DT in discriminating the classes in Flevoland, Foulum and Winnipeg datasets due to higher PA. Moreover, due to higher UA, all the tree ensembles have obtained more reliable results than the individual DT for the classes in all three datasets.

Figures 8–10 also indicated higher PA and UA values of all the tree ensembles, especially BFF and CFF, compared to the individual DT for most of the classes. For example, for Flevoland dataset, stembeans, forest, potatoes, wheat, lucerne, peas, grass, and beet, for Foulum dataset, rye, forest, urban, and wheat, and for Winnipeg dataset, canola, oats, and wheat classes have been classified more accurately (thanks to higher PA values) and more reliably (thanks to higher UA values) by all or most of the tree ensembles when compared to the individual DT.

Figures 8(a)–10(a) demonstrated that the tree ensembles were significantly more efficient (with around 20–45% higher PA values) when compared to the individual DT

Table 4. Evaluation results of all algorithms for all datasets (highest values among the methods for each metric were bolded. SD: standard deviation).

		DT	BT	RF	BFF	CFF
OA (%)	Flevoland	76.39	78.76	84.41	**85.49**	**84.89**
{mean over nTrees}	Foulum	57.89	78.89	85.24	**87.31**	**92.59**
	Winnipeg	71.25	80.75	82.86	**85.15**	**86.87**
OA (%)	Flevoland	–	0.78	1.16	**0.57**	**0.33**
{SD over nTrees}	Foulum	–	1.39	0.93	**0.58**	**0.39**
	Winnipeg	–	1.32	1.89	**1.25**	**0.68**
Mean PA (%)	Flevoland	78.53	80.41	85.82	**87.11**	**85.96**
{mean over classes)	Foulum	63.67	81.07	87.05	**88.38**	**91.79**
	Winnipeg	69.61	75.31	77.92	**80.18**	**81.01**
Mean UA (%)	Flevoland	77.06	79.21	**86.28**	**87.29**	85.56
{mean over classes)	Foulum	62.69	82.80	89.34	**89.66**	**91.10**
	Winnipeg	67.07	78.61	82.54	**86.81**	**85.58**

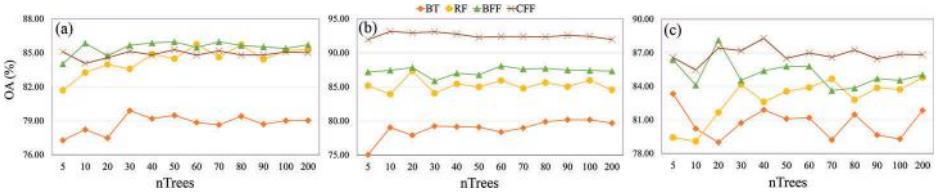

Figure 6. OA (%) plots obtained by BT (■), RF (●), BFF (▲), and CFF (×) for: (a) Flevoland, (b) Foulum, and (c) Winnipeg.

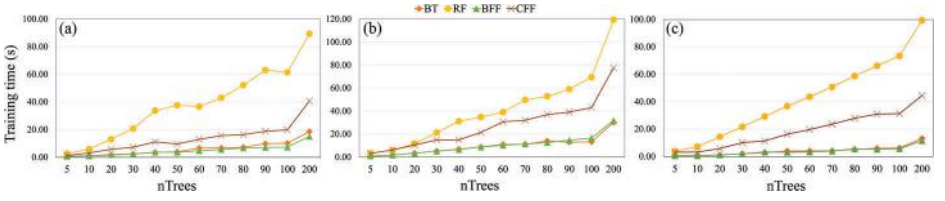

Figure 7. Training time (s) for BT (■), RF (●), BFF (▲), and CFF (×) for: (a) Flevoland, (b) Foulum, and (c) Winnipeg.

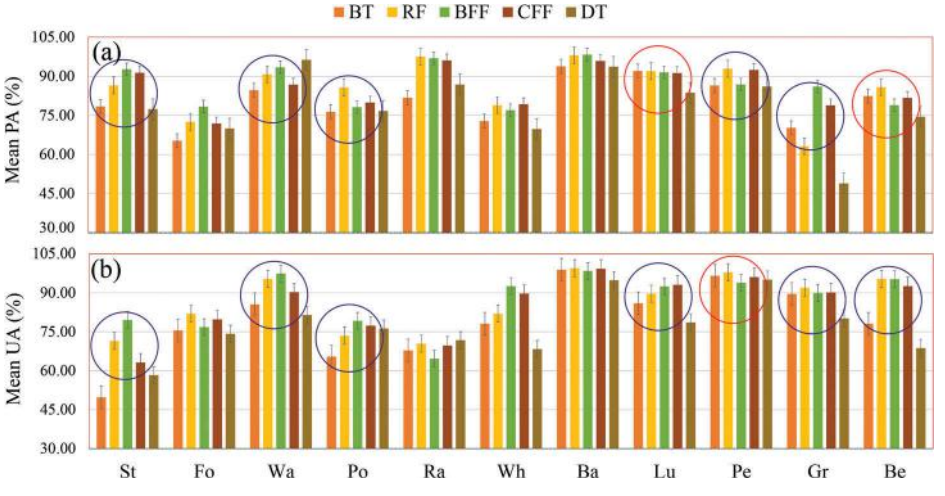

Figure 8. (a) PA (%) and (b) UA (%) Plots averaged over nTrees for the tree ensembles and the individual DT algorithms for Flevoland dataset.

in separating the grass lands for Flevoland, the oats, rye, and urban areas for Foulum, and the canola, corn and oats crops for Winnipeg. Moreover, from Figures 8(b)–10(b), higher reliability of the ensembles (with around 15–60% higher UA values) compared to the individual DT was reported in separating water, wheat, lucerne, and beet for Flevoland, as well as rye, forest, urban, and wheat for Foulum and the canola, oats, and broadleaf for Winnipeg.

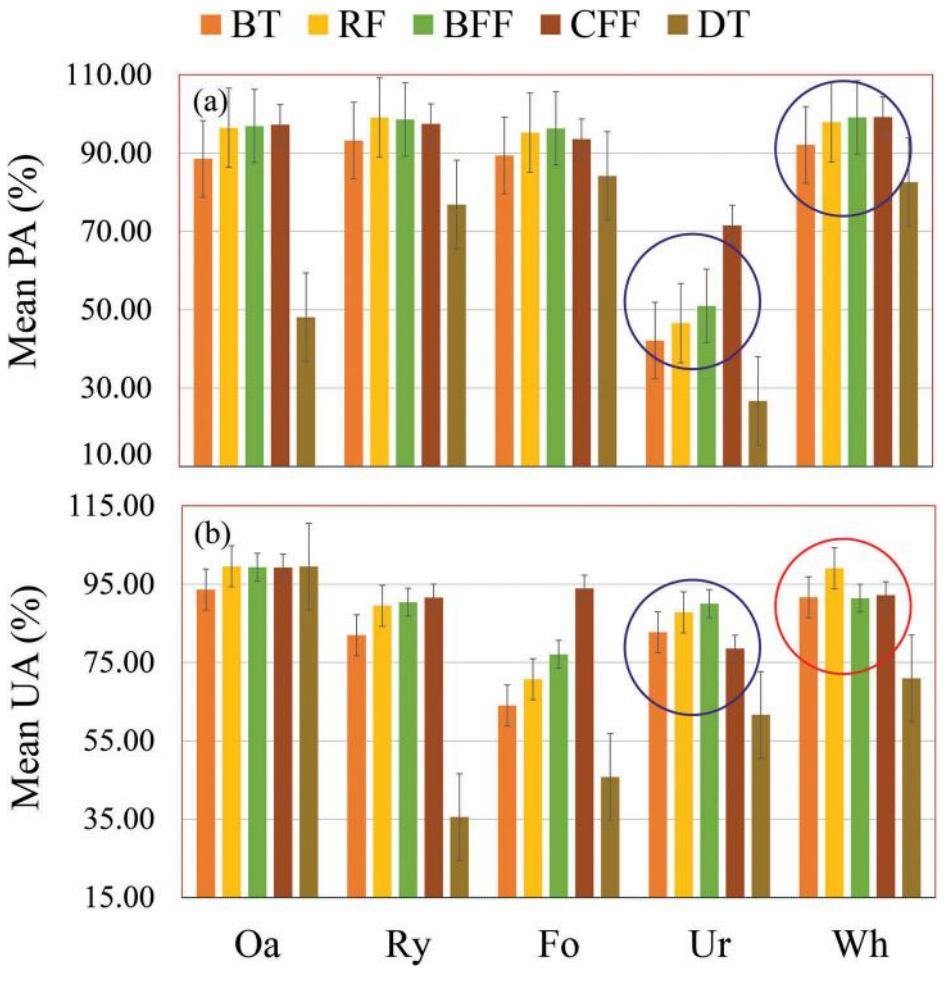

Figure 9. (a) PA (%) and (b) UA (%) Plots averaged over nTrees for the tree ensembles and the individual DT algorithms for Foulum dataset.

Overall, the tree ensembles outperformed the individual DT. However, for a few classes in the Flevoland and Winnipeg datasets, less accurate classification results were obtained by the ensembles. For example, all the tree ensembles had lower efficiency (lower PA values) than the individual DT in separating water class in Flevoland dataset, although they had more reliability (higher UA values). By contrast, the individual DT was more reliable in separating soybeans in Winnipeg dataset although with a lowest efficiency. For forage class in Winnipeg dataset, both RF and CFF had lower efficiency and lower reliability, BFF had lower efficiency, and BT had lower reliability than the individual DT. Although, the individual DT obtained slightly better efficiency (around 4% higher PA) than BT, RF, and CFF in classification of the broadleaf class in Winnipeg, it had the worst reliability (around 90% lower UA than all the tree ensembles).

Based on these results, we can conclude that the tree ensemble systems can often obtain more accurate and more reliable results as compared to the individual DT from the PolSAR data in a relatively high-dimensional feature space.

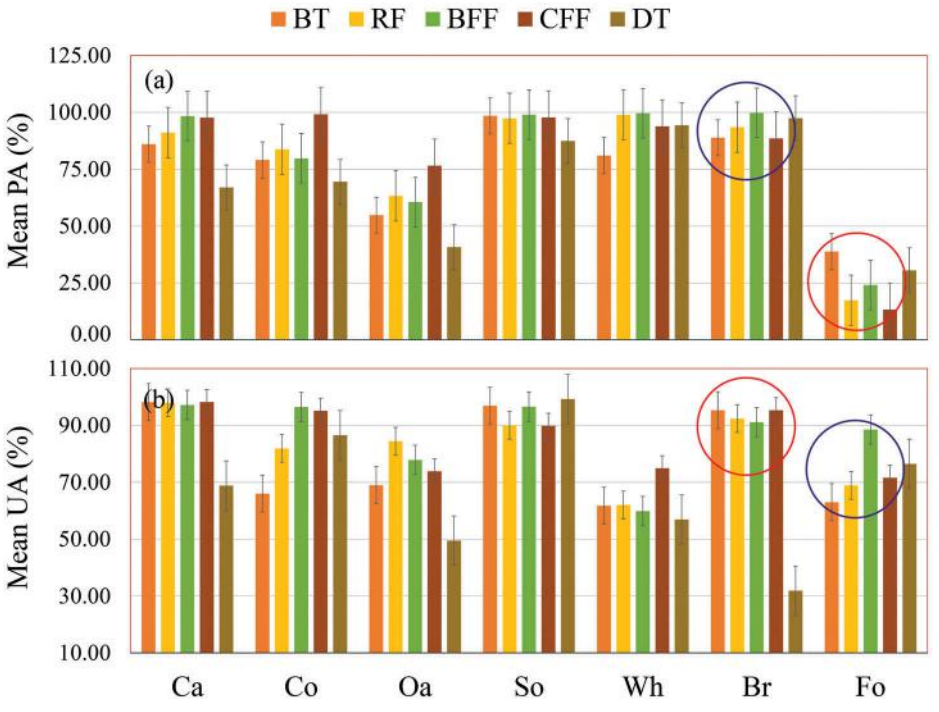

Figure 10. (a) PA (%) and (b) UA (%) Plots averaged over nTrees for the tree ensembles and individual DT algorithms for Winnipeg dataset.

4.2. *Classical* versus *proposed tree ensembles*

From Table 4, the proposed ensembles, i.e. BFF and CFF, had higher OA mean values than BT (4–14%) and RF (0.5–8%) respectively, for all three datasets. In addition, for all the nTrees of all datasets, BFF and CFF had much higher OA than BT (around 2–13% higher). Furthermore, they had higher OA than RF for all the nTrees in the Foulum dataset (approximately 0.5–6% higher), and for most of them in the Flevoland and Winnipeg datasets (see Figure 6). Regarding comparison of two proposed ensembles together, CFF often obtained higher OA than BFF for most of the nTrees in the Foulum and Winnipeg and vice versa in Flevoland.

In term of robustness, both BFF and CFF were more stable than BT and RF for all datasets, especially for Foulum, thanks to their lower OA standard deviations (see Table 4). This issue could be also clearly observed in Figure 6 where the trend lines of both BFF and CFF were closer to the constant horizontal line as compared to RF and BT. Once more between BFF and CFF, the latter was more stable in all three datasets. Another result was that the mean PA and the mean UA values of both BFF and CFF were higher than those of BT and RF for all three datasets, the exception being the mean UA of the CFF for Flevoland (see Table 4).

Regarding the results of minority classes, Figure 8 indicated that in Flevoland dataset, the proposed ensembles had more PA values for the classification of five of seven minority classes (i.e. stembeans, water, potatoes, peas, and grass). They had also more

UA values in the classification of six of seven minority classes (i.e. stem beans, water, potatoes, lucerne, grass, and beet) than all or some classical ensembles. Figure 9 also confirmed the relative success of the proposed ensembles in the classification of minority classes in Foulum dataset, i.e. urban and wheat. However, in Winnipeg dataset, one of two minority classes was better classified by the proposed ensembles, and another was better classified by the classical ensembles (Figure 10). In Figures 8–10, blue and red circles indicate the successful and unsuccessful performances of the proposed ensembles, respectively, *versus* the classical ensembles in the classification of minority classes. Moreover, for certain classes, where PA or UA of the classical ensembles was lower than that of the individual DTs, the proposed ensembles could obtain higher values than the individual DT. This issue could be clearly seen in classifying of water in Flevoland and corn and forage in Winnipeg datasets. Accordingly, we can claim that two proposed tree ensembles, i.e. BFF and CFF, are more efficient and more reliable than two classical tree ensembles, i.e. BT and RF, for classification of imbalanced PolSAR data, especially in detecting minority classes.

Regarding the training time, both BFF and BT performed 4–10 times faster than RF. The CFF was also two times faster than RF, but 2–3 times was slower than BT and BFF. The higher speed of BFF compared to CFF may be due to the use of cost-sensitive learning in the CFF training step, rather than the equal-cost learning in BFF. Cost-sensitive learning normally takes a longer time to train the CFF. In Figure 7, we can also clearly observe that the computational time for training the RF model was increasing much quicker than the two proposed ensembles. Whereas, the time plots of both BFF and CFF had a much lower slope at almost all the nTrees. There may be two reasons for this observation, the first being the use of a filter feature selection algorithm in BFF and CFF ensembles rather than a wrapper algorithm employed in the RF. Wrapper algorithms are usually much slower than the filter algorithms, due to the use of the classifier for evaluating the selected subsets. As is previously said, RSFS employed in the RF uses *k*-NN classifier which is a time-consuming algorithm. Second, the pruning process used in two proposed ensembles can accelerate their training phase.

4.3. *Impact of training set size*

In another effort, we tried to study and compare the impact of different sizes of training samples on the classification by tree ensembles and the individual tree. For this purpose, we considered 10%, 20%, 40%, 60%, and 80% of all samples for the training data. Figure 11 presents the experimental results of Flevoland dataset. For all the ensembles, nTrees was considered equal to 5 in all sizes.

Referring to standard division of the methods, Figure 11 again confirmed higher stability of two proposed tree ensembles as compared to classical ensembles (4.3 and 5.2 vs. 6.1 and 7.5). In particular, the superiority of BFF compared to both BT and RF for various training set sizes was obvious. Among the tree ensembles, BT was the most unstable algorithm having the highest difference between the minimum and maximum of OA (i.e. around 18%). Furthermore, two proposed ensembles had higher capability (around 8% higher in OA) than two classical ensembles, especially BT, in classifying the PolSAR data when there was the smaller number of training samples. This issue was obvious in Figure 11 for the results of 10% and 20% (red circles).

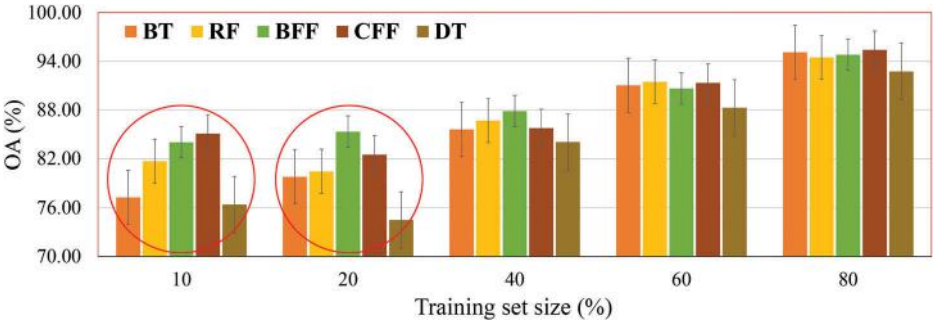

Figure 11. Impact of different training set sizes on classification accuracy by the tree ensembles (nTrees = 5) and the individual DT (experiment on Flevoland).

In detail, when 10% of sample data were considered as training samples, the OA values of the tree ensembles were nearly 9% higher than the OA of the individual DT. For 20%, the difference varied from 6% to 11%. The larger the training samples size, the higher the OA obtained by the individual DT. For example, the OA of the individual DT was nearly same to the OA of the tree ensembles for 80%. Therefore, comparing the results of small sizes (10% and 20%) and those of larger sizes (60% and 80%), confirms the higher efficiency and capability of the ensembles systems compared to individual classifiers in classification of high-dimensional data when confronting with the limited number of training samples.

4.4. *OMV and WMV* versus *MV*

We have presented two ambiguities in the MV fusion procedure. Therefore, another experiment was undertaken to compare MV and OMV procedures at the decision fusion step in the BFF algorithm at all the nTrees in Winnipeg dataset. The OA values obtained by two procedures are presented in Table 5. The mean OA over the nTrees for the two procedures is reported in the last row.

According to this table, the superiority of OMV over MV could be clearly seen at all the nTrees. Generally, OMV has obtained OA values 0.1–3% higher than MV. Its average was also 1% higher than MV. The superior performance of OMV was more obvious in the smaller nTrees (e.g. 5, 10, and 20). The larger the nTrees used, the smaller the difference between the MV and OMV. As result, OMV seems to be able to fuse the results more objectively than the MV method.

In addition to comparing MV and OMV, Table 6 presents the comparison of WMV and MV at the decision fusion step in the CFF algorithm, at all the nTrees in Foulum dataset. From this table, similar to OMV, the WMV was more successful than MV at all

Table 5. OA (%) of BFF with MV and BFF with OMV (experiment on Winnipeg).

nTrees	5	10	20	30	40	50	60	70	80	90	100	200	Mean
MV	84.48	81.08	85.38	83.61	84.56	85.45	85.46	83.35	83.45	84.51	84.33	83.83	84.16
OMV	86.38	84.11	88.12	84.53	85.39	85.79	85.80	83.63	83.84	84.70	84.53	83.95	85.06
	+1.90	**+3.03**	**+2.74**	+0.92	+0.83	+0.34	+0.34	+0.28	+0.39	+0.19	+0.20	+0.12	+0.90

The highest difference values are set in bold.

Table 6. OA (%) of CFF with MV and CFF with WMV (experiment on Foulum).

nTrees	5	10	20	30	40	50	60	70	80	90	100	200	Mean
MV	90.12	91.36	91.23	92.07	90.89	91.14	91.36	90.89	91.57	91.96	91.78	91.13	91.29
WMV	91.95	93.14	92.94	93.08	92.76	92.28	92.40	92.40	92.34	92.57	92.41	91.95	92.52
	+1.83	**+1.78**	**+1.71**	**+1.01**	**+1.87**	**+1.14**	**+1.04**	**+1.51**	+0.77	+0.61	+0.63	+0.82	+1.23

The highest difference values are set in bold.

the nTrees with 0.6–1.8% higher in OA values, and 1.2% higher in mean OA value. In the same way as OMV, the WMV had the smaller difference with the MV at nTrees >80. However, in contrast to OMV, the WMV has obtained higher OA than MV at most of nTrees <80. It was greater than 1%. Consequently, we concluded that WMV is more competent than MV, when classes have different number of samples or imbalanced distribution.

5. Conclusion

This paper addressed two problems of curse of dimensionality and imbalanced data as they applied specifically to land-cover classification using PolSAR data. The curse of dimensionality occurs when the ratio of training sample size to the number of input features is low. While the imbalanced data problem occurs when the number of samples for one or several land-cover classes is much lower than the others.

To solve these two problems, the classical ensemble systems, specifically BT and random forest (RF), as the-state-of-the-art approaches in machine learning theory, were used. Two modified ensembles were then proposed based on the idea of RF algorithm. These proposed ensembles, namely BFF and cost-sensitive filter-based forest (CFF), employed the concepts of filter feature selection, imbalanced data solutions, and extended majority voting algorithms in their structures.

The proposed ensembles had much better efficiency and reliability compared to BT, while BFF had approximately the same run time. In addition, these ensembles were generally more successful and much faster than the classic RF. Unlike RF and BT, which often gave impressive results with large ensembles, both BFF and CFF performed well with smaller ensembles. BFF and CFF also had fewer parameters to be tuned (i.e. the number of trees) and they were less sensitive to this parameter than RF and BT. This result was likely because they obtained more stable results than two others in all the nTrees. Thus, the proposed ensembles resulted in better classifications especially for minority classes in imbalanced data, and were faster. Therefore, we can suggest these proposed ensembles rather than BT and RF for classifying of multi-temporal, multi-source, high-dimensional and imbalanced data thanks to achieving higher accuracy with fewer and sparser trees

Determining the optimum weight of each class for CFF is an important key, which requires more studies. Additional research is needed as well, to evaluate the use of OMV in fusing in the context of different classifiers.

Acknowledgements

The authors would like to thank NASA Jet Propulsion Laboratory for sharing UAVSAR of Winnipeg, Manitoba, Canada and Agriculture and Agri-Food Canada for the reference map of this area, as well as, the European Space Agency for sharing the EMISAR and AIRSAR data and the corresponding reference data. This research was partially supported by the Iranian National Science Foundation (INSF).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Ainsworth, T. L., J. P. Kelly, and J. S. Lee. 2009. "Classification Comparisons between Dual-Pol, Compact Polarimetric and Quad-Pol SAR Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (5): 464–471. doi:10.1016/j.isprsjprs.2008.12.008.

Alberga, V. 2007. "A Study of Land Cover Classification Using Polarimetric SAR Parameters." *International Journal of Remote Sensing* 28 (17): 3851–3870. doi:10.1080/01431160601075541.

Alberga, V., G. Satalino, and D. K. Staykova. 2008. "Comparison of Polarimetric SAR Observables in Terms of Classification Performance." *International Journal of Remote Sensing* 29 (14): 4129–4150. doi:10.1080/01431160701840182.

Atwood, D. K., D. Small, and R. Gens. 2012. "Improving PolSAR Land Cover Classification with Radiometric Correction of the Coherency Matrix." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5: 848–856. Accessed 20 June 2016. http://earth.eo.esa.int/PolSARpro/datasets.html. Accessed April 2014. http://uavsar.jpl.nasa.gov/cgi-bin/data.pl

Bai, Y., D. Peng, X. Yang, L. Chen, and W. Yang. 2014. "Supervised Feature Selection for Polarimetric SAR Classification." 2014 12th International Congress on Signal Processing (ICSP), IEEE, Hangzhou, China. 1006–1010.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Bruzzone, L., M. Marconcini, U. Wegmüller, and A. Wiesmann. 2004. "An Advanced System for the Automatic Classification of Multitemporal SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 42: 1321–1334. doi:10.1109/TGRS.2004.826821.

Cloude, S. R., and E. Pottier. 1996. "A Review of Target Decomposition Theorems in Radar Polarimetry." *IEEE Transactions on Geoscience and Remote Sensing* 34 (2): 498–518. doi:10.1109/36.485127.

Cloude, S. R., and E. Pottier. 1997. "An Entropy Based Classification Scheme for Land Applications of Polarimetric SAR." *IEEE Transactions on Geoscience and Remote Sensing* 35 (1): 68–78. doi:10.1109/36.551935.

Congalton, R. G. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37 (1): 35–46. doi:10.1016/0034-4257(91)90048-B.

Dargahi, A., Y. Maghsoudi, and A. A. Abkar. 2013. "Supervised Classification of Polarimetric SAR Imagery Using Temporal and Contextual Information." *Remote Sensing Spatial Information Science*, Tehran, Iran, October 5–8, XL-1/W3:107–110.

Deschamps, B., H. McNairn, J. Shang, and X. Jiao. 2012. "Towards Operational Radar-Only Crop Type Classification: Comparison of a Traditional Decision Tree with a Random Forest Classifier." *Canadian Journal of Remote Sensing* 38 (1): 60–68. doi:10.5589/m12-012.

Du, P., A. Samat, B. Waske, S. Liu, and Z. Li. 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.

Du, P., A. Samat, P. Gamba, and X. Xie. 2014. "Polarimetric SAR Image Classification by Boosted Multiple-Kernel Extreme Learning Machines with Polarimetric and Spatial Features." *International Journal of Remote Sensing* 35 (23): 7978–7990. doi:10.1080/2150704X.2014.978952.

Du, P., J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu. 2012. "Multiple Classifier System for Remote Sensing Image Classification: A Review." *Sensors* 12 (4): 4764–4792. doi:10.3390/s120404764.

Ferro-Famil, L., E. Pottier, and J. S. Lee. 2001. "Unsupervised Classification of Multifrequency and Fully Polarimetric SAR Images Based on H/A/Alpha-Wishart Classifier." *IEEE Transactions on Geoscience and Remote Sensing* 39 (11): 2332–2342. doi:10.1109/36.964969.

Freeman, A., and S. L. Durden. 1998. "A Three-Component Scattering Model for Polarimetric SAR Data." *IEEE Transactions on Geoscience and Remote Sensing* 36 (3): 963–973. doi:10.1109/36.673687.

Fukuda, S., and H. Hirosawa. 2001. "Support Vector Machine Classification of Land Cover: Application to Polarimetric SAR Data." *Geoscience Remote Sensing Symposium 2001. IGARSS '01. IEEE 2001 International* 1: 187–189.

Gao, L., and Y. Ban. 2008. "Investigating the Performance of SAR Polarimetric Features in Land-Cover Classification." In *International Society for Photogrammetry and Remote Sensing*, Part B6b. Beijing, Vol. XXXVII: 3–11.

Habibi, M., M. R. Sahebi, Y. Maghsoudi, and S. Ghayourmanesh. 2016. "Classification of Polarimetric SAR Data Based on Object-Based Multiple Classifiers for Urban Land-Cover." *Journal of the Indian Society of Remote Sensing* 44 (6): 855–863. doi:10.1007/978-0-387-84858-7.

Haddadi, G. A., M. Reza Sahebi, and A. Mansourian. 2011. "Polarimetric SAR Feature Selection Using a Genetic Algorithm." *Canadian Journal of Remote Sensing* 37 (1): 27–36. doi:10.5589/m11-013.

Hara, Y., R. G. Atkins, S. H. Yueh, R. T. Shin, and J. A. Kong. 1994. "Application of Neural Networks to Radar Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 32 (1): 100–109. doi:10.1109/36.285193.

Hastie, T. J., R. J. Tibshirani, and J. H. Friedman. 2009. *The Elements Of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. doi:10.1007/978-0-387-84858-7.

He, H., and E. A. Garcia. 2009. "Learning from Imbalanced Data." *IEEE Transactions on Knoweledge and Data Engineering* 21 (9): 1263–1284. doi:10.1109/TKDE.2008.239.

Jafari, M., Y. Maghsoudi, and M. J. Valadan Zoej. 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. doi:10.1109/JSTARS.2014.2387374.

Khoshelham, K., C. Nardinocchi, E. Frontoni, A. Mancini, and P. Zingaretti. 2010. "Performance Evaluation of Automated Approaches to Building Detection in Multi-Source Aerial Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1): 123–133. doi:10.1016/j.isprsjprs.2009.09.005.

Khosravi, I., and M. Mohammad-Beigi. 2014. "Multiple Classifier Systems for Hyperspectral Remote Sensing Data Classification." *Journal of the Indian Society of Remote Sensing* 42 (2): 423–428. doi:10.1007/s12524-013-0327-7.

Krogager, E. 1990. "New Decomposition of the Radar Target Scattering Matrix." *Electronics Letters* 26 (18): 1525–1527. doi:10.1049/el:19900979.

Lardeux, C., P. L. Frison, C. Tison, J. C. Souyris, B. Stoll, B. Fruneau, and J. P. Rudant. 2009. "Support Vector Machine for Multifrequency SAR Polarimetric Data Classification." *IEEE Transactions on Geoscience and Remote Sensing* 47 (12): 4143–4152. doi:10.1109/TGRS.2009.2023908.

Lee, J. S., and E. Pottier. 2009. *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL: CRC press. doi:10.1201/9781420054989.

Lee, J. S., M. R. Grunes, and G. De Grandi. 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., M. R. Grunes, and R. Kwok. 1994. "Classification of Multi-Look Polarimetric SAR Imagery Based on Complex Wishart Distribution." *International Journal of Remote Sensing* 15 (11): 2299–2311.

Longadge, R., and S. Dongre. 2013. "Class Imbalance Problem in Data Mining Review." *International Journal of Computer Science and Network* 2 (1): 83–87.

Lonnqvist, A., Y. Rauste, M. Molinier, and T. Hame. 2010. "Polarimetric SAR Data in Land Cover Mapping in Boreal Zone." *IEEE Transactions on Geoscience and Remote Sensing* 48 (10): 3652–3662. doi:10.1109/TGRS.2010.2048115.

Ma, X., H. Shen, J. Yang, L. Zhang, and P. Li. 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. doi:10.1109/JSTARS.2013.2265331.

Maghsoudi, Y., M. Collins, and D. G. Leckie. 2012. "Polarimetric Classification of Boreal Forest Using Nonparametric Feature Selection and Multiple Classifiers." *Internationa Journal of Applied Earth Observation and Geoinformation* 19: 139–150. doi:10.1016/j.jag.2012.04.015.

Mangai, U. G., S. Samanta, S. Das, and P. R. Chowdhury. 2010. "A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification." *IETE Technical Review* 27 (4): 293–307. doi:10.4103/0256-4602.64604.

McNairn, H., A. Kross, D. Lapen, R. Caves, and J. Shang. 2014. "Early Season Monitoring of Corn and Soybeans with TerraSAR-X and RADARSAT-2." *Internationa Journal of Applied Earth Observation and Geoinformation* 28: 252–259. doi:10.1016/j.jag.2013.12.015.

Mishra, P., D. Singh, and Y. Yamaguchi. 2011. "Land Cover Classification of PALSAR Images by Knowledge Based Decision Tree Classifier and Supervised Classifiers Based on SAR Observables." *Progress in Electromagnetics Research* 30: 47–70. doi:10.2528/PIERB11011405.

Moriyama, T., S. Uratsuka, T. Umehara, M. Satake, A. Nadai, H. Maeno, and Y. Yamaguchi. 2004. "A Study on Extraction of Urban Areas from Polarimetric Synthetic Aperture Radar Image." In *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International*, Anchorage, AK. Vol. 1: 703–706. doi:10.1109/IGARSS.2004.1369127.

Nghiem, S. V., S. H. Yueh, R. Kwok, and F. K. Li. 1992. "Symmetry Properties in Polarimetric Remote Sensing." *Radio Science* 27 (5): 693–711. doi:10.1029/92RS01230.

Niu, X., and Y. Ban. 2014. "A Novel Contextual Classification Algorithm for Multitemporal Polarimetric SAR Data." *IEEE Geoscience and Remote Sensing Letters* 11 (3): 681–685. doi:10.1109/LGRS.2013.2274815.

Pohjalainen, J., O. Räsänen, and S. Kadioglu. 2015. "Feature Selection Methods and Their Combinations in High-Dimensional Classification of Speaker Likability, Intelligibility and Personality Traits." *Computer Speech Language* 29 (1): 145–171. doi:10.1016/j.csl.2013.11.004.

Qi, Z., A. G.-O. Yeh, X. Li, and X. Zhang. 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. doi:10.1016/j.isprsjprs.2015.02.004.

Qi, Z., A. G.-O. Yeh, X. Li, and Z. Lin. 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.

Räsänen, O., and J. Pohjalainen. 2013. "Random Subset Feature Selection in Automatic Recognition of Developmental Disorders, Affective States, and Level of Conflict from Speech." In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, Lyon, France. 210–214.

Salehi, M., M. R. Sahebi, and Y. Maghsoudi. 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.

Sánchez-Lladó, F. J., G. Pajares, and C. López-Martínez. 2011. "Improving the Wishart Synthetic Aperture Radar Image Classifications through Deterministic Simulated Annealing." *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (6): 845–857. doi:10.1016/j.isprsjprs.2011.09.007.

Senliol, B., G. Gulgezen, L. Yu, and Z. Cataltepe. 2008. "Fast Correlation Based Filter (FCBF) with a Different Search Strategy." International Symposium on Computer and Information Sciences, Istanbul, Turkey, October 27–29. 27–29. doi:10.1109/iscis.2008.4717949

Shah Hosseini, R., I. Entezari, S. Homayouni, M. Motagh, and B. Mansouri. 2011. "Classification of Polarimetric SAR Images Using Support Vector Machines." *Canadian Journal of Remote Sensing* 37 (2): 220–233. doi:10.5589/m11-029.

Shang, J., H. McNairn, B. Deschamps, and X. Jiao. 2011. "In-Season Crop Inventory Using Multi-Angle and Multi-Pass RADARTSAT-2 SAR Data over the Canadian Prairies." In *SPIE Optical Engineering Applications*, San Diego, CA. Vol. 8156, 815604–815604. doi:10.1117/12.894211

Shokrollahi, M., and H. Ebadi. 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: 1017–1024. doi:10.1007/s12524-016-0559-4.

Song, C. J., S. Yun, and L. Hui. 2004. "Classification of Polarimetric SAR Imagery Based on Target Decomposition and Neural Network Classifier." In *Proceeding of ACRS '04*, Chiang Mai, Thailand, November 2004.

Souissi, B., A. P. Doulgeris, and T. Eltoft. 2007. "Classification Comparisons between Compact Polarimetric and Quad-Pol SAR Imagery." *POLINSAR*, April 2015. Vol. 729.

Souissi, B., M. Ouarzeddine, and A. Belhadj-Aissa. 2014. "Optimal SVM Classification for Compact Polarimetric Data Using Stokes Parameters." *Journal Mathematical Model Algorithms Operational Researcher* 13: 433–446. doi:10.1007/s10852-013-9244-6.

Tamiminia, H., S. Homayouni, H. McNairn, and A. Safari. 2017. "A Particle Swarm Optimized Kernel-Based Clustering Method for Crop Mapping from Multi-Temporal Polarimetric L-Band SAR Observations." *International Journal of Applied Earth Observation and Geoinformation* 58: 201–212. doi:10.1016/j.jag.2017.02.010.

Uhlmann, S., and S. Kiranyaz. 2014. "Integrating Color Features in Polarimetric SAR Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 52: 2197–2206. doi:10.1109/TGRS.2013.2258675.

Waske, B., and M. Braun. 2009. "Classifier Ensembles for Land Cover Mapping Using Multitemporal SAR Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (5): 450–457. doi:10.1016/j.isprsjprs.2009.01.003.

Waske, B., and S. Van Der Linden. 2008. "Classifying Multilevel Imagery from SAR and Optical Sensors by Decision Fusion." *IEEE Transactions on Geoscience and Remote Sensing* 46 (5): 1457–1466. doi:10.1109/TGRS.2008.916089.

Xiangwei, X., K. Ji, H. Zou, and J. Sun. 2013. "Feature Selection and Weighted SVM Classifier-Based Ship Detection in PolSAR Imagery." *International Journal of Remote Sensing* 34 (22): 7925–7944. doi:10.1080/01431161.2013.827812.

Xu, L., J. Li, and A. Brenning. 2014. "A Comparative Study of Different Classification Techniques for Marine Oil Spill Identification Using RADARSAT-1 Imagery." *Remote Sensing of Environment* 141 (2014): 14–23. doi:10.1016/j.rse.2013.10.012.

Yamaguchi, Y., T. Moriyama, M. Ishido, and H. Yamada. 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.

Yu, L., and H. Liu. 2003. "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution." In *Proceedings of the twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, Vol. 3, 856–863.

Zeng, Z., H. Zhang, R. Zhang, and C. Yin. 2015. "A Novel Feature Selection Method Considering Feature Interaction." *Pattern Recognition* 48 (8): 2656–2666. doi:10.1016/j.patcog.2015.02.025.