


Multiple-feature-driven co-training method for crop mapping based on remote sensing time series imagery

Duo Jia , Peichao Gao , Changxiu Cheng & Sijing Ye

To cite this article: Duo Jia , Peichao Gao , Changxiu Cheng & Sijing Ye (2020) Multiple-feature-driven co-training method for crop mapping based on remote sensing time series imagery, International Journal of Remote Sensing, 41:20, 8096-8120

To link to this article: <https://doi.org/10.1080/01431161.2020.1771790>



View supplementary material 



Published online: 15 Aug 2020.



Submit your article to this journal 



View related articles 



View Crossmark data 



Multiple-feature-driven co-training method for crop mapping based on remote sensing time series imagery

Duo Jia^{a,b,c,d}, Peichao Gao^{a,b,c,d}, Changxiu Cheng^{a,b,c,d,e} and Sijing Ye^{a,b,c,d}

^aState Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China; ^bKey Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Beijing, China; ^cFaculty of Geographical Science, Beijing Normal University, Beijing, China; ^dCenter for Geodata and Analysis, Beijing Normal University, Beijing, China; ^eNational Tibetan Plateau Data Center, Beijing, China

ABSTRACT

Remote sensing time series imagery (RSTSI) provides a useful tool for crop mapping, as it provides crucial spectral, temporal, and spatial (STS) features. However, its high dimensionality coupled with the limited number of training samples leads to an ill-posed classification problem and the Hughes phenomenon. To solve this problem, this study presents a multiple-feature-driven co-training method (MFDC) for accurately mapping crop types based on RSTSI with a limited number of training samples. In MFDC, four complementary pre-defined views, which represent STS features, are generated for the utilization of multiple features. Then, to enhance the classifier's generalization ability, a novel labelled sample augmentation method that combines the Breaking Tiles algorithm and co-training is proposed. Third, to ensure the effectiveness of ensemble learning in co-training as well as to further speed up the learning process, a multi-view semi-supervised feature learning algorithm that expands the single view semi-supervised learning algorithm to multiple views is proposed and embedded in co-training. Finally, a weighted majority vote method is utilized to obtain the classification results. The experimental results for study areas in the United States indicate that the proposed method can accurately map crop types with a limited number of labelled training samples without a significant computational cost.


ARTICLE HISTORY

Received 15 September 2019
Accepted 30 April 2020

1. Introduction

Remote sensing provides a useful tool for identification, monitoring, and mapping of land cover because of its capacity to provide consistent and repeatable measurements at an appropriate spatial scale (Verbesselt et al. 2010). However, it is quite difficult to accurately map complicated land cover categories with similar spectral characteristics at certain points in time (Senf et al. 2015). For example, some crop types show very small spectral differences at certain times of a year. However, crops usually have different seasonal variations; therefore, to improve the accuracy, it is vital to capture their phenological dynamics and combine them with fine spatial information (Senf et al. 2015; Prishchepov

CONTACT Changxiu Cheng  chengcx@bnu.edu.cn  State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China

 Supplemental data for this article can be accessed [here](#).

© 2020 Informa UK Limited, trading as Taylor & Francis Group

et al. 2012) as well as detailed spectral information (Griffiths, Nendel, and Hostert 2019). Remote sensing time series imagery (RSTSI) is an important means of obtaining spectral-temporal-spatial (STS) features. Compared to single-date images, which reflect land surfaces at certain points in time, RSTSI has been shown to have a high capacity for the characterization of environmental phenomena and vegetation dynamics, as well as discrete change events (Suess et al. 2018; Jönsson et al. 2018; De Alban et al. 2018; Arévalo, Olofsson, and Woodcock 2020; Henits, Jürgens, and Mucsi 2016; Yan et al. 2015).

An often-used strategy to map crop types is using RSTSI as the input of some state-of-the-art supervised classification algorithms, such as Random Forest (RF) (Estel et al. 2015; Rodriguez-Galiano et al. 2012; Clark et al. 2010; Yan and Roy 2015; Salehi, Daneshfar, and Davidson 2017), Support Vector Machine (SVM) (Zhu and Liu 2014; Hackman, Gong, and Wang 2017), and deep learning (Interdonato et al. 2019; Zhong, Lina, and Zhou 2019; Pelletier, Webb, and Petitjean 2019). To further improve the classification accuracy, the original spectral features (Forkuor et al. 2017; Hu et al. 2018; Kussul et al. 2017), the temporal features (Sexton et al. 2013; Jia et al. 2014; Müller et al. 2015), the statistics of the temporal spectral features (Forkuor et al. 2017; Yu and Shang 2017), and spatial features (Rodriguez-Galiano et al. 2012; Yan and Roy 2014) have been used as input to the classifiers. However, the training sample size and feature dimension are two factors that affect the generalization ability of the classifier. The ratio of training sample size to feature dimension determines the generalization ability of the classifier (Baraldi, Bruzzone, and Blonda 2005). It is generally considered that an appropriate training sample size should be about 10 times more than the number of feature dimensions (Hughes 1968). However, obtaining labelled training samples is time-consuming, difficult, and costly, while, when considering multiple features of RSTSI as input, the number of feature dimensions is high. This imbalance easily leads to ill-posed classification problems (Hughes 1968).

Semi-supervised learning has been demonstrated to be an effective way to solve this ill-posed classification problem. In this approach, the useful information in the unlabelled samples is mined to improve the generalization ability of the classifier with a limited number of training samples (Shahshahani and Landgrebe 1994). Self-training and co-training are two typical examples of semi-supervised learning methods. Self-training iteratively selects reliable samples with predicted labels to increase the training sample size without significant extra cost, and these additional samples are further optimized to reduce the cumulative error in the iteration process (Tan et al. 2015; Di and Crawford 2011). To select more useful unlabelled samples, Dópido et al. (2013) applied active learning to a self-training framework, in which informative unlabelled samples were selected first. Then, the constraint of the neighbourhood information was utilized to reduce the error rate in label prediction. To further improve the label prediction accuracy, Kim, Park, and Lee (2017) leveraged past land cover maps as supplementary information and designed pre-defined rules to predict the labels. However, the errors of past land cover maps add some uncertainties. Instead of incorporating supplementary information, Ma, Wang, and Wang (2016) constructed a framework of joint global and local decision-making for predicting selected unlabelled samples. Tan et al. (2014) utilized the Mean Shift segmentation method to first extract the spatial information, and used the latter to optimize the final classification result.

Co-training is another widely used semi-supervised learning algorithm (Blum and Mitchell 1998). Unlike self-training, which applies the learning process on a single view,

co-training considers unlabelled samples as an information interaction among different views during learning, each view represents a kind of information of unlabelled samples. First, a classifier is trained on each view using the initial labelled training samples. Then, each classifier selects unlabelled samples with reliably predicted labels and adds these samples to another classifiers' training samples. By iteratively increasing the number of training samples for each classifier, the classifiers' generalization ability can be improved. However, co-training relies on complementarity and independence of the multi-views (Nigam and Ghani 2000), which reduces its applicability. To relax this independence requirement, Zhi-Hua and Ming (2005) proposed a tri-training algorithm that divided the labelled samples into three parts (views) by bootstrap. An unlabelled sample for each view would be then labelled when the two other classifiers agreed on a label (Zhi-Hua and Ming 2005; Tan et al. 2016). Gu and Jin (2017) further expanded this approach to more views for performance enhancement. In the above-mentioned tri-training algorithms, the final classification results were generated by using ensemble learning. However, the effectiveness of the algorithms was not studied. To increase the differences among the views to ensure the effectiveness of ensemble learning, Tan et al. (2016) improved the traditional tri-training by selecting three out of four classifiers with considerable differences based on a diversity measurement.

Compared to self-training, which performs on a single view without exploiting information from other views, co-training's multi-view approach makes it more suitable for the STS features of RSTSI. Nonetheless, co-training is still limited in the following aspects: First, although the independence requirement of multi-views can be relaxed through the tri-training approach, its assumption that the accumulated labelling noise can be compensated with a large amount of unlabelled data makes a large number of unlabelled sample lack representativeness, meanwhile leading to an increase of training times. Second, the current strategy of introducing diverse measurements to ensure the effectiveness of ensemble learning in the co-training is time-consuming.

In light of the aforementioned problems, in this study a novel multiple-feature-driven co-training (MFDC) method is proposed for accurately mapping crop types using the STS features of RSTSI. The improvement of the MFDC lies in two strategies: First, to enhance the representativeness and accuracy of selected unlabelled samples, a labelled sample augmentation algorithm is proposed. Second, to ensure the effectiveness of ensemble learning as well as to speed up the learning process, a multi-view semi-supervised feature (MVSSF) learning approach is proposed, which enlarges the single view semi-supervised learning algorithm to multiple views. The main novelties of the proposed MFDC method include the following: (1) A co-training algorithm capable of making full use of the STS features of RSTSI is utilized. (2) MFDC can accurately map crop types using a limited number of training samples without significant computational cost. (3) To the best of our knowledge, this study introduces for the first time multi-view feature learning in a co-training algorithm so as to ensure the effectiveness of ensemble learning as well as to speed up the learning process.

2. Study area and data sets

2.1. Study area

For this study, we conducted experiments examining two agricultural areas in the USA (Figure 1). One site involved images from 2010 from Kansas State, in the Central Great

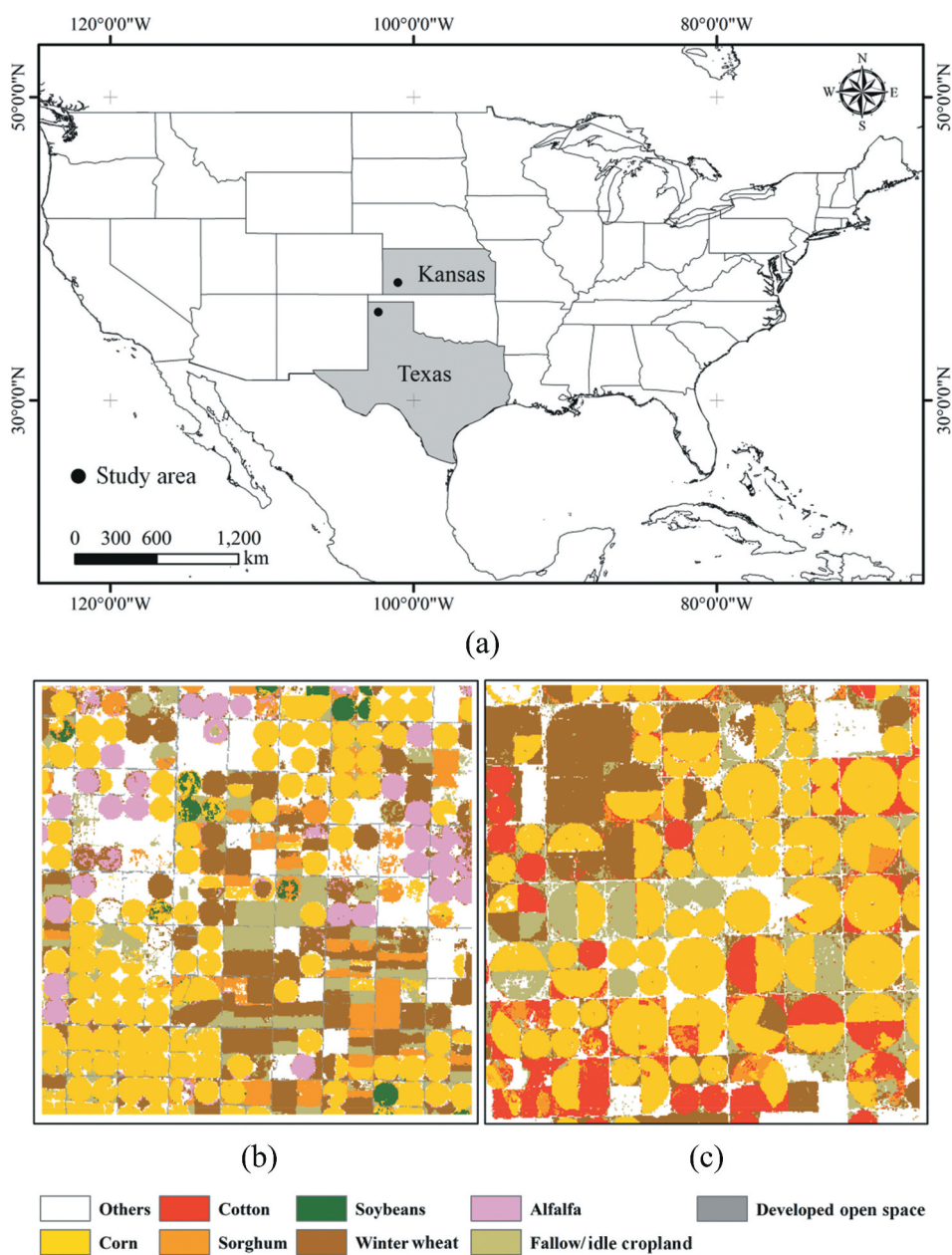


Figure 1. Location of two study areas and reference datasets. (a) Location of two study areas, (b) Reference datasets of study area 1, (c) Reference datasets of study area 2.

Plains area of the USA, covering an area of $13.5 \text{ km} \times 13.5 \text{ km}$. Kansas State is dominated by agriculture, covering 46.9% (10.0 million ha) of its total area (Wardlow and Egbert 2008). Mapping crop types in this study site was complicated because of the similarity in spectra of different crop types, making it suitable for testing how well the proposed MFDC method can recognize complex crop types.

To further prove the effectiveness, generalization, and transferability of MFDC, an experiment was also conducted with 2017 images from another study site located in Texas State, USA, covering an area of 12 km × 12 km. This site encompasses corn, cotton, pasture and grassland, Sorghum, and winter wheat plantations.

2.2. Reference dataset

The National Agricultural Statistics Service Cropland Data Layers (CDL), featuring the highest accuracy in current crop-type classification for the two study areas for 2010 and 2017 were regarded as the reference data (Table 1). Furthermore, as the accuracy of CDL varies for non-crop types, pasture and grassland data were not considered.

2.3. Dataset processing

Landsat Surface Reflectance Data (LSRD) with 30 m spatial resolution as well as sub-pixel geolocation accuracy for 2010 and 2017 were downloaded from the United States Geological Survey Earth Resources Observation and Science Centre. The two study areas were covered by Landsat (Path 30, Row 34 and Path 31, Row 35, respectively, in the Worldwide Reference System-2). The LSRD of Landsat 5 and Landsat 7 were generated from the Landsat Ecosystem Disturbance Adaptive Processing System (Masek et al. 2006). The LSRD of Landsat 8 was generated from the Landsat Surface Reflectance Code (Vermote et al. 2016). In this study, subtle differences in the spectral ranges of the corresponding bands between Landsat 5/7 and Landsat 8 for the study area in Texas State were neglected because they did not have a significant impact on the spectral variability (Li, Jiang, and Feng 2014; Flood 2014).

Moderate-resolution Imaging Spectroradiometer (MODIS) Nadir Bidirectional Reflectance Distribution Function-Adjusted Reflectance 16 days composite data (MCD43A4 collection 6) with 500 m spatial resolution for 2010 and 2017 were obtained from the Land Processes Distributed Active Archive Centre to generate the fused Landsat-MODIS reflectance time series for crop classification, and MODIS tiles of h10v05 and h18v04 were used to cover the two study areas.

The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Feng, Schwaller, and Hall 2006) was used to generate RSTSI for the two study areas. For details

Table 1. Statistic of reference sample for the two study areas.

Label	Class name	Reference sample number	
		Study area 1	Study area 2
Class 1	Corn	59958	66497
Class 2	Cotton	–	18550
Class 3	Sorghum	15638	5769
Class 4	Soybeans	3370	–
Class 5	Winter wheat	37537	27833
Class 6	Alfalfa	16733	–
Class 7	Fallow/idle cropland	18521	17651
Class 8	Developed open space	5988	–

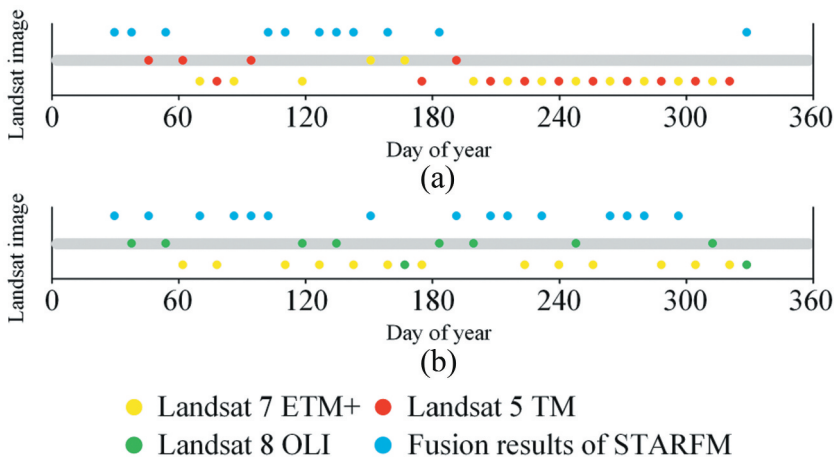


Figure 2. RSTSI in two study areas obtained using STARFM. Dots in the grey box indicate Landsat images that were used as input in the STARFM. Dots below the grey box represent Landsat images that were only used for constructing RSTSI. (a) Study area in Kansas, (b) Study area in Texas.

on STARFM, we refer the reader to (Feng, Schwaller, and Hall 2006). The constructed dense regular RSTSI of the two study areas each consisted of 38 images for the years 2010 and 2017 (Figure 2).

3. Methodology

3.1. Overview

The flow of the proposed MFDC method consists of the following steps (Figure 3): First, to make full use of the STS features of RSTSI, four complementary pre-defined views are generated (Section 3.1). Then, a novel labelled sample augmentation framework is used to enhance the classifier's generalization ability by iteratively increasing the number of training samples (Section 3.2). Third, MVSSF algorithm that expands the single view semi-supervised learning algorithm to multiple views, is applied to co-training to perform multi-view feature selection to ensure the effectiveness of ensemble learning without significant extra cost (Section 3.3). Finally, a weighted majority vote method is applied to obtain the final classification results (Section 3.4). In the following sections, we present a more detailed explanation of the proposed MFDC method.

3.2. Multi-view generation using multiple features

For effective utilization of the STS features of RSTSI, four pre-defined relatively complementary views were used as input for the co-training algorithm (Table 2), where each view represented a class of features.

View 1 represents the spectral index time series that consists of the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI) time series, which were utilized to record phenological changes. The NDVI time series has been widely accepted as a sensitive indicator of phenological variations, biomass changes

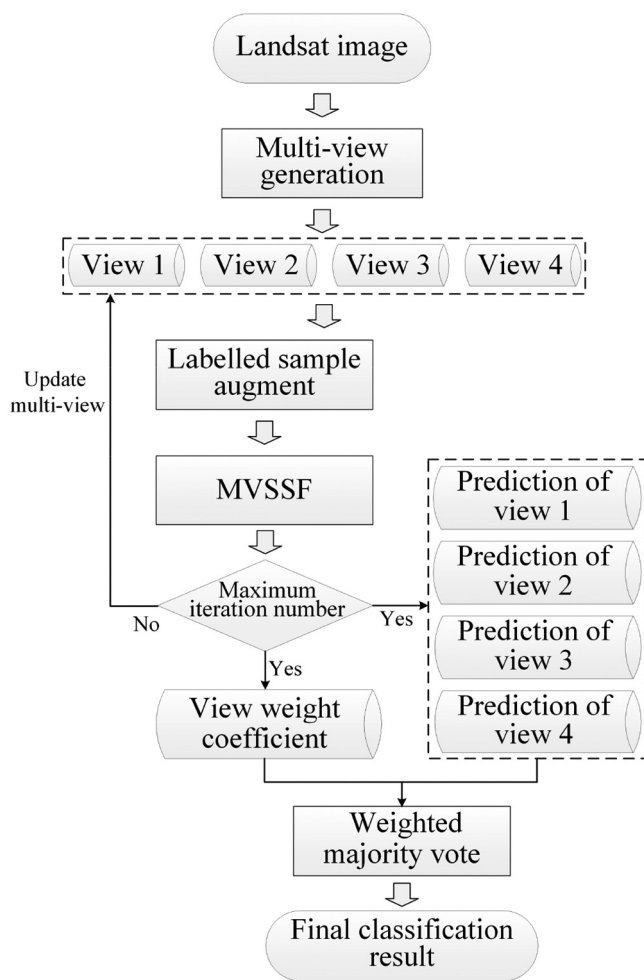


Figure 3. Framework of the proposed multiple-feature-driven co-training method.

Table 2. List of multiple feature, which were categorized into four views.

View order	Feature category	Feature	Number of feature
View1	Spectral index time series feature	NDVI time series, NDWI time series	76
View2	Tasselled cap transformation feature	Brightness, greenness, wetness	81
View3	Time series metrics feature	Standard deviation, quantile range, quantile value, amplitude, maximum, mean, median, minimum	64
View4	Multi-seasonal textural feature	Variance, homogeneity, contrast, dissimilarity, entropy, second moment, correlation	192

of vegetation, and crop classification (Shao et al. 2016; Sun et al. 2016; Sexton et al. 2013). NDWI is also referred to as the leaf area water absent index, and can contribute to improving crop mapping (Hao et al. 2015). It has also been widely utilized as an input variable for land cover classification (Fisher, Flood, and Danaher 2016; Dronova et al. 2015; Davranche, Lefebvre, and Poulin 2010).

View 2 represents tasselled cap transformation features, including the brightness, greenness, and wetness. These features, which are a typical method to enhance discrimination among spectral features, are beneficial for the accurate characterization of land cover types and are particularly useful for the identification of similar cover types (Sexton et al. 2013). A total of 27 clear Landsat top-of-atmosphere images were characterized using brightness, greenness, and wetness. The tasselled cap transformation features do not contain complete temporal information, so the correlation between view 1 and view 2 is alleviated to some extent.

View 3 represents the time series metrics features. Multi-temporal spectral data captured at various dates in the same or consecutive years can be summarized by statistical metrics (e.g., average, variability), which in turn function as descriptive or predictive variables and have shown to be a viable means for the discrimination of land cover types (Gómez, White, and Wulder 2016) (Gebhardt et al. 2014; Petitjean, Inglada, and Gancarski 2012). In this study, eight statistical metrics of NDVI, NDWI, and six Landsat reflectance band time series, including band 1–5, band 7 for Thematic Mapper (TM)/Enhanced Thematic Mapper Plus (ETM+) and band 2–7 for Operational Land Imager (OLI), were selected (Table 2). Note that although view 3 is derived from view 1, the correlation of the two views is alleviated using the proposed MVSSF (Section 3.3).

Although spectral and temporal features are vital for identifying land cover (Zhu and Woodcock 2014), these features are sensitive to variations in the vegetation density (Gómez, White, and Wulder 2016). Moreover, the above-mentioned views are generated based on the pixel scale that neglect spatial information, which was mitigated through the creation of view 4. It represents a multi-seasonal textural feature, with a proven potential for mapping land cover (Rodriguez-Galiano et al. 2012; Akar and Güngör 2015). A clear Landsat image was selected for each season and further eight textural features, including the mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlations with TM/ETM+'s band 1–5, band 7 and OLI's band 2–7, were calculated. Co-occurrence matrix measures (Haralick, Shanmugam, and Dinstein 1973) were utilized for describing the remote sensing texture, whose window size was 3×3 pixels.

3.3. Labelled sample augmentation

As previously described, the process of increasing the number of samples by iteration, defined as labelled sample augmentation in this paper, is a critical step to enhance the generalization ability of the classifier. Here, a novel labelled sample augmentation framework was designed for application on the four pre-defined views. The flowchart of the proposed scheme consists of the following steps (Figure 4):

First, the posterior probability set of each view was generated using an SVM and initially labelled training samples X^{Label} . Since standard SVM classifiers do not provide a posterior probability, the posterior probabilities of a multi-class classification were obtained using the voting method based on the posterior probabilities of binary classification, as proposed by Platt (Platt 2000). $P(\mathbf{y}_t = l | \mathbf{x}_t)$ is defined as follows:

$$P(\mathbf{y}_t = l | \mathbf{x}_t) = \frac{\sum_{j'=1, j' \neq l}^c P_{l, j'}(\mathbf{y}_t = l | \mathbf{x}_t)}{\sum_{k'=1}^c \sum_{j'=1, j' \neq k'}^c P_{k', j'}(\mathbf{y}_t = l | \mathbf{x}_t)} \quad (1)$$

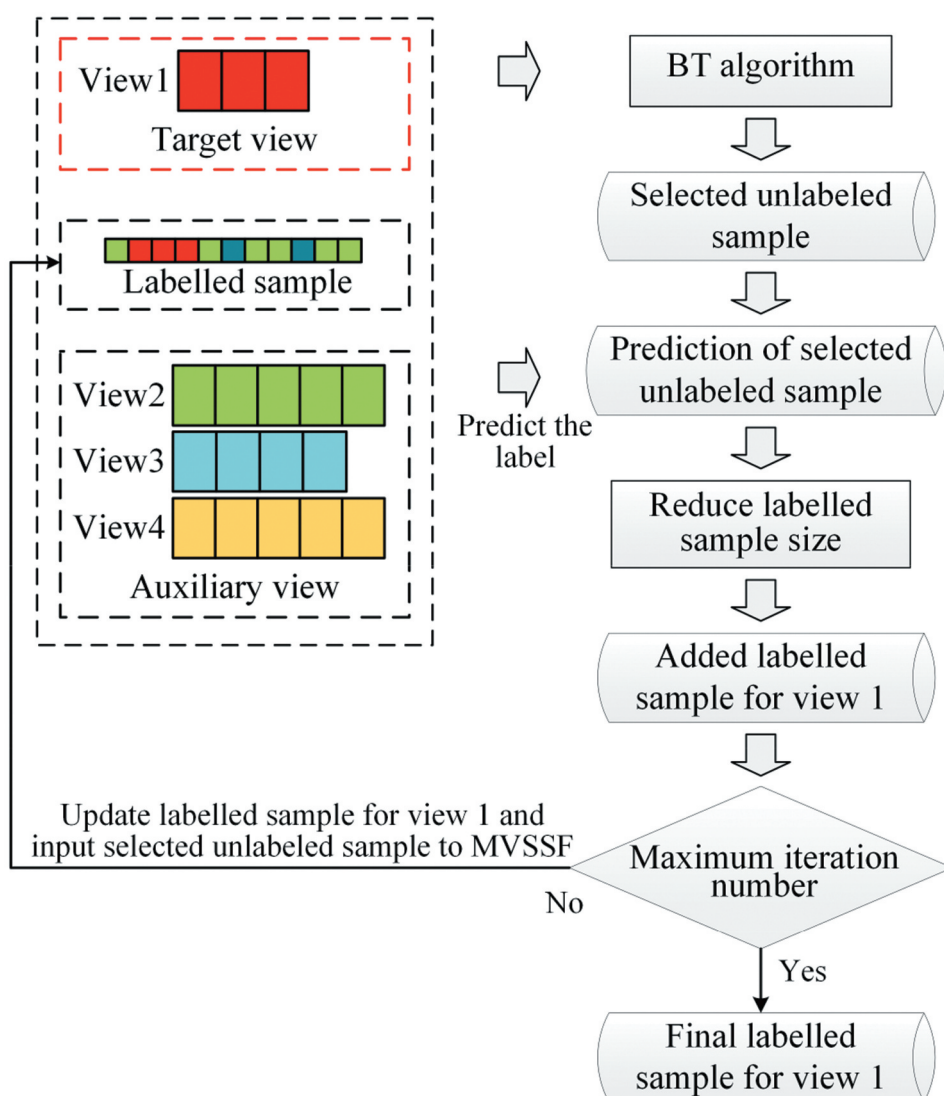


Figure 4. Framework of the proposed labelled sample augmentation, taking view 1 as an example. The process for the other views is the same as that of view 1.

where $P_{l,j'}(\mathbf{y}_t = l | \mathbf{x}_t)$ represents the posterior probability of sample \mathbf{x}_t being assigned to class l as calculated independently using a binary SVM, which discriminates only between class l and class j' , $P_{k',j'}(\mathbf{y}_t = l | \mathbf{x}_t)$ represents the posterior probability of sample \mathbf{x}_t being assigned to class k' as calculated independently using a binary SVM, which discriminates only between class k' and class j' , c is the total number of classes.

To further improve the representativeness of the selected unlabelled samples, the Breaking Tiles (BT) algorithm (Bruzzone and Persello 2009) was utilized to achieve more diversity in the unlabelled samples of each view. The BT rule is defined as follows:

$$\hat{\mathbf{x}}_t^{\text{BT}} = \underset{\mathbf{x}_t \in X_u}{\operatorname{argmin}} \left\{ \max_{l \in \{1, \dots, c\}} P(\mathbf{y}_t = l | \mathbf{x}_t) - \max_{l \in \{1, \dots, c\} \setminus \{l^+\}} P(\mathbf{y}_t = l | \mathbf{x}_t) \right\} \quad (2)$$

where $l^+ = \underset{l \in \{1, \dots, c\}}{\operatorname{argmax}} P(\mathbf{y}_t = l | \mathbf{x}_t)$ is the most probable class for sample \mathbf{x}_t , and $P(\mathbf{y}_t = l | \mathbf{x}_t)$

represents the posterior probability of \mathbf{x}_t being assigned to class l . X_u represents unlabelled samples.

Subsequently, the i -th view's selected unlabelled samples are labelled if the other three classifiers agree on the labelling of these samples. To further reduce the number of the selected unlabelled samples, thus reducing computation time and guaranteeing the diversity of the unlabelled samples, a sample is selected if the label predicted by the other three classifiers is different from the label predicted by the first classifier. The reasoning is that these samples are labelled with high confidence but do not improve the generalization ability, as they do not add to the diversity.

Finally, the selected unlabelled samples were added to the initial labelled samples to increase the labelled sample size, and were then input to the MVSSF to optimize the multi-view features (Section 3.3).

3.4. Multi-view semi-supervised feature learning

MVSSF was designed to perform multi-view feature selection to alleviate the dependency between different views to ensure the effectiveness of ensemble learning and to speed up the learning process.

Inspired by the Structural Feature Selection with Sparsity method (Ma et al. 2012), in MVSSF we expand a single-view semi-supervised feature selection to multi-view. The global objective function of MVSSF is:

$$\min \sum_{i=1}^4 \mathbf{u}_i^T \left(\operatorname{Tr}(\mathbf{F}_i^T \mathbf{L}_i \mathbf{F}_i) + \operatorname{Tr}((\mathbf{F}_i - \mathbf{Y})^T \mathbf{U}(\mathbf{F}_i - \mathbf{Y})) + \mu \mathbf{X}_i^T \mathbf{W}_i + 1_{n_i} \mathbf{b}_i^T - \mathbf{F}_{iF}^2 \right) + \gamma \mathbf{W}_{i2,1} \quad (3)$$

subject to $\sum_{i=1}^4 \mathbf{u}_i = 1, \quad \mathbf{u}_i \geq 0$

where $X = \{X_i \in \mathbb{R}^{d_i \times n_i}\}_{i=1}^m$ denotes the set of views which consists of each view's selected unlabelled samples, and X_i denotes the i -th view. d_i denotes the number of features in the feature space of the i -th view, and n_i denotes the number of selected unlabelled samples of the i -th view. $\mathbf{W}_i \in \mathbb{R}^{d_i \times c}$ denotes a projection matrix of the i -th view used for feature selection, c is the class number, while μ , γ , and r are pre-defined regularization parameters. $\|\cdot\|_F$ denotes the Frobenius norm, and $\|\cdot\|_{2,1}$ denotes the $L_{2,1}$ norm. Each terms are explained below.

The manifold regularization term $\operatorname{Tr}(\mathbf{F}_i^T \mathbf{L}_i \mathbf{F}_i)$ is utilized to address each view's high-dimensional data with the manifold structure, where $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_{n_i}]^T \in \mathbb{R}^{n_i \times c}$ is the predicted label. The graph Laplacian matrix $\mathbf{L}_i \in \mathbb{R}^{n_i \times n_i}$ is defined as

$$\mathbf{L}_i = \mathbf{D}_i - \mathbf{S}_i \quad (4)$$

where \mathbf{D}_i is a diagonal matrix with the diagonal elements $\mathbf{D}_i^{p,p} = \sum_q \mathbf{S}_i^{p,q}$, $p, \mathbf{S}_i^{p,q} \in \mathbb{R}^{n_i \times n_i}$ denotes the similarity matrix as

$$\mathbf{S}_i^{p,q} = \begin{cases} 1, & x^p \text{ and } x^q \text{ are } k \text{ nearest neighbors} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The relation between different views is highlighted by the term $\text{Tr}((\mathbf{F}_i - \mathbf{Y})^T \mathbf{U}(\mathbf{F}_i - \mathbf{Y}))$, where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n_i}]^T \in \{0, 1\}^{n_i \times c}$ is the label matrix, constructed by fixed use of the initial labelled training samples X^{Label} . If $X_{i'}$ belongs to the j' -th class, $\mathbf{Y}_{i',j'} = 1$ and $\mathbf{Y}_{i',j'} = 0$; otherwise, $\mathbf{y}_{i'}$ is set to a vector with all zeros. $\mathbf{U} \in \mathbb{R}^{n_i \times n_i}$ is a selecting diagonal matrix whose diagonal element $\mathbf{U}_{t,t} = \infty$ if X_t is labelled and $\mathbf{U}_{t,t} = 1$ otherwise.

The label consistency constraint term $X_i^T \mathbf{W}_i + \mathbf{1}_{n_i} \mathbf{b}_i^T - \mathbf{F}_{iF}^2$ is added to relate the selected features of each view to the concepts of the initial labelled training samples. Here, $\mathbf{b}_i \in \mathbb{R}^c$ is the bias term of the i -th view.

Because the above three terms all contain predicted labels, we further apply the linear weighting of each view's regularization term with regard to \mathbf{u}_i^r as the view weight coefficients. Here, r adjusts the sparsity of the view weight coefficients, where a smaller value highlights the weight of important views.

The regularization term $\|\mathbf{W}_i\|_{2,1}$ guarantees sparsity to be suitable for feature selection.

Once the i -th view's selected unlabelled samples are generated by the labelled sample augmentation, these samples, along with the initial labelled training samples X^{Label} , are utilized to solve the optimization problem (Equation (3)) and obtain the selected features for each view and the view weight coefficients \mathbf{u}_i . The optimization process of the MVSSF is summarized in Table 3. More details for the variables updating can be found in the supplemental material.

3.5. Classification based on the weighted majority vote method

With the increase in the labelled training sample number, and the optimization of multi-view features achieved through labelled sample augmentation and MVSSF respectively, the classifier's generalization ability on each view is enhanced.

Table 3. Multi-view semi-supervised feature learning process.

Step	Process
	Input: X, μ, γ, r, S_i Output: selected features for each view and \mathbf{u}_i
1:	Construct graph Laplacian matrix \mathbf{L}_i
2:	Compute the selecting matrix \mathbf{U}
3:	Initialize \mathbf{u}_i as $\frac{1}{4}$
4:	Initialize \mathbf{W}_i randomly
5:	While not convergent to
6:	For $i = 1$ to 4 do
7:	Update \mathbf{b}_i^T using equations (A1).
8:	Update \mathbf{W}_i using equations (A4).
9:	Update \mathbf{u}_i using equations (A7).
10:	Sort each view's feature according to $\ \mathbf{W}_i(\hat{k}, :)\ _2$ in descending order and select the top- \hat{k} ranked entries to obtain the selected feature for each view
11:	End for
12:	End while
13:	Return selected feature for each view and \mathbf{u}_i

Therefore, upon reaching the maximum iterations number, a weighted majority vote method is utilized to predict the final label. The combination rule is as follows:

$$C(X) = \max_{c'=1,2,\dots,c} \sum_{i=1}^4 \mathbf{u}_i P(c'|X_i) \quad (6)$$

where $C(X)$ is the final class label of the the set of views, \mathbf{u}_i denotes the view weight coefficients obtained using MVSSF and $P(c'|X_i)$ is the posterior probability of sample to be classified as class c' on the i -th view.

4. Experiment

4.1. Experimental strategy

In this section, comparison experiments are presented to demonstrate the effectiveness of MFDC. MFDC, along with other semi-supervised learning frameworks, was implemented with 10, 15, 20, 25, and 100 initial labelled training samples per class. The different semi-supervised learning frameworks used for comparison are as follows:

- (1) Self-training performance on multi-view, denoted as Mv-Self. Specifically, during the labelled sample augmentation process, the BT algorithm was utilized to select unlabelled samples for each view. Then, the label of the selected unlabelled samples was predicted by an initial classification map while the final classification result was generated through a majority vote.
- (2) Multi-view co-training, denoted as Mv-Co. During the augmentation process, an initial candidate sample set for each view was generated by selecting the samples with consistently predicted labels in other views. Furthermore, some samples were randomly selected from the initial candidate sets and added to the training samples of each view. The final classification result was generated through a majority vote.
- (3) Multi-view co-training, denoted as Mv-Self-Co. During the augmentation process, the BT algorithm was utilized to select unlabelled samples from each view, which were then labelled upon agreement with the other three classifiers. The final classification result was generated through a majority vote.
- (4) The proposed MFDC algorithm.

MFDC was also compared with the use of SVM with principal component analysis (PCA) features to verify the superiority of using STS features over using spectral features only. To achieve this, the spectral bands of all the clear Landsat images were first stacked together, followed by PCA. The first p' components with 90% information were selected as input for the SVM. The number of input features of study area 1 and study area 2 were 14 and 16, respectively. The details of the algorithms are shown in [Table 4](#).

The performance of the above-mentioned algorithms was evaluated by calculating their overall accuracy (OA) and average accuracy (AA). The initial training samples were

Table 4. Detail of the algorithms involved in the experiments.

Algorithm	Labelled sample augment	Feature learning	Feature
Mv-Self	BT	None	STS
Mv-Co	Co-training	None	STS
Mv-Self-Co	BT and co-training	None	STS
MFDC	BT and co-training	MVSSF	STS
SVM-PCA	None	PCA	Original spectral

selected randomly per class from the reference dataset. The final accuracy results are the average of ten independent runs.

4.2. Parameter setup

In the experiments, the parameters were set as follows:

- (1) Labelled sample augmentation: the maximum number of iterations was set to 20. The number of the selected unlabelled samples in each iteration for all the semi-supervised learning frameworks was 100.
- (2) Classifier parameter: the linear kernel function type was used for the SVM. The penalty factor was 0.5, which was obtained by comparing results obtained with the factor ranging from 2^{-8} to 2^8 , using SVM as the classifier with 25 labelled training samples per class.
- (3) MVSSF: The k nearest neighbour range was set to 5. r was fixed to 2. We compared all literature-reported values for the regularization parameters μ and γ (Ma et al. 2012), namely {0.001, 0.01, 0.1, 1, 10, 100 and 1000}, to obtain the best results.

4.3. Experimental results

The initial and final classification accuracies of the different semi-supervised learning frameworks for the two study areas are shown in Tables 5 and 6. Meanwhile, SVM-PCA, which simply uses PCA-transformed features, was regarded as a benchmark scenario to test the effectiveness of STS features.

In area 2, the initial classification accuracies of the semi-supervised learning frameworks are higher than those of SVM-PCA, which indicates the superiority of STS features over the original spectral features. However, in area 1, the accuracy of semi-supervised learning is significantly lower than that of SVM-PCA when the initial training sample is 10. This could be related to the fact that, compared to the original spectral features, the higher dimensionality of the STS features outweighs the fact that the small number of training samples. In this case, the use of PCA is an effective way to achieve higher accuracy compared to STS features. Nevertheless, the accuracy of SVM-PCA is still unsatisfactory. The STS features mitigate the issues of the ill-posed classification problem through the multi-view co-training method. As shown in Tables 5 and 6, the significant improvement in accuracy of the semi-supervised learning frameworks, compared to that of SVM-PCA verify this potential of the STS features.

Table 5. OA and AA \pm standard deviation of different semi-supervised learning frameworks for study area 1. The column labelled Initial presents classification using only the initial labelled samples.

Initial number	Method	OA		AA	
		Initial	Final	Initial	Final
10	SVM-PCA	0.7904 \pm 0.0110	-	0.6971 \pm 0.0190	-
	Mv-Self	0.7795 \pm 0.0377	0.8552 \pm 0.0052	0.6717 \pm 0.03429	0.7693 \pm 0.0122
	Mv-Co		0.8638 \pm 0.0014		0.8051 \pm 0.0093
	Mv-Self-Co		0.8773 \pm 0.0006		0.8211 \pm 0.0024
	MFDC		0.8777 \pm 0.0011		0.8269 \pm 0.0032
15	SVM-PCA	0.7959 \pm 0.0108	-	0.7018 \pm 0.0135	-
	Mv-Self	0.8000 \pm 0.0129	0.8590 \pm 0.0041	0.6959 \pm 0.0156	0.7746 \pm 0.0105
	Mv-Co		0.8645 \pm 0.0020		0.8021 \pm 0.0089
	MV-Self-Co		0.8776 \pm 0.0007		0.8181 \pm 0.0028
	MFDC		0.8778 \pm 0.0006		0.8277 \pm 0.0022
20	SVM-PCA	0.8043 \pm 0.0111	-	0.7089 \pm 0.0093	-
	Mv-Self	0.8087 \pm 0.0108	0.8617 \pm 0.0044	0.7054 \pm 0.0080	0.7821 \pm 0.0108
	Mv-Co		0.8650 \pm 0.0023		0.7999 \pm 0.0093
	MV-Self-Co		0.8778 \pm 0.0008		0.8188 \pm 0.0029
	MFDC		0.8784 \pm 0.0009		0.8247 \pm 0.0028
25	SVM-PCA	0.8115 \pm 0.0100	-	0.7111 \pm 0.0107	-
	Mv-Self	0.8133 \pm 0.0083	0.8631 \pm 0.0038	0.7154 \pm 0.0061	0.7834 \pm 0.0104
	Mv-Co		0.8665 \pm 0.0013		0.7974 \pm 0.063
	Mv-Self-Co		0.8778 \pm 0.0008		0.8176 \pm 0.0033
	MFDC		0.8784 \pm 0.0007		0.8232 \pm 0.0026
100	SVM-PCA	0.8332 \pm 0.0071	-	0.7349 \pm 0.0060	-
	Mv-Self	0.8306 \pm 0.0046	0.8676 \pm 0.0030	0.7319 \pm 0.0074	0.7829 \pm 0.0030
	Mv-Co		0.8692 \pm 0.0014		0.7832 \pm 0.0048
	Mv-Self-Co		0.8783 \pm 0.0009		0.8041 \pm 0.0026
	MFDC		0.8788 \pm 0.0009		0.8092 \pm 0.0026

Table 6. OA and AA \pm standard deviation of different semi-supervised learning frameworks for study area 2. The column labelled Initial presents classification using the initial training labelled samples.

Initial number	Method	OA		AA	
		Initial	Final	Initial	Final
10	SVM-PCA	0.5889 \pm 0.0521	-	0.4070 \pm 0.0673	-
	Mv-Self	0.6915 \pm 0.0211	0.8376 \pm 0.0083	0.5560 \pm 0.0497	0.7277 \pm 0.0641
	Mv-Co		0.8513 \pm 0.0027		0.7617 \pm 0.0269
	MV-Self-Co		0.8711 \pm 0.0015		0.7800 \pm 0.0141
	MFDC		0.8716 \pm 0.0030		0.7920 \pm 0.0137
15	SVM-PCA	0.6756 \pm 0.0476	-	0.4950 \pm 0.1027	-
	Mv-Self	0.7151 \pm 0.0375	0.8380 \pm 0.0090	0.6030 \pm 0.0543	0.7142 \pm 0.0403
	Mv-Co		0.8523 \pm 0.0012		0.7654 \pm 0.0288
	MV-Self-Co		0.8711 \pm 0.0026		0.7937 \pm 0.0099
	MFDC		0.8729 \pm 0.0020		0.7863 \pm 0.0170
20	SVM-PCA	0.6988 \pm 0.0259	-	0.5478 \pm 0.0831	-
	Mv-Self	0.7486 \pm 0.0237	0.8409 \pm 0.0112	0.6365 \pm 0.0367	0.7073 \pm 0.0440
	Mv-Co		0.8531 \pm 0.0031		0.7566 \pm 0.0221
	MV-Self-Co		0.8714 \pm 0.0027		0.7915 \pm 0.0107
	MFDC		0.8731 \pm 0.0034		0.7928 \pm 0.0159
25	SVM-PCA	0.7043 \pm 0.0190	-	0.5585 \pm 0.0291	-
	Mv-Self	0.7643 \pm 0.0191	0.8411 \pm 0.0062	0.6594 \pm 0.0195	0.7197 \pm 0.0362
	Mv-Co		0.8539 \pm 0.0021		0.7598 \pm 0.0200
	MV-Self-Co		0.8723 \pm 0.0020		0.7873 \pm 0.0195
	MFDC		0.8729 \pm 0.0021		0.7955 \pm 0.0171
100	SVM-PCA	0.7742 \pm 0.0113	-	0.6967 \pm 0.0105	-
	Mv-Self	0.8139 \pm 0.0063	0.8573 \pm 0.0019	0.7120 \pm 0.0087	0.7489 \pm 0.0075
	Mv-Co		0.8618 \pm 0.0034		0.7473 \pm 0.0129
	MV-Self-Co		0.8734 \pm 0.0028		0.7694 \pm 0.0185
	MFDC		0.8734 \pm 0.0036		0.7925 \pm 0.0099

In terms of comparison of different semi-supervised learning frameworks, it is clear that MFDC performs well and stably compared with the other semi-supervised learning frameworks. As shown in Table 5, compared to the initial accuracy, MFDC achieves the highest improvement in OA for area 1, when using 10, 15, 20, 25, and 100 samples per class as the initial labelled samples. Similar behaviour can be observed in Table 6, where the MFDC achieves the highest classification accuracies in study area 2. A comparative analysis of the data presented in Tables 5 and 6, indicates a more striking improvement in the initial classification performance when using 10, 15, 20, 25, and 100 samples per class as the initial labelled samples. To summarize, the proposed MFDC can achieve more accurate classification results using a limited number of training label samples.

The classification maps corresponding to the average OAs of ten runs obtained using the various methods are shown in Figures 5 and 6. As seen in the dotted box, compared to SVM-PCA, the initial classification results of semi-supervised learning frameworks

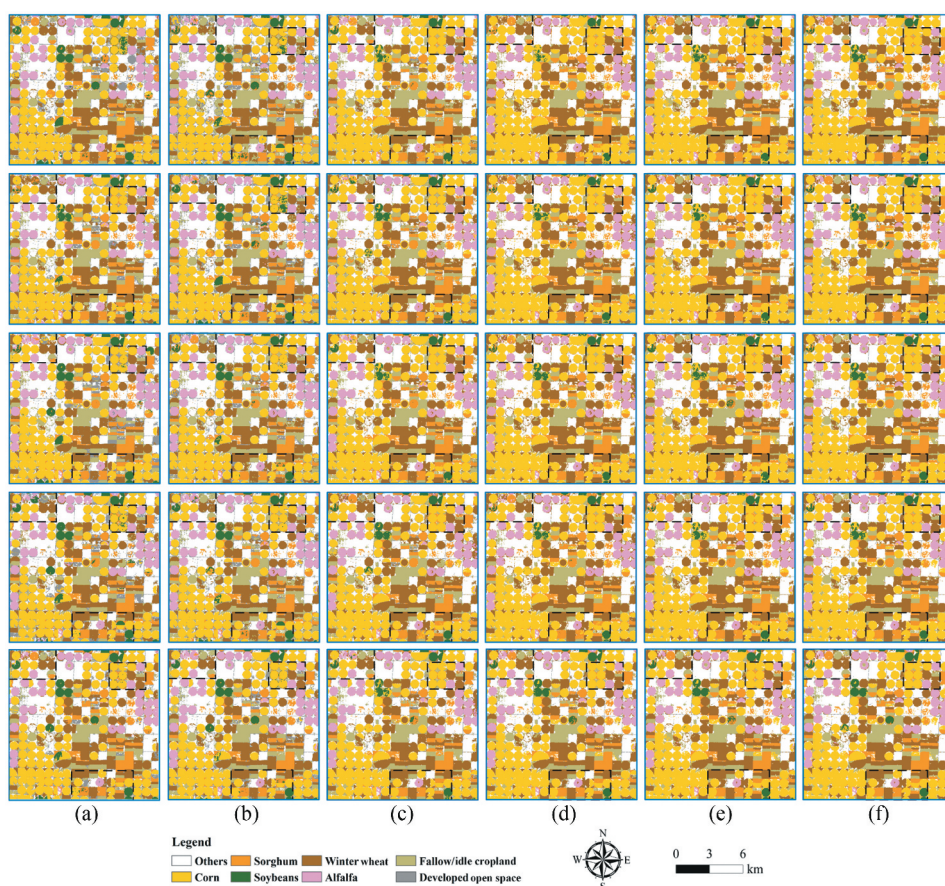


Figure 5. Classification maps of study area 1 obtained by different learning frameworks with different labelled training sample sizes per class. A dotted box highlights areas with significant differences in the classification maps. The first line to the fifth line represent the labelled training sample number of 10, 15, 20, 25, and 100. (a) SVM-PCA, (b) Initial classification map, (c) Mv-Self, (d) Mv-Co, (e) Mv-Self-Co, (f) MFDC.

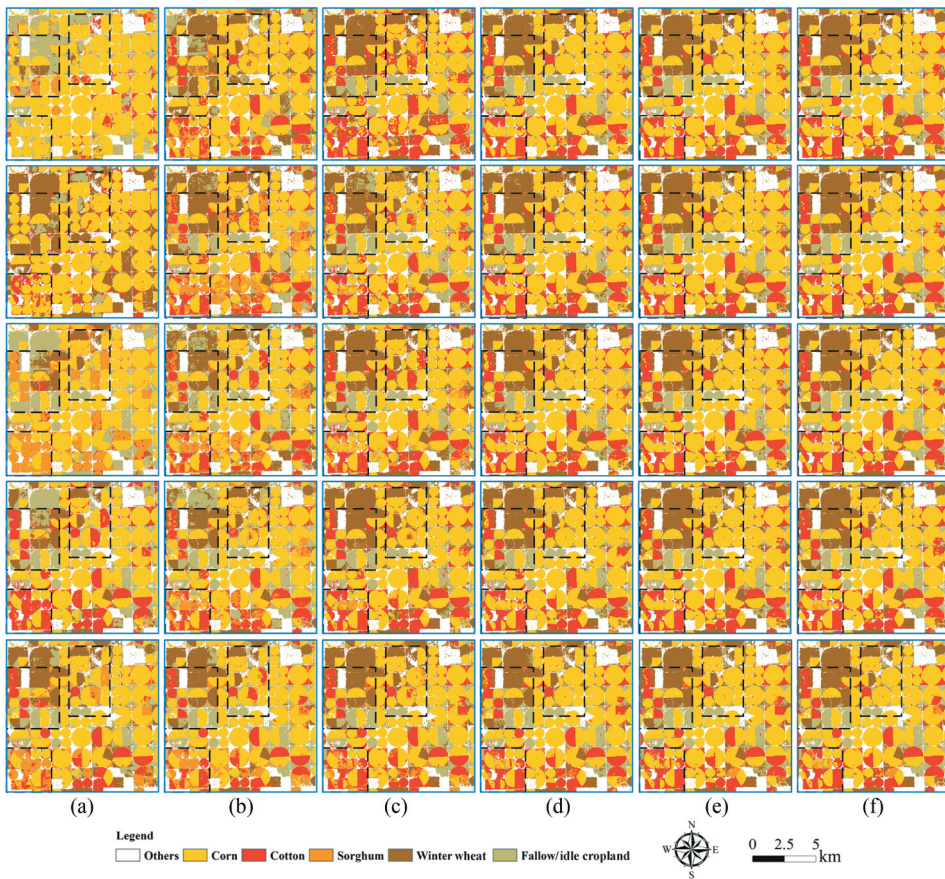


Figure 6. Classification maps of study area 2 obtained by different learning frameworks with different labelled training sample sizes per class. A dotted box highlights areas with significant differences in the classification maps. The first line to the fifth line represent the labelled training sample number of 10, 15, 20, 25, and 100. (a) SVM-PCA, (b) Initial classification map, (c) Mv-Self, (d) Mv-Co, (e) Mv-Self-Co, (f) MFDC.

generate smoother results, especially in study area 2. Additionally, the final classification maps obtained by the semi-supervised learning frameworks are more consistent with the reference dataset, as shown in Figure 1, in some mixed regions of multiple crop types than the initial classification results. With regard to the different semi-supervised learning frameworks, similar results to those presented in Tables 5 and 6 can be observed, where Mv-Self generates the worst results. The classification maps obtained by Mv-Self-Co and MFDC are more consistent with the reference dataset.

To compare the different semi-supervised learning frameworks in further detail, the accuracies of the above-mentioned frameworks in the iteration process were evaluated. As shown in Figures 7 and 8, in general, the OAs of all views increase with the number of iterations, which indicates that the iteratively selected unlabelled samples are conducive to the classification process. The MFDC achieves higher OAs than the other semi-supervised learning frameworks, which indicates its superiority. Specifically, compared to Mv-Co and Mv-Self-Co, the accuracy of Mv-Self is unsatisfactory. This can be attributed to the fact that,

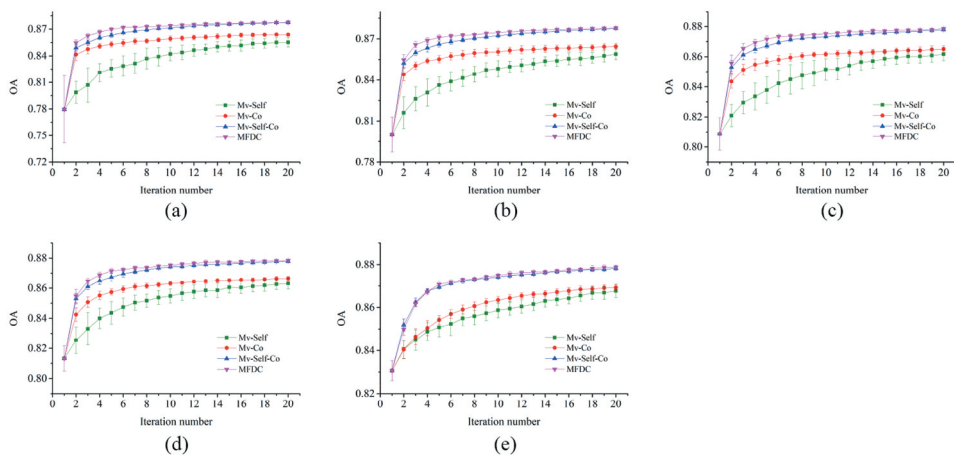


Figure 7. Iteration process in different semi-supervised learning frameworks for study area 1 conducted with different labelled training sample numbers. The error bars indicate the standard deviations. (a) Initial number: 10, (b) Initial number: 15, (c) Initial number: 20, (d) Initial number: 25, (e) Initial number: 100.

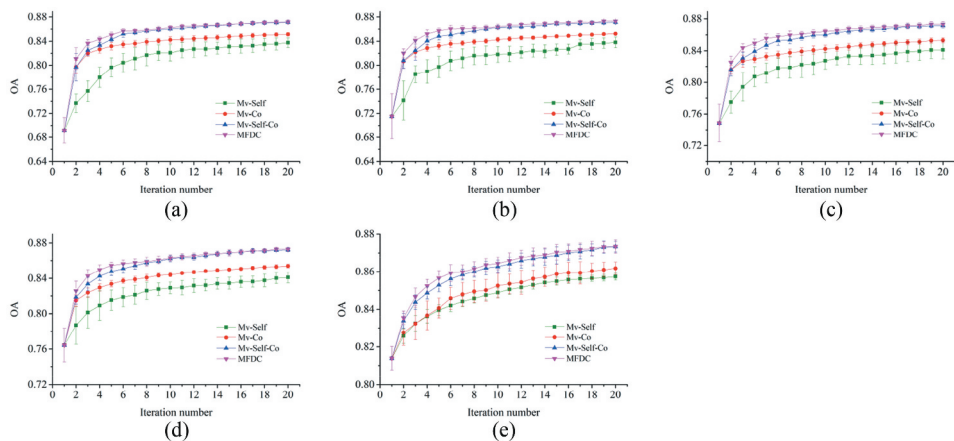


Figure 8. Iteration process in different semi-supervised learning frameworks for study area 2 for different labelled training sample numbers. The error bars indicate the standard deviations. (a) Initial number: 10, (b) Initial number: 15, (c) Initial number: 20, (d) Initial number: 25, (e) Initial number: 100.

although the unlabelled samples selected by the BT algorithm are informative, the addition of these unlabelled samples to the classifier will increase the cumulative error. With regard to Mv-Co, for which the accuracy is slightly higher than that of Mv-Self, the improvement in the accuracy may be related to the effective utilization of information between the views. However, due to the lack of unlabelled sample diversity, there is no significant improvement in the generalization ability of classifier. Mv-Self-Co and MFDC outperform Mv-Co and Mv-Self on all indexes, which indicates the effectiveness of the combination of the BT algorithm with the traditional co-training algorithm. Compared to the Mv-Self-Co, MFDC shows

a higher accuracy, which can be attributed to the ability of the proposed MVSSF to ensure the effectiveness of ensemble learning in co-training.

As described above, the STS features are superior to the original spectral features, as they provide more detailed information for classification. Moreover, the STS features extracted from the four pre-defined views can be utilized effectively by the combination of labelled sample augmentation and the proposed MVSSF algorithm.

5. Discussion

5.1. Effectiveness of multi-view semi-supervised feature learning

MVSSF is proposed and was embedded in co-training to perform multi-view feature selection to alleviate the dependency between different views to ensure the effectiveness of ensemble learning and to reduce the computational cost. The experimental results verified that the MVSSF algorithm ensure the effectiveness of ensemble learning, to further verify that the MVSSF algorithm improves the efficiency of the semi-supervised learning frameworks effectively, the corresponding running times were compared. Timing was performed using MATLAB R2016b on a desktop PC equipped with an Intel (R) Core (Trade Mark) i7-8700 CPU @3.20 GHz, 16.00 GB Ram. As shown in Table 7, there is a trade-off between the running time and accuracy; in general, the longer the calculation time of the algorithm, the better the accuracy. Although Mv-Self incurs the lowest computational cost in most cases, its classification performance is obviously lower than that of other frameworks, as mentioned previously. MFDC achieves very high classification accuracies for a low computational cost, especially compared to Mv-Self-Co. Therefore, the MVSSF algorithm improves the efficiency of semi-supervised learning.

Table 7. Running time \pm standard deviation (s) of different semi-supervised learning frameworks.

Initial number	Method	Running time (s)	
		Study area 1	Study area 2
10	Mv-Self	395.5519 \pm 12.7598	295.3951 \pm 8.7989
	Mv-Co	490.2486 \pm 21.4323	446.3365 \pm 8.7989
	Mv-Self-Co	576.1367 \pm 12.1198	463.2013 \pm 9.7850
	MFDC	454.3984 \pm 3.1104	363.2256 \pm 9.7850
15	Mv-Self	415.8331 \pm 7.7610	308.3851 \pm 5.7525
	Mv-Co	546.8348 \pm 12.4625	468.8220 \pm 8.5640
	Mv-Self-Co	607.8910 \pm 10.9665	484.3865 \pm 12.5028
	MFDC	491.3087 \pm 6.5493	376.5986 \pm 4.7117
20	Mv-Self	441.6154 \pm 8.1235	322.3924 \pm 4.8149
	Mv-Co	583.0950 \pm 12.5041	485.9817 \pm 8.0268
	Mv-Self-Co	645.1630 \pm 9.8411	506.9458 \pm 13.2872
	MFDC	513.9933 \pm 10.2052	391.1194 \pm 6.7013
25	Mv-Self	463.8812 \pm 7.7165	337.8055 \pm 4.5821
	Mv-Co	596.9593 \pm 12.2454	500.5225 \pm 8.5093
	Mv-Self-Co	676.5910 \pm 11.4239	520.7545 \pm 9.2917
	MFDC	526.9529 \pm 15.9470	406.5405 \pm 10.5227
100	Mv-Self	675.5836 \pm 11.6975	474.7314 \pm 7.7222
	Mv-Co	852.2023 \pm 9.3716	627.8263 \pm 6.0559
	Mv-Self-Co	878.3109 \pm 7.6532	651.0536 \pm 5.7707
	MFDC	586.7979 \pm 2.8871	429.4411 \pm 3.3492

5.2. Advantages of the multiple-feature-driven Co-training method

The MFDC algorithm was proposed for accurately mapping crop types based on RSTSI using limited numbers of labelled training samples. The experimental results verified the superiority of the proposed algorithm. Compared to the original spectral features, the STS features obtained using four pre-defined views mitigate the issues of the ill-posed classification problem through multi-view co-training. Furthermore, the proposed labelled sample augmentation iteratively selects reliable samples with predicted labels to the training samples to increase the training sample size, while MVSSF relieves the imbalance between the feature dimensions and the limited number of training samples by reducing the number of feature dimensions.

Compared to other semi-supervised learning frameworks, the proposed MFDC can achieve higher performance without significant extra computational cost. The superiority can be attributed to the following aspects:

The proposed novel labelled sample augmentation of MFDC enhances the information and representativeness of the selected unlabelled samples. The BT algorithm, which searches the samples at the edge of the hyperplane to add information to the labelled training samples (Bruzzone and Persello 2009) was introduced to enhance representativeness. However, the accuracy is likely to decrease with the accumulation of uncertainty due to a lack of effective strategies to ensure the reliability of the prediction results. The general strategy is to introduce an additional algorithm (Tan et al. 2015; Ma, Wang, and Wang 2016) and auxiliary data (Kim, Park, and Lee 2017) in the learning process to predict the label directly or optimize the prediction results (Tan et al. 2014; Dópido et al. 2013). However, these strategies lead to an inevitable increase in the computational cost, as well as some additional uncertainties. In contrast, MFDC improves the accuracy of the predicted labels of selected unlabelled samples by increasing the number of views instead of introducing additional complex strategies and auxiliary data. Therefore, the introduction of additional uncertainties and computational cost can be avoided.

5.3. Adaptability of the multiple-feature-driven Co-training method

Although the effectiveness of the MFDC method was verified in the two study areas, it is worth discussing the applicability of the proposed MFDC method to other cases.

In the experiment, we limited the number of training samples per class to 10 samples, which we consider to be the minimum, taking into account the relatively small extent of the two study areas and the complexity of crop types. When the complexity of crop cover is high, more representative training samples are required.

The proposed algorithm was tested in the two study areas to a relatively small spatial extent. To determine the influence of spatial extent on the accuracy of the MFDC method, the method was further applied in a larger study area with an extent of over 10 times that of study area 1. This study area is located in Kansas State and covers 2025 km². The location of this study area is shown in the supplemental material. CDL data for 2014 were regarded as the reference data. LSRD and MCD43A4 collection 6 were utilized to generate RSTSI in the extended study area for 2014 using STARFM. The parameter setup of the MFDC method was the same as that presented in 4.2. Clearly, in the larger study area, the accuracy of the proposed MFDC is still substantial (Table 8). Compared to the initial

Table 8. OA \pm standard deviation and AA \pm standard deviation of MFDC in the larger study area. The column labelled Initial presents classification using the initial labelled training samples.

Initial number	OA		AA	
	Initial	Final	Initial	Final
10	0.7234 \pm 0.0645	0.8473 \pm 0.0045	0.6416 \pm 0.0199	0.7496 \pm 0.0458
15	0.7496 \pm 0.0218	0.8482 \pm 0.0034	0.6555 \pm 0.0120	0.7455 \pm 0.0453
20	0.7613 \pm 0.0213	0.8501 \pm 0.0035	0.6622 \pm 0.0130	0.7612 \pm 0.0377
25	0.7734 \pm 0.0170	0.8515 \pm 0.0026	0.6687 \pm 0.0070	0.7567 \pm 0.0417

accuracy, MFDC still achieves a significant improvement in accuracy. Thus, the proposed algorithm is applicable to larger areas.

In some cases, the class imbalance, which indicates the proportion between different categories to be extracted, differs greatly in general. The proposed algorithm is not aimed at the problem of class-imbalance. Thus, it is difficult for the MFDC to provide satisfactory results in the presence of both class-imbalance and a limited number of training samples. In this case, we recommend that the MFDC should be applied in a hierarchical way, which involves merging similar categories and mapping these merged categories first, followed by further mapping of detailed categories.

5.4. Limitations and future work

The proposed method can accurately map crop types using a limited number of training samples without significant computational cost. However, the applicability of the MFDC still has some limitations:

The accuracy of MVSSF is sensitive to the number of selected features. Additional efforts should be made to improve the robustness of MVSSF in the presence of limited numbers of samples. These efforts include the application of a more effective graph construction method, such as Sparse Graph Regularization (SGR) and SGR with total variation (TV-SGR) (Xue et al. 2017), to express the manifold structure with fewer samples. The predicted label of the added sample F_i can also be utilized to augment the labelled samples in the process of feature learning.

The proposed MVSSF requires a relatively large number of pre-defined parameters and is strongly sensitive to some of these parameters. For example, the sensitivity analysis of regularization parameters μ and γ shows that different parameter settings have a significant impact on the accuracy. Fine-tuning all these parameters to select the optimal value is time-consuming, which limits the practicality of the proposed algorithm.

Finally, in this study, agricultural areas with a single-season cropping system were selected, and the effectiveness and efficiency of the proposed MFDC method were tested. Undeniably, more systematic analysis is necessary to determine the algorithm's adaptability to other study areas with other complex land cover types.

6. Conclusions

In this study, we propose a novel co-training method for accurately mapping crop types using RSTSI with a limited number of labelled training samples. In the proposed method, a labelled sample augmentation method that combines the BT algorithm

and co-training is designed to enhance the representativeness and label estimation accuracy of the selected unlabelled samples. Additionally, the MVSSF, which expands the single view semi-supervised learning algorithm to multiple views, is proposed and embedded in the framework to ensure the effectiveness of ensemble learning as well as to speed up the learning process of co-training. To the best of our knowledge, the present study is the first instance of multi-view feature learning being introduced into a co-training algorithm. The results of the experiments, conducted on study areas in the United States, confirm the effectiveness and superiority of MFDC, with some limitations on its practicality. Future improvements include the utilization of a more effective graph construction method and labelled sample augmentation in MVSSF free of complex parameter tuning, as well as further verification of the applicability of the algorithm to other complex land cover types.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive comments and the high-performance computing support from the Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University (Available online: <https://gda.bnu.edu.cn/>).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the second Tibetan Plateau Scientific Expedition and Research Program (STEP) under Grant [2019QZKK0608] and National Natural Science Foundation of China under Grant [41901316].

References

- Akar, Ö., and O. Güngör. 2015. "Integrating Multiple Texture Methods and NDVI to the Random Forest Classification Algorithm to Detect Tea and Hazelnut Plantation Areas in Northeast Turkey." *International Journal of Remote Sensing* 36 (2): 442–464. doi:10.1080/01431161.2014.995276.
- Arévalo, P., P. Olofsson, and C. E. Woodcock. 2020. "Continuous Monitoring of Land Change Activities and Post-disturbance Dynamics from Landsat Time Series: A Test Methodology for REDD+ Reporting." *Remote Sensing of Environment* 238: 111051. doi:10.1016/j.rse.2019.01.013.
- Baraldi, A., L. Bruzzone, and P. Blonda. 2005. "Quality Assessment of Classification and Cluster Maps without Ground Truth Knowledge." *IEEE Transactions on Geoscience and Remote Sensing* 43 (4): 857–873. doi:10.1109/TGRS.2004.843074.
- Blum, A., and T. Mitchell. 1998. "Combining Labeled and Unlabeled Data with Co-training." Paper presented at the Conference on Computational Learning Theory, Madison, Wisconsin, USA.
- Bruzzone, L., and C. Persello. 2009. "Active Learning for Classification of Remote Sensing Images." Paper presented at the 2009 IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009.
- Clark, M. L., T. Mitchell Aide, H. Ricardo Grau, and G. Riner. 2010. "A Scalable Approach to Mapping Annual Land Cover at 250 M Using MODIS Time Series Data: A Case Study in the Dry Chaco

- Ecoregion of South America." *Remote Sensing of Environment* 114 (11): 2816–2832. doi:[10.1016/j.rse.2010.07.001](https://doi.org/10.1016/j.rse.2010.07.001).
- Davranche, A., G. Lefebvre, and B. Poulin. 2010. "Wetland Monitoring Using Classification Trees and SPOT-5 Seasonal Time Series." *Remote Sensing of Environment* 114 (3): 552–562. doi:[10.1016/j.rse.2009.10.009](https://doi.org/10.1016/j.rse.2009.10.009).
- De Alban, D., J. M. Grant Connette, P. Oswald, and L. Edward Webb. 2018. "Combined Landsat and L-Band SAR Data Improves Land Cover Classification and Change Detection in Dynamic Tropical Landscapes." *Remote Sensing* 10 (2): 306. doi:[10.3390/rs10020306](https://doi.org/10.3390/rs10020306).
- Di, W., and M. M. Crawford. 2011. "Active Learning via Multi-View and Local Proximity Co-Regularization for Hyperspectral Image Classification." *IEEE Journal of Selected Topics in Signal Processing* 5 (3): 618–628. doi:[10.1109/JSTSP.2011.2123077](https://doi.org/10.1109/JSTSP.2011.2123077).
- Dópido, I., J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas Dias, and J. A. Benediktsson. 2013. "Semisupervised Self-Learning for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 51 (7): 4032–4044. doi:[10.1109/TGRS.2012.2228275](https://doi.org/10.1109/TGRS.2012.2228275).
- Dronova, I., P. Gong, L. Wang, and L. Zhong. 2015. "Mapping Dynamic Cover Types in a Large Seasonally Flooded Wetland Using Extended Principal Component Analysis and Object-based Classification." *Remote Sensing of Environment* 158: 193–206. doi:[10.1016/j.rse.2014.10.027](https://doi.org/10.1016/j.rse.2014.10.027).
- Estel, S., T. Kuemmerle, C. Alcántara, C. Levers, A. Prishchepov, and P. Hostert. 2015. "Mapping Farmland Abandonment and Recultivation across Europe Using MODIS NDVI Time Series." *Remote Sensing of Environment* 163: 312–325. doi:[10.1016/j.rse.2015.03.028](https://doi.org/10.1016/j.rse.2015.03.028).
- Feng, G. J., M. M. Schwaller, and F. Hall. 2006. "On the Blending of the Landsat and MODIS Surface Reflectance: Predicting Daily Landsat Surface Reflectance." *IEEE Transactions on Geoscience and Remote Sensing* 44 (8): 2207–2218. doi:[10.1109/TGRS.2006.872081](https://doi.org/10.1109/TGRS.2006.872081).
- Fisher, A., N. Flood, and T. Danaher. 2016. "Comparing Landsat Water Index Methods for Automated Water Classification in Eastern Australia." *Remote Sensing of Environment* 175: 167–182. doi:[10.1016/j.rse.2015.12.055](https://doi.org/10.1016/j.rse.2015.12.055).
- Flood, N. 2014. "Continuity of Reflectance Data between Landsat-7 ETM+ and Landsat-8 OLI, for Both Top-of-Atmosphere and Surface Reflectance: A Study in the Australian Landscape." *Remote Sensing* 6 (9): 7952–7970. doi:[10.3390/rs6097952](https://doi.org/10.3390/rs6097952).
- Forkuor, G., C. Conrad, J. Michael Thiel, B. Zoungrana, and E. Jérôme Tondoh. 2017. "Multiscale Remote Sensing to Map the Spatial Distribution and Extent of Cropland in the Sudanian Savanna of West Africa." *Remote Sensing* 9 (8): 523. doi:[10.3390/rs9080839](https://doi.org/10.3390/rs9080839).
- Gebhardt, S., T. Wehrmann, M. A. M. Ruiz, P. Maeda, J. Bishop, M. Schramm, R. Kopeinig, et al. 2014. "MAD-MEX: Automatic Wall-to-Wall Land Cover Monitoring for the Mexican REDD-MRV Program Using All Landsat Data." *Remote Sensing* 6 (5): 3923. doi:[10.3390/rs6053923](https://doi.org/10.3390/rs6053923).
- Gómez, C., J. C. White, and M. A. Wulder. 2016. "Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 116: 55–72. doi:[10.1016/j.isprsjprs.2016.03.008](https://doi.org/10.1016/j.isprsjprs.2016.03.008).
- Griffiths, P., C. Nendel, and P. Hostert. 2019. "Intra-annual Reflectance Composites from Sentinel-2 and Landsat for National-scale Crop and Land Cover Mapping." *Remote Sensing of Environment* 220: 135–151. doi:[10.1016/j.rse.2018.10.031](https://doi.org/10.1016/j.rse.2018.10.031).
- Gu, S., and Y. Jin. 2017. "Multi-train: A Semi-supervised Heterogeneous Ensemble Classifier." *Neurocomputing* 249: 202–211. doi:[10.1016/j.neucom.2017.03.063](https://doi.org/10.1016/j.neucom.2017.03.063).
- Hackman, K. O., P. Gong, and J. Wang. 2017. "New Land-cover Maps of Ghana for 2015 Using Landsat 8 and Three Popular Classifiers for Biodiversity Assessment." *International Journal of Remote Sensing* 38 (14): 4008–4021. doi:[10.1080/01431161.2017.1312619](https://doi.org/10.1080/01431161.2017.1312619).
- Hao, P., Y. Zhan, L. Wang, Z. Niu, and M. Shakir. 2015. "Feature Selection of Time Series MODIS Data for Early Crop Classification Using Random Forest: A Case Study in Kansas, USA." *Remote Sensing* 7 (5): 5347–5369. doi:[10.3390/rs70505347](https://doi.org/10.3390/rs70505347).
- Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics SMC-3*, no. 6: 610–621. doi:[10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314).

- Henits, L., C. Jürgens, and L. Mucsi. 2016. "Seasonal Multitemporal Land-cover Classification and Change Detection Analysis of Bochum, Germany, Using Multitemporal Landsat TM Data." *International Journal of Remote Sensing* 37 (15): 3439–3454. doi:10.1080/01431161.2015.1125558.
- Hu, T., X. Huang, L. Jiayi, and L. Zhang. 2018. "A Novel Co-training Approach for Urban Land Cover Mapping with Unclear Landsat Time Series Imagery." *Remote Sensing of Environment* 217: 144–157. doi:10.1016/j.rse.2018.08.017.
- Hughes, G. 1968. "On the Mean Accuracy of Statistical Pattern Recognizers." *IEEE Transactions on Information Theory* 14 (1): 55–63. doi:10.1109/TIT.1968.1054102.
- Interdonato, R., D. Ienco, R. Gaetano, and K. Ose. 2019. "DuPLO: A DUal View Point Deep Learning Architecture for Time Series Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 149: 91–104. doi:10.1016/j.isprs.2019.01.011.
- Jia, K., S. Liang, N. Zhang, X. Wei, G. Xingfa, X. Zhao, Y. Yao, and X. Xie. 2014. "Land Cover Classification of Finer Resolution Remote Sensing Data Integrating Temporal Features from Time Series Coarser Resolution Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 93: 49–55. doi:10.1016/j.isprs.2014.04.004.
- Jönsson, P., Z. Cai, A. Eli Melaas, M. Friedl, and L. Eklundh. 2018. "A Method for Robust Estimation of Vegetation Seasonality from Landsat and Sentinel-2 Time Series Data." *Remote Sensing* 10 (4): 635. doi:10.3390/rs10040635.
- Kim, Y., N.-W. Park, and K.-D. Lee. 2017. "Self-Learning Based Land-Cover Classification Using Sequential Class Patterns from past Land-Cover Maps." *Remote Sensing* 9 (9): 921. doi:10.3390/rs9090921.
- Kussul, N., M. Lavreniuk, S. Skakun, and A. Shelestov. 2017. "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data." *IEEE Geoscience and Remote Sensing Letters* 14 (5): 778–782. doi:10.1109/LGRS.2017.2681128.
- Li, P., L. Jiang, and Z. Feng. 2014. "Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors." *Remote Sensing* 6 (1): 310–329. doi:10.3390/rs6010310.
- Ma, X., H. Wang, and J. Wang. 2016. "Semisupervised Classification for Hyperspectral Image Based on Multi-decision Labeling and Deep Feature Learning." *ISPRS Journal of Photogrammetry and Remote Sensing* 120: 99–107. doi:10.1016/j.isprs.2016.09.001.
- Ma, Z., F. Nie, Y. Yang, J. R. R. Uijlings, N. Sebe, and A. G. Hauptmann. 2012. "Discriminating Joint Feature Analysis for Multimedia Data Understanding." *IEEE Transactions on Multimedia* 14 (6): 1662–1672.
- Masek, J. G., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, G. Feng, J. Kutler, and L. Teng-Kui. 2006. "A Landsat Surface Reflectance Dataset for North America, 1990–2000." *IEEE Geoscience and Remote Sensing Letters* 3 (1): 68–72. doi:10.1109/LGRS.2005.857030.
- Müller, H., P. Rufin, P. Griffiths, A. J. B. Siqueira, and P. Hostert. 2015. "Mining Dense Landsat Time Series for Separating Cropland and Pasture in a Heterogeneous Brazilian Savanna Landscape." *Remote Sensing of Environment* 156: 490–499. doi:10.1016/j.rse.2014.10.014.
- Nigam, K., and R. Ghani. 2000. "Analyzing the Effectiveness and Applicability of Co-Training." Paper presented at the International Conference on Information & Knowledge Management, McLean, Virginia, USA.
- Pelletier, C. I., G. Webb, and F. Petitjean. 2019. "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series." *Remote Sensing* 11 (5): 523. doi:10.3390/rs11050523.
- Petitjean, F., J. Inglada, and P. Gancarski. 2012. "Satellite Image Time Series Analysis Under Time Warping." *IEEE Transactions on Geoscience and Remote Sensing* 50 (8): 3081–3095. doi:10.1109/TGRS.2011.2179050.
- Platt, J. 2000. "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods." *Advances in Large Margin Classifiers* 10: 61–74.
- Prishchepov, A. V., V. C. Radeloff, M. Dubinin, and C. Alcantara. 2012. "The Effect of Landsat ETM/ETM + Image Acquisition Dates on the Detection of Agricultural Land Abandonment in Eastern Europe." *Remote Sensing of Environment* 126: 195–209. doi:10.1016/j.rse.2012.08.017.

- Rodriguez-Galiano, V. F., M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan. 2012. "Random Forest Classification of Mediterranean Land Cover Using Multi-seasonal Imagery and Multi-seasonal Texture." *Remote Sensing of Environment* 121: 93–107. doi:10.1016/j.rse.2011.12.003.
- Salehi, B., B. Daneshfar, and A. M. Davidson. 2017. "Accurate Crop-type Classification Using Multi-temporal Optical and Multi-polarization SAR Data in an Object-based Image Analysis Framework." *International Journal of Remote Sensing* 38 (14): 4130–4155. doi:10.1080/01431161.2017.1317933.
- Senf, C., P. J. Leitão, D. Pflugmacher, S. van der Linden, and P. Hostert. 2015. "Mapping Land Cover in Complex Mediterranean Landscapes Using Landsat: Improved Classification Accuracies from Integrating Multi-seasonal and Synthetic Imagery." *Remote Sensing of Environment* 156: 527–536. doi:10.1016/j.rse.2014.10.018.
- Sexton, J. O., D. L. Urban, M. J. Donohue, and C. Song. 2013. "Long-term Land Cover Dynamics by Multi-temporal Classification across the Landsat-5 Record." *Remote Sensing of Environment* 128: 246–258. doi:10.1016/j.rse.2012.10.010.
- Shahshahani, B. M., and D. A. Landgrebe. 1994. "The Effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon." *IEEE Transactions on Geoscience and Remote Sensing* 32 (5): 1087–1095. doi:10.1109/36.312897.
- Shao, Y., R. S. Lunetta, B. Wheeler, J. S. Iliames, and J. B. Campbell. 2016. "An Evaluation of Time-series Smoothing Algorithms for Land-cover Classifications Using MODIS-NDVI Multi-temporal Data." *Remote Sensing of Environment* 174: 258–265. doi:10.1016/j.rse.2015.12.023.
- Suess, S., S. van der Linden, A. Okujeni, P. Griffiths, P. J. Leitão, M. Schwieder, and P. Hostert. 2018. "Characterizing 32 Years of Shrub Cover Dynamics in Southern Portugal Using Annual Landsat Composites and Machine Learning Regression Modeling." *Remote Sensing of Environment* 219: 353–364. doi:10.1016/j.rse.2018.10.004.
- Sun, C., Y. Liu, S. Zhao, M. Zhou, Y. Yang, and L. Feixue. 2016. "Classification Mapping and Species Identification of Salt Marshes Based on a Short-time Interval NDVI Time-series from HJ-1 Optical Imagery." *International Journal of Applied Earth Observation and Geoinformation* 45: 27–41. doi:10.1016/j.jag.2015.10.008.
- Tan, K., H. Jun, L. Jun, and D. Peijun. 2015. "A Novel Semi-supervised Hyperspectral Image Classification Approach Based on Spatial Neighborhood Information and Classifier Combination." *ISPRS Journal of Photogrammetry and Remote Sensing* 105: 19–29. doi:10.1016/j.isprsjprs.2015.03.006.
- Tan, K., E. Li, Q. Du, and P. Du. 2014. "An Efficient Semi-supervised Classification Approach for Hyperspectral Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 97: 36–45. doi:10.1016/j.isprsjprs.2014.08.003.
- Tan, K., J. Zhu, Q. Du, L. Wu, and D. Peijun. 2016. "A Novel Tri-Training Technique for Semi-Supervised Classification of Hyperspectral Images Based on Diversity Measurement." *Remote Sensing* 8 (9): 749. doi:10.3390/rs8090749.
- Verbesselt, J., R. Hyndman, A. Zeileis, and D. Culvenor. 2010. "Phenological Change Detection while Accounting for Abrupt and Gradual Trends in Satellite Image Time Series." *Remote Sensing of Environment* 114 (12): 2970–2980. doi:10.1016/j.rse.2010.08.003.
- Vermote, E., C. Justice, M. Claverie, and B. Franch. 2016. "Preliminary Analysis of the Performance of the Landsat 8/OLI Land Surface Reflectance Product." *Remote Sensing of Environment* 185: 46–56. doi:10.1016/j.rse.2016.04.008.
- Wardlow, B. D., and S. L. Egbert. 2008. "Large-area Crop Mapping Using Time-series MODIS 250 M NDVI Data: An Assessment for the U.S. Central Great Plains." *Remote Sensing of Environment* 112 (3): 1096–1116. doi:10.1016/j.rse.2007.07.019.
- Xue, Z., P. Du, J. Li, and H. Su. 2017. "Sparse Graph Regularization for Hyperspectral Remote Sensing Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 55 (4): 2351–2366. doi:10.1109/TGRS.2016.2641985.
- Yan, E., G. Wang, H. Lin, C. Xia, and H. Sun. 2015. "Phenology-based Classification of Vegetation Cover Types in Northeast China Using MODIS NDVI and EVI Time Series." *International Journal of Remote Sensing* 36 (2): 489–512. doi:10.1080/01431161.2014.999167.

- Yan, L., and D. P. Roy. 2014. "Automated Crop Field Extraction from Multi-temporal Web Enabled Landsat Data." *Remote Sensing of Environment* 144: 42–64. doi:[10.1016/j.rse.2014.01.006](https://doi.org/10.1016/j.rse.2014.01.006).
- Yan, L., and D. P. Roy. 2015. "Improved Time Series Land Cover Classification by Missing-observation-adaptive Nonlinear Dimensionality Reduction." *Remote Sensing of Environment* 158: 478–491. doi:[10.1016/j.rse.2014.11.024](https://doi.org/10.1016/j.rse.2014.11.024).
- Yu, B., and S. Shang. 2017. "Multi-Year Mapping of Maize and Sunflower in Hetao Irrigation District of China with High Spatial and Temporal Resolution Vegetation Index Series." *Remote Sensing* 9 (8): 855. doi:[10.3390/rs9080855](https://doi.org/10.3390/rs9080855).
- Zhi-Hua, Z., and L. Ming. 2005. "Tri-training: Exploiting Unlabeled Data Using Three Classifiers." *IEEE Transactions on Knowledge and Data Engineering* 17 (11): 1529–1541. doi:[10.1109/TKDE.2005.186](https://doi.org/10.1109/TKDE.2005.186).
- Zhong, L., H. Lina, and H. Zhou. 2019. "Deep Learning Based Multi-temporal Crop Classification." *Remote Sensing of Environment* 221: 430–443. doi:[10.1016/j.rse.2018.11.032](https://doi.org/10.1016/j.rse.2018.11.032).
- Zhu, X., and D. Liu. 2014. "Accurate Mapping of Forest Types Using Dense Seasonal Landsat Time-series." *ISPRS Journal of Photogrammetry and Remote Sensing* 96: 1–11. doi:[10.1016/j.isprsjprs.2014.06.012](https://doi.org/10.1016/j.isprsjprs.2014.06.012).
- Zhu, Z., and C. E. Woodcock. 2014. "Automated Cloud, Cloud Shadow, and Snow Detection in Multitemporal Landsat Data: An Algorithm Designed Specifically for Monitoring Land Cover Change." *Remote Sensing of Environment* 152: 217–234. doi:[10.1016/j.rse.2014.06.012](https://doi.org/10.1016/j.rse.2014.06.012).