POLSAR DATA ONLINE CLASSIFICATION BASED ON MULTI-VIEW LEARNING

Xiangli Nie[a] *Shuguang Ding*[b] *Bo Zhang* [b] *Hong Qiao*[a,c] *Xiayuan Huang*[a,⋆]

[a]The State Key Lab of Management and Control for Complex System,Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[b] Institute of Applied Mathematics, AMSS, Chinese Academy of Sciences, Beijing, 100190, China
[c] CAS Centre for Excellence in Brain Science and Intelligence Technology, Shanghai 200031, China

ABSTRACT

Polarimetric synthetic aperture radar (PolSAR) plays an indispensable part in remote sensing. With its development and application, rapid and accurate online classification for PolSAR data becomes more and more important. PolSAR data can be depicted by different features such as polarimetric, texture and color features, which can be considered as multiple views. In this paper, we propose an online multi-view learning method based on the passive aggressive algorithm, named OMVPA, for PolSAR data real-time classification. The OMVPA method makes full use of the consistency and complementary properties of different views. Experimental results on real PolSAR data demonstrate that the proposed method maintain a smaller mistake rate compared with other methods.

***Index Terms*—** Multi-view learning, online classification, passive-aggressive (PA) algorithm, polarimetric synthetic aperture radar (PolSAR).

1. INTRODUCTION

Polarimetric synthetic aperture radar (PolSAR) is a new form of radar system and has wide applications in agroforestry, geology, oceanography and military fields as an advanced instrument for remote sensing [1]. With the development and application of PolSAR systems, rapid and accurate online classification for PolSAR data becomes more and more important.

For PolSAR online classification problems, samples arrive in a sequence and the number of samples grows unboundedly as the time increases. Thus, the conventional batch algorithms, which retrain a classifier on the whole data set when a new sample arrives, can not handle these learning tasks efficiently. To improve the computational efficiency and solve the storage problems, many online learning algorithms [2–6] were proposed for its scalability and efficiency dealing with the sequential data. Different from the batch algorithms, online algorithms have several characteristics: 1) only one new sample is available in each trial; 2) the current predictor should be updated immediately if the new sample is misclassified; 3) the algorithms can be processed efficiently without reusing all the visible samples. Among these algorithms, the online Passive-Aggressive (PA) algorithm is commonly used for its reasonable performance and low computational cost [7]. In this paper, we focus on the PA algorithms.

PolSAR data classification such as built-up areas extraction and the land use/land cover classification, has attracted a lot of attentions. In the past two decades, many methods have been proposed for PolSAR data classification, and they can be categorized according to the features such as the statistical, scattering and texture information [8]. The statistical characteristics-based methods include the supervised Wishart classifier [9] and unsupervised H/A/alpha-Wishart classifier [10]. The scattering mechanisms classification approaches include the Freeman-Durden decomposition-based method [11]. The texture features-based methods include the classification approach using the decision tree and multiscale texture [12] and the evolutionary radial basis function network based classifier [13]. The color features extracted from the Pauli decomposition images are introduced for PolSAR classification in [14] and experimental results show that the additional color features can improve the classification performance. Recently, the deep convolutional neural networks based method was proposed in [15], which can automatically extract translational-invariant spatial features.

However, these classification methods generate the classifier on the entire training data set, and the corresponding classifiers would not be updated when a newly arrived sample is misclassified or they have to be retained on the whole data set. Besides, most of the above approaches just use one kind of feature or some of them may consider several feature information but concatenate the features from different views into a single view. This concatenation do not take full advantage of the relationships among the multiple views which are

This work was partly supported by the National Natural Science Foundation of China under Grants 61379093, 61602483 and 91648205, by the Beijing Natural Science Foundation under Grant 4174107 and by the Early Career Development Award of SKLMCCS. ⋆ Corresponding author. Email address: (xiangli.nie and xiayuan.huang)@ia.ac.cn.

obtained from different feature subsets.

To address the above two issues, in this paper, we propose an online multi-view learning algorithm based on the PA algorithm for PolSAR data real-time classification, named OMVPA. PolSAR data can be characterized by multiple features such as the polarimetric, color and texture features, which can be considered as different views. In order to explore the consistency and complementary properties of these views, we give a new model that ensemble the classifiers of different views and enforces the agreement among different predictors based on the coregularization algorithm. Experimental results on real PolSAR data have shown the proposed online multi-view classification algorithm is more effective than single-view learning algorithm.

The rest of this paper is organized as follows. Section 2 presents the related methods and the proposed online multi-view learning model and algorithm. Experimental results on real PolSAR data are shown in Section 3. Some concluding remarks are given in Section 4.

2. THE PROPOSED ONLINE MULTI-VIEW LEARNING ALGORITHM

In the following sections, lowercase letters (e.g., λ_1, λ_2, d, r) and bold lowercase letters (e.g., $\boldsymbol{w}$, $\mathbf{x}$) refer to scalars and vectors, respectively. $(\boldsymbol{w} \cdot \mathbf{x})$ denotes the inner product between $\boldsymbol{w}$ and $\mathbf{x}$, and $\|\mathbf{x}\|$ refers to the L_2 norm of $\mathbf{x}$.

2.1. Related Methods

The online Passive-Aggressive (PA) learning algorithm [4] is set to minimize the distance between the learnt classifier and the previous classifier and the loss of the learnt classifier suffered on the current instance simultaneously. Three optimization problems with one using hard margin and two based on soft margins are formulated. The latter two can better handle the non-separable and noisy case. However, the family of online PA algorithm only works on single view datasets. Recently, a two-view online learning algorithm [7] was proposed, which minimizes the changes in the two view weights and disagreements between the two classifiers based on different views. In this algorithm, the prediction function $f(x_t^{(1)}, x_t^{(2)}) = \mathrm{sign}(\frac{1}{2}\boldsymbol{w}_t^{(1)} \cdot \mathbf{x}_t^{(1)} + \frac{1}{2}\boldsymbol{w}_t^{(2)} \cdot \mathbf{x}_t^{(2)})$, which is the equally weighted sum of $f(x_t^{(1)})$ and $f(x_t^{(2)})$. This setting is not suitable for the situation that the two views are not equal important for online learning. In [16], an adaptive two-view online learning algorithm was given and a weighting parameter η is used to adjust the importance of the two views. However, this weight is also used to trade off the disagreement between the two views and the corresponding term is defined as $|\eta \boldsymbol{w}_t^{(1)} \cdot \mathbf{x}_t^{(1)} - (1-\eta)\boldsymbol{w}_t^{(2)} \cdot \mathbf{x}_t^{(2)})|$. This disagreement factor is not reasonable because the introduced weight may increase the disagreement of the two views. Besides, the coefficients

Algorithm 1: Online Multi-view Learning Algorithm for Binary Classification

Input: Four positive scalars λ_1, d, r and c.
Initialization: $\boldsymbol{w}^{(1)} = \mathrm{rand}(1, \mathrm{length}(\mathbf{x}_1^{(1)}))$, $\boldsymbol{w}^{(2)} = \mathrm{rand}(1, \mathrm{length}(\mathbf{x}_1^{(2)}))$.
for $t = 1, 2, ...$ **do**
  Receive instances: $\mathbf{x}_t^{(1)} \in \mathbb{R}^m$ and $\mathbf{x}_t^{(2)} \in \mathbb{R}^n$.
  Compute the sign: $s = \mathrm{sign}((\boldsymbol{w}_t^{(1)} \cdot \mathbf{x}_t^{(1)}) - (\boldsymbol{w}_t^{(2)} \cdot \mathbf{x}_t^{(2)}))$.
  Compute: $f_t = r(\boldsymbol{w}_t^{(1)} \cdot \mathbf{x}_t^{(1)}) + (1-r)(\boldsymbol{w}_t^{(2)} \cdot \mathbf{x}_t^{(2)})$.
  Predict: $\hat{y}_t = \mathrm{sign}(f_t)$.
  Receive the correct label: $y_t \in \{-1, 1\}$
  Compute the loss: $l_t = \max\{0, 1 - y_t f_t\}$.
  if $l_t > 0$ **then**
    Compute the parameter τ according to (6).
    Update $\boldsymbol{w}_{t+1}^{(1)}$ and $\boldsymbol{w}_{t+1}^{(2)}$ according to (3) and (4).
  end if
end for

before $\|\boldsymbol{w}^{(1)} - \boldsymbol{w}_t^{(1)}\|^2$ and $\|\boldsymbol{w}^{(2)} - \boldsymbol{w}_t^{(2)}\|^2$ are equal and there is no parameter to trade off the two terms, which may enforce the update of the two-view weight vectors to be close to each other. Its performance would be not good when there is a large difference between the two views. Motivated by the above problems, in the following subsection, we give a new online classification algorithm based on multi-view learning.

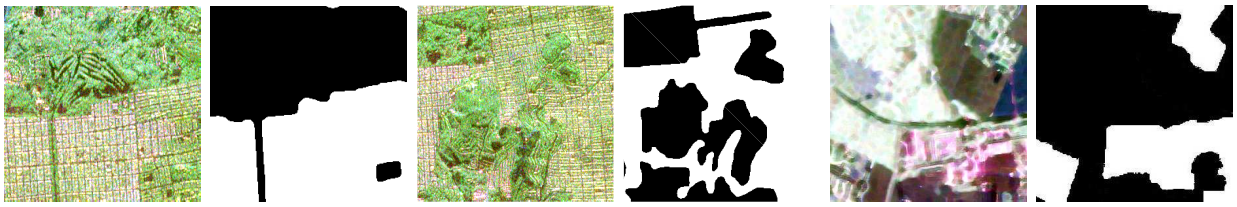
2.2. The Proposed Online Classification Algorithm

For simplicity, we focus on binary classification problem of two views in this paper. Consider the instance-label pair $(\mathbf{x}_t, y_t)$ on round t, where $\mathbf{x}_t = (\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)})$ is the two-view instance and $y_t \in \{+1, -1\}$ is the common label. The goal is to learn a linear prediction function $f = \mathrm{sign}(r(\boldsymbol{w}^{(1)} \cdot \mathbf{x}_t^{(1)}) + (1-r)(\boldsymbol{w}^{(2)} \cdot \mathbf{x}_t^{(2)}))$ where $r \in (0, 1)$ is a weight parameter to adjust the importance of different views.

The proposed constrained optimization problem for online multi-view learning based on PA algorithm, named OMVPA, can be written as follows:

$$\begin{aligned}\{\boldsymbol{w}_{t+1}^{(1)}, \boldsymbol{w}_{t+1}^{(2)}\} = \underset{\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}}{\mathrm{argmin}} & \frac{\lambda_1}{2}\|\boldsymbol{w}^{(1)} - \boldsymbol{w}_t^{(1)}\|^2 + \frac{\lambda_2}{2}\|\boldsymbol{w}^{(2)} - \boldsymbol{w}_t^{(2)}\|^2 \\ & + d|(\boldsymbol{w}^{(1)} \cdot \mathbf{x}_t^{(1)}) - (\boldsymbol{w}^{(2)} \cdot \mathbf{x}_t^{(2)})| + c\xi^2 \\ \text{s.t.} \quad & l(\boldsymbol{w}; (\mathbf{x}_t, y_t)) \le \xi. \end{aligned} \tag{1}$$

where λ_1 and λ_2 are trade-off parameters in the two views, d is a coupling parameter and c is a positive parameter. In this objective function, the first two terms require $\boldsymbol{w}_{t+1}$ must stay as close as possible to $\boldsymbol{w}_t$. The third term enforces the agreement among predictors on a new coming sample. The last term can scale the objective function quadratically. Here,



(a) (b) (c) (d) (e) (f)

Fig. 1: Real PolSAR data. (a) The Pauli decomposition image of sanf1. (b) The ground truth map of sanf1. (c) The Pauli decomposition image of sanf2. (d) The ground truth map of sanf2. (e) The Pauli decomposition image of oberp. (f) The ground truth map of oberp.

Table 1: AVERAGE ONLINE MISTAKE RATES ON REAL POLSAR DATA

Dataset/Methods	PA_Pol	PA_coltex	PA_Cat	AdaPA	OMVPA
sanf1	0.0782±0.0006	0.0875±0.0008	0.0551±0.0006	0.0617±0.0008	**0.0461±0.0005**
sanf2	0.2167±0.0010	0.2291±0.0013	0.2032±0.0011	0.1723±0.0004	**0.1536±0.0006**
oberp	0.2861±0.0007	0.2998±0.0014	0.2750±0.0007	0.2682±0.0009	**0.2163±0.0011**

the loss is defined by the hinge-loss function $l(\boldsymbol{w};(\mathbf{x}_t,y_t)) = \max\{0, 1-y_t(r(\boldsymbol{w}^{(1)}\cdot\mathbf{x}_t^{(1)})+(1-r)(\boldsymbol{w}^{(2)}\cdot\mathbf{x}_t^{(2)}))\}$.

The Lagrangian function of the optimization problem (1) can be written as follows:

$$L(\boldsymbol{w},\xi,\tau) = \frac{\lambda_1}{2}\|\boldsymbol{w}^{(1)}-\boldsymbol{w}_t^{(1)}\|^2 + \frac{\lambda_2}{2}\|\boldsymbol{w}^{(2)}-\boldsymbol{w}_t^{(2)}\|^2 + d|(\boldsymbol{w}^{(1)}\cdot\mathbf{x}_t^{(1)})-(\boldsymbol{w}^{(2)}\cdot\mathbf{x}_t^{(2)})| + c\xi^2 + \tau[1-y_t(r(\boldsymbol{w}^{(1)}\cdot\mathbf{x}_t^{(1)})+(1-r)(\boldsymbol{w}^{(2)}\cdot\mathbf{x}_t^{(2)}))-\xi] \quad (2)$$

where $\tau \geq 0$ is a Lagrangian multiplier. Letting $s := \text{sign}((\boldsymbol{w}^{(1)}\cdot\mathbf{x}_t^{(1)})-(\boldsymbol{w}^{(2)}\cdot\mathbf{x}_t^{(2)}))$ and setting the partial derivatives of L with respect to $\boldsymbol{w}^{(1)}$ to zero gives,

$$\frac{\partial L}{\partial \boldsymbol{w}^{(1)}} = \lambda_1(\boldsymbol{w}^{(1)}-\boldsymbol{w}_t^{(1)}) + (ds-\tau r y_t)\mathbf{x}_t^{(1)} = 0.$$

Then we have

$$\boldsymbol{w}_{t+1}^{(1)} = \boldsymbol{w}_t^{(1)} + \frac{1}{\lambda_1}(\tau r y_t - ds)\mathbf{x}_t^{(1)} \quad (3)$$

Similarly, for the other view we get

$$\boldsymbol{w}_{t+1}^{(2)} = \boldsymbol{w}_t^{(2)} + \frac{1}{\lambda_2}(\tau(1-r)y_t + ds)\mathbf{x}_t^{(2)} \quad (4)$$

Taking the partial derivatives of L with respect to ξ and setting it to zero gives,

$$\xi = \frac{\tau}{2c}. \quad (5)$$

Plugging (3), (4) and (5) into (2), we have

$$L(\tau) = -\frac{1}{2\lambda_1}(\tau r y_t - ds)^2\|\mathbf{x}_t^{(1)}\|^2 + \tau r - (\tau r y_t - ds)(\boldsymbol{w}^{(1)}\cdot\mathbf{x}_t^{(1)}) - (\tau(1-r)y_t+ds)(\boldsymbol{w}^{(2)}\cdot\mathbf{x}_t^{(2)}) + \tau(1-r) - \frac{1}{2\lambda_2}(\tau(1-r)y_t+ds)^2\|\mathbf{x}_t^{(2)}\|^2 - \frac{\tau^2}{4c}$$

Setting the partial derivatives of L with respect to τ to zero gives,

$$\frac{\partial L}{\partial \tau} = 1 - r y_t(\boldsymbol{w}_t^{(1)}\cdot\mathbf{x}_t^{(1)}) - (1-r)y_t(\boldsymbol{w}_t^{(2)}\cdot\mathbf{x}_t^{(2)}) + ds y_t(r\|\mathbf{x}_t^{(1)}\|^2/\lambda_1 - (1-r)\|\mathbf{x}_t^{(2)}\|^2/\lambda_2) - (r^2\|\mathbf{x}_t^{(1)}\|^2/\lambda_1 + (1-r)^2\|\mathbf{x}_t^{(2)}\|^2/\lambda_2 + 1/2c)\tau$$

So we get

$$\tau = \frac{l(\boldsymbol{w}_t;(\mathbf{x}_t,y_t)) + ds y_t\left[\frac{r}{\lambda_1}\|\mathbf{x}_t^{(1)}\|^2 - \frac{1-r}{\lambda_2}\|\mathbf{x}_t^{(2)}\|^2\right]}{\frac{r^2}{\lambda_1}\|\mathbf{x}_t^{(1)}\|^2 + \frac{(1-r)^2}{\lambda_2}\|\mathbf{x}_t^{(2)}\|^2 + \frac{1}{2c}}. \quad (6)$$

The pseudo-code of the above algorithm is presented in Algorithm 1. The positive parameters λ_1, d, r and c can be selected by cross validation. The weight vectors $\boldsymbol{w}^{(1)}$ and $\boldsymbol{w}^{(2)}$ are initialized randomly from unitary uniform distributions.

3. EXPERIMENTAL RESULTS

In this section, we test the proposed method on three real PolSAR datasets, named sanf1, sanf2 and oberp, which are the subsets of the San Francisco Bay data and the Oberpfaffenhofen Area data, respectively. They are collected by AIRSAR and E-SAR, respectively, and can be obtained from the website: `http://earth.eo.esa.int/polsarpro/datasets.html`. Fig. 1 (a), (c) and (e) show the Pauli decomposition images of sanf1, sanf2 and oberp, respectively. The sizes of the three images are 200×200, 300×300, and 300×300, respectively. In Fig. 1 (b) and (d), the ground objects are comprise of grass and buildings which are labeled as black and white in the ground truth maps, respectively. In Fig. 1 (f), the suburban areas are labeled as white while

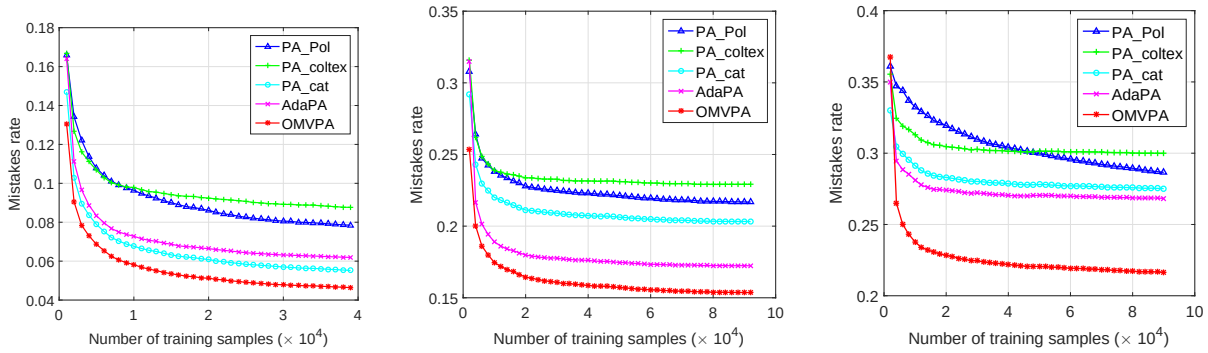


Fig. 2: Comparison of error evolution curves. (a) Results on sanf1. (b) Results on sanf2. (c) Results on oberp.

the other regions including the woodland and farmland are labeled as black, which can be regarded as the built-up areas extraction problem.

In our experiments, we consider the polarimetric feature and the color-texture feature of PolSAR data, which can be regarded as two views. The above features are computed in a sliding window with size 11×11 in order to reduce the effect of speckle noise [17]. The polarimetric features include 36 parameters, which consist of the elements from covariance and coherent matrices and the components of Cloude, Krogager, Freeman-Durden, Huynen and VanZyl decomposition [14]. The color-texture features include 44 parameters with 10 color parameters and 34 texture parameters. The color features are made up of the three color components of Pauli decomposition, the hue-saturation-value components and the weight of the most dominant color. The texture features include the quantized 10 bin LBP histogram and the first and second order moments of gabor wavelets with three scales and four orientations [14].

The performance of the proposed method is compared with the results of PA algorithm and that of the adaptive two-view PA algorithm denoted as AdaPA. The PA algorithm is carried out on the polarimetric feature, color-texture feature and their concatenation feature datasets, respectively. The corresponding results are denoted as PA_Pol, PA_coltex and PA_Cat, respectively. In order to get the best performance, the parameters in AdaPA and the proposed algorithm are chosen by cross validation. As for AdaPA, we set $d \in \{0.001, 0.01, 0.1\}$, $r \in (0, 1)$ and $c \in [0.1, 0.5]$. For OMVPA, we set $\lambda_1 \in [0.5, 1.5]$, $\lambda_2 = 1$, $d \in \{0.001, 0.01, 0.1\}$, $r \in (0, 1)$ and $c \in [0.01, 0.1]$.

Table 1 presents the average mistake rates of different methods on the three datasets. To better evaluate the performance, each experiment is repeated 10 times and the results are averaged over the 10 repetitions. For the PA algorithm, as expected, the average mistake rate on the concatenation feature dataset is lower than that on each view feature dataset. It can be seen that the AdaPA algorithm has a higher average error than PA_Cat on the sanf1 dataset, which may be caused by the unreasonable weighted disagreement constraint. The proposed OMVPA method has the lowest average mistake rate and achieves about 1%, 2% and 5% improvements compared with the results of PA_Cat and AdaPA, respectively.

Fig. 2 shows the evolution of mistake rates as a function of the number of online rounds. We observe that the error rate of OMVPA consistently decreases with the data stream size and the average mistake rates of almost all these methods tend to be stable when the number of samples reaches 3×10^4.

4. CONCLUSION

In this paper, an online multi-view learning method based on passive aggressive algorithm, called OMVPA, has been proposed for PolSAR data real-time classification. The polarimetric and color-texture features are considered as different views. The proposed OMVPA method ensembles the classifiers of different views and enforces the agreement among different predictors by exploring the consistency and complementary properties of these views. The optimization problem can be solved by Lagrange multiplier method and the update steps have analytical solutions. Experimental results on real PolSAR data demonstrate that the proposed method can achieve a lower error rate compared with other methods.

References

[1] J.S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, CRC press, 2009.

[2] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," 2003.

[3] J. Kivinen, A. J. Smola, and R. C. Williamson, "Online learning with kernels," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2165–2176, 2004.

[4] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, "Online passive-aggressive algorithms," *Journal of Machine Learning Research*, vol. 7, no. Mar, pp. 551–585, 2006.

[5] D. Wang, H. Qiao, B. Zhang, and M. Wang, "Online support vector machine based on convex hull vertices selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 4, pp. 593–609, 2013.

[6] S. G. Ding, X. L. Nie, H. Qiao, and B. Zhang, "A fast algorithm of convex hull vertices selection for online classification," *IEEE Transactions on Neural Networks and Learning Systems, Doi. 10.1109/TNNLS.2017.2648038*, 2016.

[7] T. Nguyen, K.Y. Chang, and S.C. Hui, "Two-view online learning," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2012, pp. 74–85.

[8] B. Hou, C. Chen, X.J. Liu, and L.C. Jiao, "Multilevel distribution coding model-based dictionary learning for PolSAR image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 11, pp. 5262–5280, 2015.

[9] J.S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery-based on complex Wishart distribution," *International Journal of Remote Sensing*, vol. 15, no. 11, pp. 2299–2311, 1994.

[10] L. Ferro-Famil, E. Pottier, and J.S. Lee, "Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 11, pp. 2332–2342, 2001.

[11] J.S. Lee, M. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 722–731, 2004.

[12] M. Simard, S. Saatchi, and G. De Grandi, "The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2310–2321, 2000.

[13] T. Ince, S. Kiranyaz, and M. Gabbouj, "Evolutionary RBF classifier for polarimetric SAR images," *Expert Systems with Applications*, vol. 39, no. 5, pp. 4710–4717, 2012.

[14] S. Uhlmann and S. Kiranyaz, "Integrating color features in polarimetric SAR image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 4, pp. 2197–2216, 2014.

[15] Y. Zhou, H.P. Wang, F. Xu, and Y.Q. Jin, "Polarimetric sar image classification using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1935–1939, 2016.

[16] T. Nguyen, K.Y. Chang, and S.C. Hui, "Adaptive two-view online learning for math topic classification," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2012, pp. 794–809.

[17] X.L. Nie, H. Qiao, and B. Zhang, "A variational model for PolSAR data speckle reduction based on the Wishart distribution," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1209–1222, 2015.