INTERNATIONAL JOURNAL OF REMOTE SENSING
2020, VOL. 41, NO. 5, 1668–1691
https://doi.org/10.1080/01431161.2019.1674458

Taylor & Francis
Taylor & Francis Group

Multitemporal ensemble learning for snow cover extraction from high-spatial-resolution images in mountain areas

Pengfeng Xiao[a,b,c], Chengxi Li[a], Luijun Zhu[d], Xueliang Zhang[a,b,c], Tengyao Ma[a] and Xuezhi Feng[a,b,c]

[a]Department of Geographic Information Science, School of Geography and Ocean Science, Nanjing University, Nanjing, Jiangsu, China; [b]Collaborative Innovation Center of South China Sea Studies, Nanjing, Jiangsu, China; [c]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, Jiangsu, China; [d]Department of Civil Engineering, Monash University, Clayton, Australia

ABSTRACT

High-spatial and -temporal resolution snow cover products in mountain areas are important to hydrological applications. The GF-1 satellite provides multispectral images with 8-m resolution and a revisit up to 2 days, which makes it possible to produce snow cover products. However, it is challenging to extract snow cover from these images because of limited spectral bands, severe mountain shadows, and dataset-shift problem in multitemporal classification. To overcome the limitations above, this study proposes a multitemporal ensemble learning framework to extract snow cover from high-spatial-resolution images in mountain areas. The principle behind ensemble learning, i.e. learning from disagreement, is extended from single image classification to multitemporal ones. We assume that multitemporal training samples selected within time-invariant classes at the same locations can be different in feature space. Such disagreements are used in multitemporal ensemble learning to improve classification accuracy. To enhance both accuracy and diversity of the multiple classifiers trained on these samples, a joint feature selection method is suggested to select the optimal multitemporal feature space and a joint parameter optimization method is designed to ensemble classifiers trained for multitemporal images. The experiments show that the performances of multitemporal ensemble classifiers are superior to that of single classifiers, confirming the effectiveness of the proposed framework.

ARTICLE HISTORY
Received 4 December 2018
Accepted 29 July 2019

1. Introduction

Accurately acquiring multitemporal snow cover has great significance for snowmelt runoff modelling and snow disaster monitoring in mountain areas (Crawford et al. 2013; Czyzowska-Wisniewski et al. 2015; Mityók et al. 2018). Until now, several satellite sensors have been used to extract snow cover, e.g. Advanced Very High Resolution Radiometer (AVHRR) (Hüsler et al. 2014), Moderate Resolution Imaging Spectroradiometer (MODIS)

CONTACT Pengfeng Xiao xiaopf@nju.edu.cn Department of Geographic Information Science, School of Geography and Ocean Science, Nanjing University, Nanjing, Jiangsu 210023, China

(Hall et al. 2002), Satellite Pour l'Observation de la Terre (SPOT) Vegetation (VGT) (Dankers and Jong 2004), Landsat Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), Operational Landsat Imager (OLI) (Crawford et al. 2013), Sentinel-2 (Zhu, Wang, and Woodcock 2015), and ZY-3 Charge Couple Device (CCD) (Zhu et al. 2014). As shown in Figure 1, most of these sensors cannot provide snow cover observation with both high-spatial and -temporal resolution, which is essential to snowmelt runoff prediction in mountain areas. Chinese GaoFen-1 (GF-1) satellite carries two Panchromatic and Multispectral Sensors (PMSs), which provide panchromatic images with 2-m resolution and multispectral images with 8-m resolution. The three-satellite constellation (GF-1, GF-2, and GF-6) provides at most a revisit of 2 days, making it possible to extract snow cover with both high-spatial and -temporal resolution.

The high-spatial resolution makes GF-1 images contain detailed land surface information, but also results in two challenges: limited spectral bands and severe mountain shadows. Like most of the high-spatial-resolution satellites, the GF-1 images have only four spectral bands: blue, green, red, and near-infrared band. As the lack of shortwave-infrared band, the most effective snow mapping feature, i.e. Normalised Difference Snow Index (NDSI) (Dozier 1989), cannot be calculated accordingly. Some alternatives based on visible bands have been proposed (Hinkler et al. 2002; Hinkler, Ørbæk, and Hansen 2003). However, they still cannot achieve satisfied extraction of snow cover in mountain areas because of the large overlap in spectral space between snow in shadow and snow-free pixels (Rosenthal and Dozier 1996). Topographic correction is an effective strategy to alleviate the influence of mountain shadows (Crawford et al. 2013; Negi, Kulkarni, and Semwal 2009; Sirguey, Mathieu, and Arnaud 2009), but the lack of Digital Elevation Model (DEM) data with the same high-spatial resolution makes the correction quite difficult. The two challenges have been preliminarily investigated in our previous study (Zhu et al. 2014), where a feature selection algorithm was integrated to find effective features in case

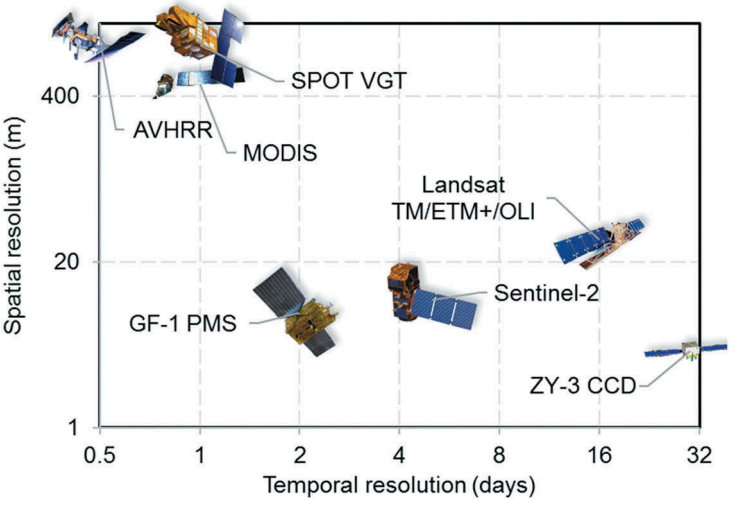

Figure 1. Spatial and temporal resolution of the main satellite sensors used to extract snow cover. Only GF-1 PMS has high-spatial and -temporal resolution simultaneously.

of lacking shortwave-infrared band and snow in shadow was treated as an independent class to alleviate the influence of mountain shadows.

The high-temporal resolution makes GF-1 images possible to extract time-series snow cover for snowmelt runoff modelling, but multitemporal classification is widely acknowledged to suffer from dataset-shift problem (Moreno-Torres et al. 2012). Specifically, snow cover acquired at different phases has different spectral characteristics thus a classifier trained based on samples selected on an image cannot be applied directly to others. This problem can be alleviated by separately selecting adequate training samples for each image, but it is time-consuming and laborious. Hence, an approach that can address this problem and extract snow cover from multitemporal images in a more efficient way is needed. Until now, some techniques have been reported, such as transfer learning (Weiss, Khoshgoftaar, and Wang 2016), active learning (Tuia, Pasolli, and Emery 2011), and multitask learning (Leiva-Murillo, Gomez-Chova, and Camps-Valls 2013). Our previous study (Zhu et al. 2016) tried to reduce the heavy dependence on the number of training samples using a mutual-learning strategy, but it can only extract snow cover from dual-temporal images. Therefore, an effective approach for simultaneously extracting snow cover from multitemporal images is needed.

Ensemble learning (Dietterich 2002) may provide a different view for solving the dataset-shift problem of multitemporal classification. In ensemble learning, different learners are built for the same classification task, with the difference coming from different training samples (Breiman 2001), feature spaces (Piao et al. 2015), or label spaces (Tsoumakas, Katakis, and Vlahavas 2011). Ensemble learning has attracted intensive interest in the remote sensing community and achieved satisfied results for single-temporal image classification (Belgiu and Drăguţ 2016; Rodriguez-Galiano et al. 2012; Samat et al. 2014). The key to the success of ensemble learning is that the learners should be accurate and diverse to effectively construct an ensemble (Sagi and Rokach 2018), while it is not critical where the difference comes from. For the multitemporal images, the spectral shift of snow cover among different phases can be a source of diversity. Accordingly, independent classifiers trained for different phases is expected to be improved using ensemble learning.

The motivation of this study is to extend the classic ensemble learning based on a single image to multitemporal ensemble learning based on multitemporal images as shown in Figure 2. The main difference between the two methods is that the classic ensemble learning constructs multiple learners for a single image and then combines them to build a classifier, with the difference among multiple spectral subsets utilized to improve the accuracy. In contrast, the proposed multitemporal ensemble learning constructs image-specific learners first, followed by a mutual-learning process using the temporal difference. Accordingly, the classic ensemble learning is a snapshot method, while the proposed framework is a multitemporal one without requiring image specific training set and training process. Note that the dataset shift of multitemporal images can be exploited to generate the diversity of learners trained on multiple views (V_1, V_2, …, V_N), which means the dataset-shift problem is an advantage in the proposed framework, rather than a disadvantage in normal multitemporal classification. Originally, multi-view refers to different descriptions of the same object, which is originated from dividing attribute sets on a single dataset (Zhao et al. 2017). Since multitemporal images are different descriptions of the same spatial extent, they can be regarded as multiple

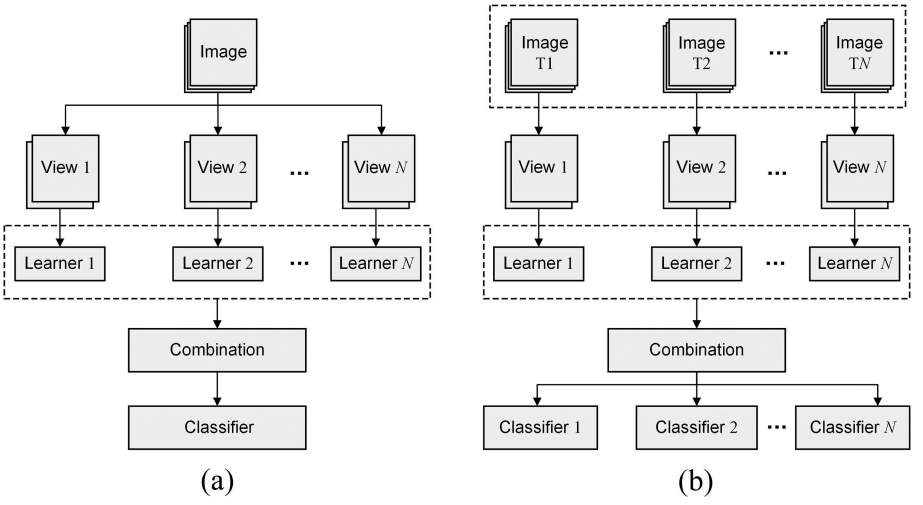

Figure 2. Extension from classic ensemble learning based on a single image (a) to multitemporal ensemble learning based on multitemporal images (b).

views of the earth surface. Thus, through extending the conception of multiple views to multitemporal images, the performance of learners trained on multitemporal images is expected to be improved by ensemble learning. Similar to the classic ensemble learning, the building of classifiers and the way to use the difference are critical to its multitemporal extension. Thus, two issues should be considered in the proposed framework: (1) multitemporal images should be able to build diverse and accurate classifiers; and (2) the difference of classifiers trained for multitemporal images can be utilized.

The main contributions of this study are: (1) extracting time-series snow cover with both high-spatial and -temporal resolutions in mountain areas by ensemble learning to overcome the challenge of dataset-shift problem; (2) extending the classic ensemble learning based on a single image to multitemporal ensemble learning based on multitemporal images by utilizing the spectral shift among different phases; and (3) improving the performance of classifiers for each phase by joint feature selection and joint parameter optimization.

2. Method

2.1. Overview of the proposed framework

Given N images acquired sequentially over the same area, the proposed framework of multitemporal ensemble learning includes three main components: (1) determination of training samples and extraction of features; (2) joint feature selection for multitemporal images; and (3) ensemble of classifiers via joint parameter optimization (Figure 3). In the first component, a location set $\{L(i, j)\}$ is selected where the label of location $L(i, j)$ is consistent in the multitemporal images. N feature vectors are extracted for each location, which are training samples for the classification. These feature vectors may be quite different because of dataset shift in different phases as results of snow evolution, variation of illumination, change of observation geometry, etc. These differences are expected to

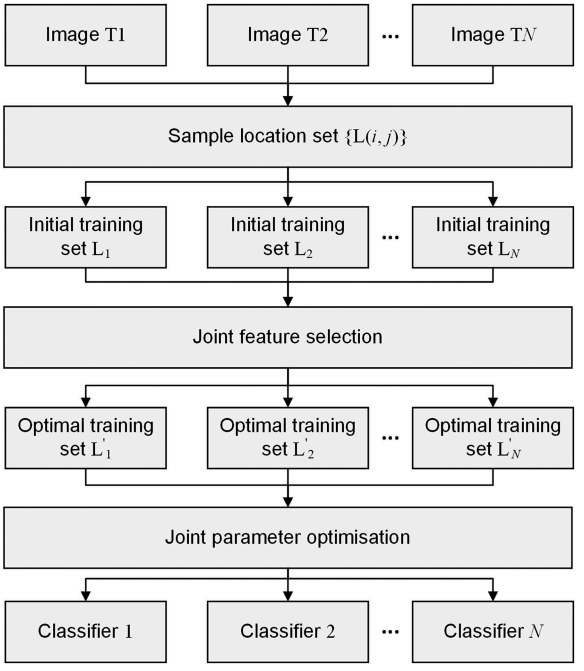

Figure 3. Flowchart of the proposed framework of multitemporal ensemble learning for multitemporal snow cover extraction.

improve the classifiers separately trained on multitemporal images. In the second component, a joint feature selection procedure is designed to find the optimal feature subsets in view of diversity and separability. In the final component, the predication accuracy and difference among *N* classifiers are used to jointly optimize the base classifiers. A brief introduction to the base classifier and the details of the last two components are provided below.

2.2. Base classifier

The support vector machine (SVM) is selected as the base classifier because of its widely acknowledged capability in remotely sensed image classification (Leiva-Murillo, Gomez-Chova, and Camps-Valls 2013). Each classifier contains multiple SVMs to implement multiclass classification in this study. In order to train an SVM, a constrained Quadratic Programming (QP) problem must be solved (Platt 1999):

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y_i y_j k(x_i, x_j) \alpha_i \alpha_j \tag{1}$$

$$subject\ to \sum_{i=1}^{m} y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, \forall i \tag{2}$$

where $k(x_i, x_j)$ is a kernel function which measures the similarity of a stored training sample x_i to the input one x_j; y_i and y_j are class labels; α is the Lagrange multiplier; and C is a penalty parameter. In this study, the Gaussian Radial Basis Function (RBF) kernel is used:

$$k(x_i, x_j) = \exp\left(-\gamma||x_i - x_j||^2\right) \tag{3}$$

where γ represents the width of the kernel. Therefore, two parameters C and γ are required in SVM, which will be jointly optimized by the classifier ensemble later.

Platt's Sequential Minimal Optimization (SMO) algorithm (Platt 1999) is selected to solve the QP problem for its simplicity and speed (Keerthi et al. 2001). The SMO algorithm decomposes the QP problem into sub-problems (Chang, Hsu, and Lin 2000). At every step, only two Lagrange multipliers are chosen to be jointly optimized and this procedure can be done analytically. Therefore, although the number of sub-problems which should be solved is raised, the overall QP problem can be solved very quickly.

2.3. Joint feature selection for multitemporal images

The aim of feature selection is to find feature space for each image, on which classifiers can be trained to be accurate and diverse. Considering the goal of accurately extracting multitemporal snow cover, finding feature space with high prediction accuracy and low redundancy is of the top priority. Accordingly, the feature selection includes two steps: (1) selecting best feature subsets for each image independently with high prediction accuracy and low redundancy; and (2) determining optimal combination considering the effectiveness of each feature subset and the diversity among multiple classifiers. The two steps are described below.

Step 1. Selecting best feature subsets for each image. In order to explore the feature space of each image, the elitist Pareto-based multi-objective evolutionary algorithm (Jiménez et al. 2015) is applied. The algorithm incorporates the statistical criteria to select several feature subsets to maximize classification accuracy as well as to minimize subset cardinality. Specifically, feature selection of each image begins with initializing population P and evaluating all the K individuals. For each generation, a pair of parents is produced by a binary tournament selection from P. Then, returns the best between two random individuals according to a rank-crowding-better function. An individual I is considered better than an individual J if the rank of I is lower than the rank of J in the population. The rank of an individual I refers to the non-domination level of I among P. Thus, slot(I) = slot (J), where the radial slot is the portion of the search space an individual belongs to, and it is defined as:

$$\text{slot}(I) = \sum_{j=1}^{n-1} d^{j-1} \left\lfloor d \frac{\alpha_j^I}{\pi/2} \right\rfloor \tag{4}$$

$$\alpha_j^I = \begin{cases} \pi/2 & \text{if } h_{j+1}^I = 0 \\ \arctan\left(\frac{h_{j+1}^I}{h_j^I}\right) & \text{if } h_{j+1}^I \neq 0 \end{cases} \tag{5}$$

where $d = \left\lfloor \sqrt[n-1]{N} \right\rfloor$ and h_j^I refers to the objective function f_j^I normalized in [0, 1].

Then, the selected pair of parents is crossed, mutated, repaired, evaluated, and eventually added to Q, an initially empty auxiliary population. This process is running until there exists M individuals in Q. Next, an auxiliary population R is obtained through the union of P and Q, where M best individuals can survive to next generation. After T iterations, M best individuals $F_1^i, F_2^i, \ldots, F_M^i (i = 1, 2, \ldots, N)$ for each image are obtained.

Step 2. Determining optimal combination of feature subsets. The optimal combination is obtained from the selected M best feature subsets for each image through a joint metric (C) considering both prediction accuracy and classifier diversity. The full search is used here considering the limited number of available feature subsets ($N \times M$). The combination with the highest C value is selected, which is the optimal feature subset for an image. The joint metric is defined as:

$$C = \lambda \frac{\sum_{i=1}^{N} E_s}{N} + (1 - \lambda) \frac{2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} A_{i,j}}{N(N-1)} \tag{6}$$

where λ is the combination weight; E_s refers to the heuristic merit of a feature subset S; N refers to the number of images; and $A_{i,j}$ refers to the disagreement measure between the ith and jth SVMs.

The heuristic merit E_s is calculated based on the Correlation-based Feature Selection (CFS) (Hall 1999), which is a filter algorithm that ranks feature subsets according to a heuristic evaluation based on correlation. The feature subsets are evaluated considering the individual prediction ability of each feature and the redundancy between them. The evaluation function of the heuristic merit for a feature subset S is:

$$E_s = \frac{k\overline{r_{fc}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{7}$$

where $\overline{r_{fc}}$ refers to the average feature-class correlation; $\overline{r_{ff}}$ refers to the average feature-feature inter-correlation; and k is the number of features. Thus, the numerator of the merit provides an indication of how predictive of the class a feature subset is, while the denominator is an indication of how much redundancy there is among the features.

The disagreement measure $A_{i,j}$ is to assess the difference between pairwise classifiers built on two images, which can be calculated as (Kuncheva and Whitaker 2003):

$$A_{i,j} = \frac{N_{rw} + N_{wr}}{N_s} \tag{8}$$

where N_{rw} is the number of samples for which the first classifier is right while the second classifier is wrong; N_{wr} is the number of samples for which the first classifier is wrong while the second classifier is right; and N_s is the number of training samples.

2.4. Ensemble of classifiers via joint parameter optimization

After feature selection, N initial SVMs can be trained on the optimal feature spaces with default parameters of C and γ. These initial classifiers can then learn from each other using the prediction difference on the training set. The genetic algorithm (GA) (Huang and Wang 2006) is used to integrate the classifiers through jointly optimizing C and γ of all SVMs.

The first step of GA is chromosome design and population initialization. In the chromosome design, a binary coding system is used to encode all C and γ with each parameter being a 16-bit binary value. In the population initialization, 20 chromosomes are randomly generated. These chromosomes are then adaptively optimized using three genetic operations: selection, crossover, and mutation.

The selection operation is used to pick good chromosomes from the current population according to the fitness function (Huang and Wang 2006):

$$F = \frac{N_u}{N_t} \tag{9}$$

where N_u is the number of pixels which keep unchanged on multiple images and N_t is the total number of pixels on multiple images. This is similar to the decision strategy of 'winner takes all'. One classifier can learn from the rest if it has inconsistent prediction with the majority.

The crossover and mutation operation explore the new regions of search space while retaining some current information (Stein et al. 2005). The crossover operator can exchange genes between two chromosomes and the mutation operator is used to improve the genetic diversity by randomly modifying genes (Engelbrecht 2006). The optimization process is terminated when the number of iterations reaches a predefined value.

3. Study area and dataset

3.1. Study area

The study area is located in Eren Habirga of middle Tianshan Mountains, Xinjiang Province, China, with a total area 1259 km^2, as shown in Figure 4. This area is in the alpine region with its elevation ranging from 2438 to 4419 m. The northwest of the area is covered with perennial snow. There is 31.4% of this area with a slope greater than 30°. Thus, the influence of mountain shadows is severe.

3.2. Dataset

The multitemporal dataset is composed of five GF-1 PMS images acquired on 3, 7, 15, 19, and 23 October 2013, named as T1, T2, T3, T4, and T5, respectively, with an image size of 4418 × 4371 pixels. The interval between every two adjacent images is 4 days, except that the interval between T2 and T3 is 8 days. This is because there was a snowfall between 7 and 15 October, thus the image on 11 October was cloudy and was not selected as the experiment data. The GF-1 PMS multispectral image consists of four bands, i.e. blue, green, red, and near-infrared (NIR), with 8-m spatial resolution. In addition, radiometric correction was implemented on each image based on the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) (Cooley et al. 2002). Subsequently, spatial registration was adopted on the five images and the Root Mean Square Error (RMSE) was controlled within 0.5 pixel.

The five images have a little offset because of the shift of satellite orbit. As the training procedure requires the multitemporal images cover the same area, a subset with 1000 × 1000 pixels in the northwest of the image was selected as the focus area to carry out an

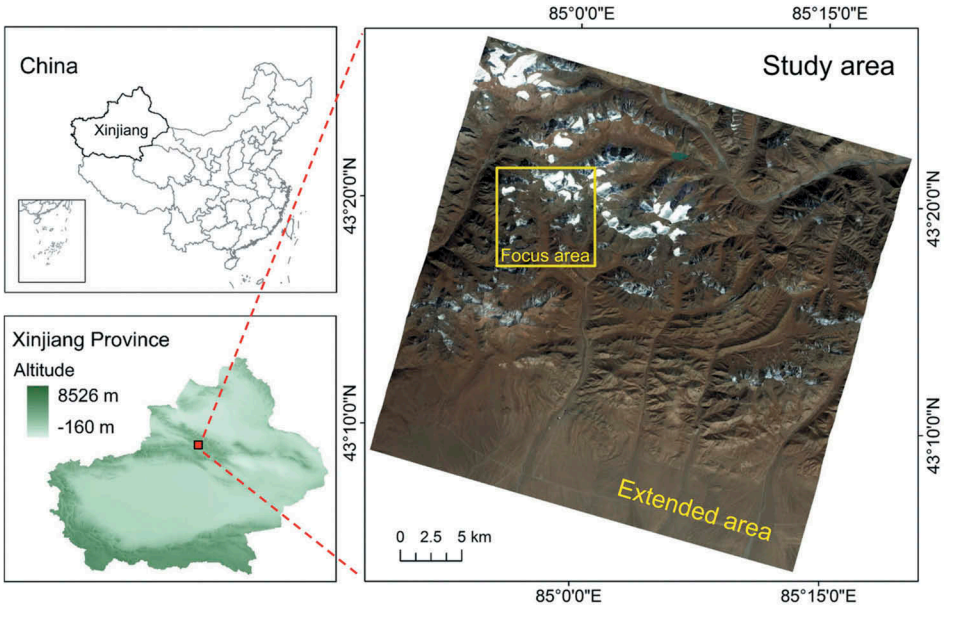

Figure 4. Illustration of the study area in Eren Habirga of middle Tianshan Mountains, Xinjiang Province, China. The study area includes a focus area with 1000 × 1000 pixels and an extended area with 4418 × 4371 pixels. The GF-1 PMS image was acquired on 3 October 2013, as shown with false colour composite (red: near-infrared band; green: red band; blue: green band).

experiment at first. Then, the trained multiple classifiers of the focus area were extended to the whole images to further validate the robustness of the proposed framework, as shown in Figure 4.

We calculated 13 candidate features for successive feature selection, including 4 spectral features, 6 index features, and 3 transformed features, as shown in Table 1. The spectral features are the four bands of GF-1 PMS image (B1, B2, B3, and B4). The index features consist of two snow indices, i.e. GF-1 Snow Index (GFSI) (Jiang et al. 2015) and Normalized Difference Snow and Ice Index (NDSII) (Hinkler et al. 2002), three vegetation indices, i.e. Normalized Difference Vegetation Index (NDVI), Ratio Vegetation Index (RVI), and Difference Vegetation Index (DVI), and one water index, i.e. Normalized Difference Water Index (NDWI). The transformed features are the first, second, and third principal components (PC1, PC2, and PC3). All the features were normalized to [0, 1] to facilitate the feature selection and ensemble learning.

Table 1. Descriptions of the 13 extracted candidate features.

Type	Number	Name	Description
Spectral features	4	B1, B2, B3, B4	Spectral bands
Index features	6	GFSI	$(R_{Blue}-R_{NIR})/(R_{Blue}+R_{NIR})$
		NDSII	$(R_{Red}-R_{NIR})/(R_{Red}+R_{NIR})$
		NDVI	$(R_{NIR}-R_{Red})/(R_{NIR}+R_{Red})$
		RVI	R_{NIR}/R_{Red}
		DVI	$R_{NIR}-R_{Red}$
		NDWI	$(R_{Green}-R_{NIR})/(R_{Green}+R_{NIR})$
Transformed features	3	PC1, PC2, PC3	Principal components

R represents the reflectance value of the band shown as a subscript.

4. Experiment and results

4.1. Experiment setup

The proposed framework is programmed using Java language. In the joint feature selection, we set the number of images $N = 5$, the number of iterations $I = 10$, the number of individuals in population $K = 100$, and the influence of combination weight $\lambda = 0.5$. In the joint parameter optimization, we set maximum generation 100, population size 20, crossover rate 0.25, and mutation rate 0.01. Besides, we set $n_C = 20$, $n_\gamma = 20$, $C \in (0.1, 1000)$, and $\gamma \in (0.001, 1)$. The termination criterion is that the generation number reaches 100.

In this study, snow in sunlight and snow in shadow are regarded as two different classes to alleviate the severe influence of mountain shadows. Hence, the classification scheme includes three classes: snow in sunlight, snow in shadow, and snow-free. We randomly selected 50 pixels for each class to form the training sample sets of the five focus images through visual interpretation, noting that all the training samples have consistent class labels on the multiple images. We also randomly selected 3000 pixels per class for each focus image to form the validation sample sets, noting that the validation samples were selected from each focus image independently. It is obvious that the number of the training samples is extremely smaller than that of the validation samples, which will illustrate the robustness of the proposed framework.

4.2. Evaluation criteria

The performance of feature selection was evaluated in terms of three metrics: (1) the number of features for each selected feature subset; (2) the disagreement measure of each pairwise classifiers built on multiple images; and (3) the Jeffreys-Matusita (J-M) distance between snow and snow-free based on validation samples. The J-M distance is a commonly used measure to assess the degree of interclass separability, which is defined as (Zheng, Yuan, and Lu 2017):

$$D_{i,j} = \sqrt{2(1 - \mathrm{e}^{-\beta})} \tag{10}$$

$$\beta = \frac{1}{8}(\mu_i - \mu_j)^{\mathrm{T}} \left(\frac{\mathbf{C}_i + \mathbf{C}_j}{2}\right)^{-1} (\mu_i - \mu_j) + \frac{1}{2}\ln\left(\frac{\frac{1}{2}|\mathbf{C}_i + \mathbf{C}_j|}{\sqrt{|\mathbf{C}_i||\mathbf{C}_j|}}\right) \tag{11}$$

where μ_i and μ_j are the mean of class i and j, respectively; and $\mathbf{C}_i$ and $\mathbf{C}_j$ are the covariance matrix of class i and j, respectively.

After feature selection and parameter optimization, the performance of snow cover extraction was evaluated using the validation sample set. It is noted that during the classification procedure, snow in shadow and snow in sunlight were regarded as two different classes. After that, these two classes were merged into one class, i.e. snow. The evaluation metrics used here are precision, recall, and F-score, which are defined as (Rittger, Painter, and Dozier 2013):

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{12}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{13}$$

$$F = \frac{2PR}{P + R} \tag{14}$$

where TP (true positive) is the number of snow pixels which are identified correctly; FP (false positive) is the number of snow-free pixels which are identified as snow; FN (false negative) is the number of snow pixels which are identified as snow-free; P (precision) is the proportion of correctly identified snow pixels against identified snow pixels; R (recall) is the proportion of correctly identified snow pixels against true snow pixels; and F (F-score) is a balance of P and R, which is one of the most useful metrics to evaluate the performance of snow cover extraction.

The tradeoff between the effectiveness of each feature subset and the diversity among multiple images is essential to construct an effective ensemble, which is of great importance to the joint parameter optimization. In this study, the combination weight λ was used to achieve this tradeoff. Hence, the influence of the combination weight λ was also evaluated. In addition, the influence of the number of phases input the proposed framework was also investigated.

4.3. Effectiveness of the joint feature selection

Five feature subsets were obtained after the joint feature selection, as shown in Table 2. The number of features included in the feature spaces of the five images are 7, 7, 8, 5, and 7, occupied 46.1%, 46.1%, 38.4%, 61.6%, and 46.1% of the initial feature spaces, respectively. In addition, the selected five feature subsets include all the three feature categories, indicating that index features and transformed features are important supplements to the original spectral features. Specifically, each selected feature subset contains B1, GFSI, and PC2, indicating that these three features are essential to distinguish snow and snow-free, and other features may provide a diversity of classifiers trained on multiple images.

Table 3 shows the J-M distance between snow (including snow in sunlight and snow in shadow) and snow-free before and after feature selection. Not surprisingly, the J-M distance between snow in shadow and snow-free is smaller than that between snow in sunlight and snow-free before feature selection, confirming the difficulty of discriminating snow in shadow from snow-free, because of the large overlap in spectral space between snow in shadow and snow-free. After feature selection, the J-M distance between snow and snow-free was increased, especially between snow in shadow and snow-free, indicating the effectiveness of including the index features and transformed features. This shows that after feature selection, the separability of snow and snow-free for multiple images was enlarged.

The disagreement of each pairwise classifiers trained on pairwise images is shown in Figure 5. After joint feature selection, the inconsistencies between different image combinations have generally improved. Specifically, the inconsistency between different image combinations reaches a maximum of 0.31, a minimum of 0.04, and an average of 0.17. The large increases are all related to T3, e.g. T2–T3, T1–T3, and T3–

Table 2. Descriptions of the selected feature subsets for the five images on the focus area.

Image	T1	T2	T3	T4	T5
Feature	B1	B1	B1	B1	B1
	B2	B2	B2	B3	B3
	B4	B3	B3	B4	B4
	GFSI	GFSI	B4	GFSI	GFSI
	RVI	NDWI	GFSI	PC2	DVI
	PC1	PC1	NDWI		NDWI
	PC2	PC2	PC1		PC2
			PC2		
Number	7	7	8	5	7

Table 3. J-M distances between snow and snow-free before and after feature selection.

Image	Status	Snow in sunlight and snow-free	Snow in shadow and snow-free
T1	Before	1.38	1.37
	After	1.41	1.41
T2	Before	1.38	1.35
	After	1.41	1.41
T3	Before	1.39	1.39
	After	1.41	1.40
T4	Before	1.38	1.36
	After	1.40	1.39
T5	Before	1.37	1.35
	After	1.41	1.41

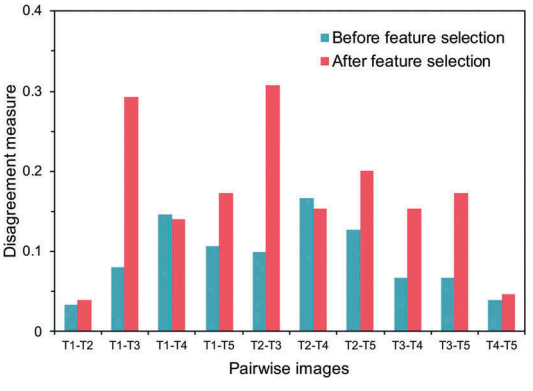

Figure 5. Disagreement of each pairwise classifiers trained on pairwise images before and after feature selection.

T5. A possible explanation is that a snowfall occurred between T2 and T3 and thus snow cover in T3 had quite different spectral characteristics from the rest, which provides more space for building diversities. This is also confirmed by the small increase observed at the combinations of T1–T2 and T4–T5, where snow cover was relatively stable and consequently it was difficult to enhance the diversity using feature selection. Despite the increased inconsistency in most cases, an opposite change was observed for T1–T4 and T2–T4, which is mainly caused by the tradeoff between accuracy and separability. In a word, on the condition that the separability of snow and snow-free was enlarged, the diversity of classifiers trained on multiple images was increased after feature selection.

4.4. Influence of the combination weight

The selected feature subsets for each image should compromise between the distinguishing ability of snow with snow-free and the diversity of classifiers trained on multiple images. In order to evaluate the influence of the combination weight λ of prediction accuracy and classifier diversity, experiments on the five images with different λ were designed. First, the average J-M distances between snow (including snow in sunlight and snow in shadow) and snow-free of the five images were calculated. Then, the average feature numbers of the five selected feature subsets were analysed. Finally, the average disagreement measures of each pairwise classifiers trained on multiple images were calculated.

Figure 6 shows the results with different combination weights λ. We can see that (1) the average J-M distances between snow and snow-free increase with the increasing of λ, (2) the average feature numbers of the selected feature subsets increase on the whole when λ increases, and (3) the average disagreement measures of each pairwise classifiers decrease on the whole when λ increases. This is because as λ increases, the weight of the prediction accuracy increases and the weight of the diversity of features decreases. Specifically, all the three measures keep stable when λ is between 0.3 and 0.7. Hence, we can indicate that the effectiveness of the joint feature selection can be satisfied if λ is not set extremely high or low (above 0.7 or below 0.3).

4.5. Performance of the multitemporal ensemble learning

The classification experiment was repeated 10 folds to evaluate the performance of the proposed framework. All of the precision, recall, and *F*-score of snow are satisfied in the classification results. The average *F*-scores for T1–T5 are 0.98, 0.97, 0.98, 0.98, and 0.98, respectively, with a standard deviation of less than 0.01. Figure 7 shows the snow cover extraction results for the five images on the focus area. It is shown less false alarms after ensemble learning than before ensemble learning. The snow cover area of T3 has a large increase compared to T2, which indicates that a snowfall occurred between T2 and T3.

The time consumption of the proposed framework of the five images on the focus area is 59 min 23 s, testing by a computer with Intel core i7-4940MX processor and 32 GB memory. The time consumption comprises four parts, i.e. selecting the best feature subsets, determining optimal feature combination, joint parameter optimization, and image classification, as shown in Table 4. The joint parameter optimization processing consumes most of the time (51 min 7 s), which can be considered as the time difference between ensemble learning method and non-ensemble learning method.

4.6. Comparison of the multitemporal ensemble learning to non-ensemble learning

In order to evaluate the effectiveness of the multitemporal ensemble learning, five SVM classifiers with default parameters settings ($C_1 = C_2 = C_3 = 1, \gamma_1 = \gamma_2 = \gamma_3 = 0.01$) were used as the benchmark of non-ensemble learning for comparison. The average precision, recall, and *F*-score of snow of the 10-fold classification are shown in Figure 8. It indicates that the accuracy increases significantly after ensemble learning. Specifically, the average precision increases, which means the positive prediction of snow increases significantly

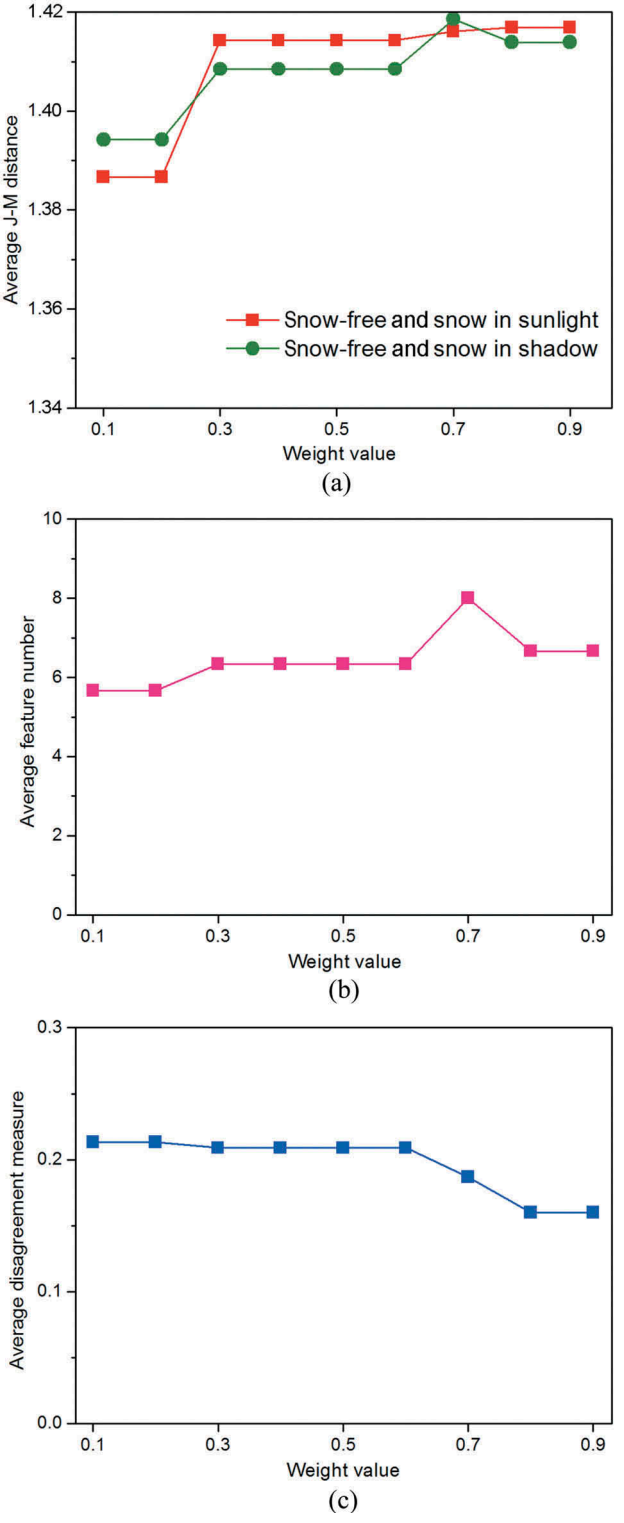

Figure 6. Influence of the combination weight λ on the five feature subsets. (a) is the average J-M distance between snow (including snow in sunlight and snow in shadow) and snow-free, (b) is the average feature number, and (c) is the average disagreement measure of each pairwise classifiers.

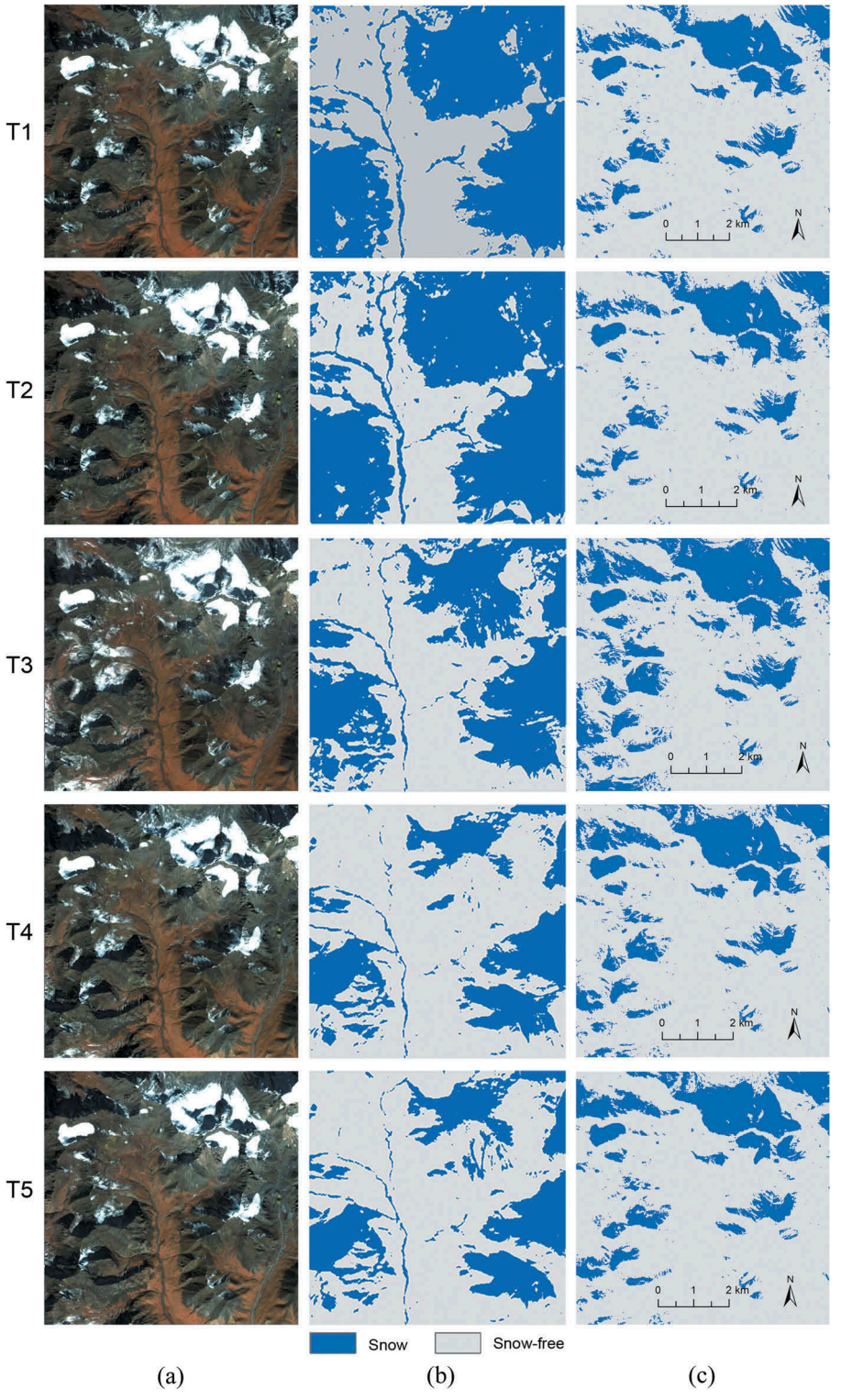

Figure 7. Five images on the focus area with false colour composite (red: near-infrared band; green: red band; blue: green band) (a) and their results of snow cover extraction before (b) and after (c) ensemble learning.

Table 4. Time consumption of the proposed framework of the five images on the focus area.

Processing	Selecting best feature subsets	Determining optimal feature combination	Joint parameter optimization	Image classification	Total
Time	1 s	7 min 34 s	51 min 7 s	41 s	59 min 23 s

after ensemble learning. However, the recall decreases slightly after ensemble learning, showing a decrease in detection rate of snow. As a balance of precision and recall, the average *F*-scores for T1–T5 increase 24%, 26%, 21%, 20%, and 20%, respectively. Specifically, before ensemble learning, the accuracies of T3, T4, and T5 are higher than those of T1 and T2, which is related to the snowfall between T2 and T3. The appearance of fresh snow makes the snow cover extraction more accurate because of their higher reflectance at visible and NIR bands. However, after ensemble learning, the accuracies of all the images reach a very high level. This shows that ensemble learning can achieve satisfied accuracy for both fresh snow and aged snow. Consequently, we can conclude that the five SVM classifiers performed better after the ensemble, which means the proposed framework can effectively improve the performance of multitemporal snow cover extraction.

4.7. *Influence of the number of image phases*

The number of image phases affects the results of joint feature selection and multitemporal snow cover extraction based on ensemble learning. Because two classifiers cannot guarantee the effectiveness of parameter optimization, the minimum number of classifiers should be three. Therefore, we set the number of image phases as 3, 4, and 5, then analyse its influence on the accuracy of snow cover extraction. The average *F*-sores are shown in Figure 9. With the increase of the number of image phases, the accuracy of multitemporal snow cover extraction increases gradually. This means the classifiers based on joint parameter optimization are more robust in the case of more phases. The reason is that the dataset-shift problem is an advantage rather than a disadvantage in the multitemporal ensemble learning. The snow cover change of multitemporal images can be exploited to generate the diversity of learners trained on multiple images. Nevertheless, in the case of three phases, the average *F*-score of the three images (T1–T3) also reaches 0.90, indicating that the snow cover extraction reaches a high accuracy even when the number of phases is small.

The change of the average *F*-scores with different number of image phases in Figure 9 can be analysed combined with Figure 8. As shown in Figure 8(c), the *F*-scores of the single SVM classifiers (before ensemble learning) in T1 and T2 are worse than those in T3, T4, and T5, which indicates that the feature quality of T1 and T2 is lower than that of T3, T4, and T5. When T1–T3 is ensembled, although the feature quality of T3 is high, but the feature quality of T1 and T2 is low, thereby affecting the classification accuracy. When T4 is added, the classification accuracy is improved significantly. When T5 is added, the low feature quality of T1 and T2 cannot influence the accuracy evidently; thus, the classification accuracy is further improved.

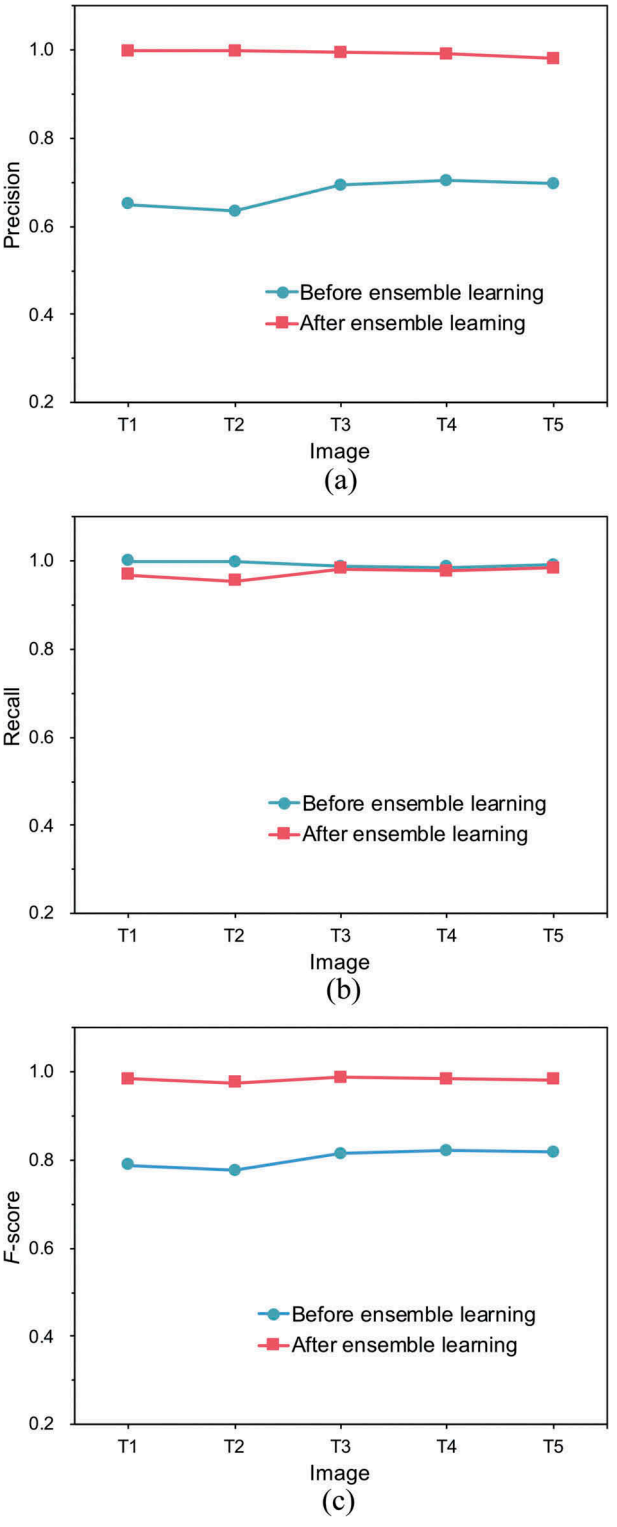

Figure 8. Accuracy comparison of snow cover extraction of the five single SVM classifiers (before ensemble learning) and optimized SVM classifiers (after ensemble learning) of the five images on the focus area. (a), (b), and (c) refer to the average precision, recall, and *F*-score of snow of the 10-fold classification, respectively.

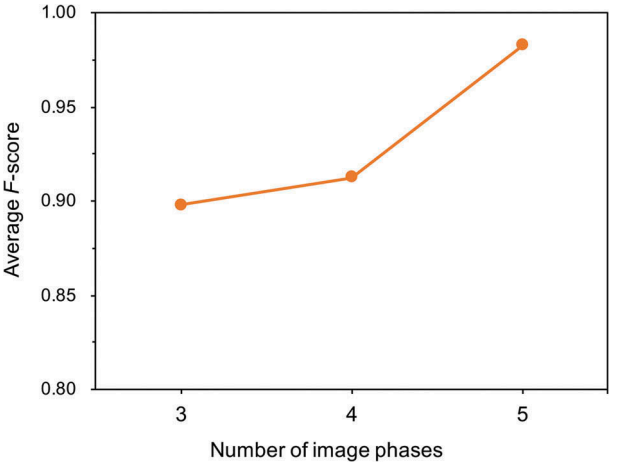

Figure 9. Average *F*-scores of snow cover extraction of three (T1–T3), four (T1–T4), and five (T1–T5) phases of images on the focus area.

4.8. Expansion from the focus area to the extended area

In order to further validate the robustness of the proposed framework, the classification experiment was implemented on the extended area based on the trained classifiers and parameters of the focus area. The validation sample set consists of 2000 pixels randomly selected from the extended area, including 500 pixels of snow in sunlight, 500 pixels of snow in shadow, and 1000 pixels of snow-free, as shown in Figure 10(a). It is noted that the validation samples were selected from each image independently.

Figure 10(b,c) is the result of the snow cover extraction for the five images on the extended area before and after multitemporal ensemble learning, respectively. They are shown that the false alarms are substantially reduced after ensemble learning. Table 5 presents the accuracies of snow cover extraction using five single SVM classifiers (before ensemble learning) and optimized SVM classifiers (after ensemble learning) for the five extended images. The precisions of snow are increased significantly after ensemble learning in all cases, coincident with the decreased false alarms observed in Figure 10(c). Most of the recalls of snow retain 1.00 before and after ensemble learning, showing quite low omission errors. The *F*-scores for T1–T5 after ensemble learning are 0.95, 0.93, 0.94, 0.97, and 0.97, respectively, which illustrates the robustness of the proposed framework expanded from the focus area to the extended area.

5. Discussions

In order to effectively extract time-series snow cover from both high-spatial and -temporal resolution satellite images, the multitemporal ensemble learning is proposed to address the issue of high-temporal resolution, and the spectral, index, and transformed features are combined to alleviate the problem of limited spectral bands of high-spatial-resolution images. Then, the tradeoff between the prediction accuracy and the classifier diversity of multiple images is achieved after feature selection. Moreover, the dataset-shift

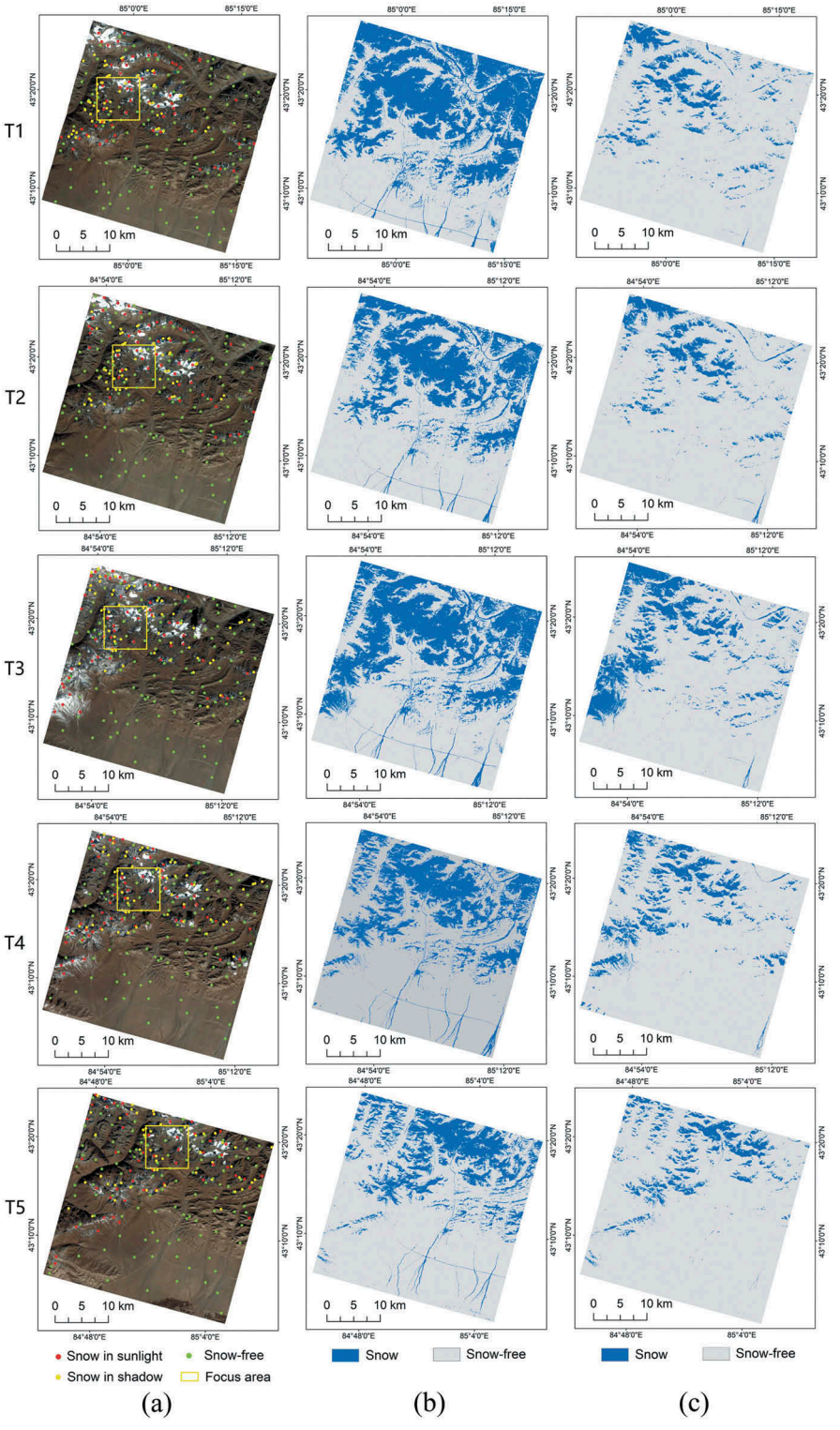

Figure 10. Five images on the extend area with false colour composite (red: near-infrared band; green: red band; blue: green band) (a) and their results of snow cover extraction before (b) and after (c) ensemble learning. The location of the focus area and the validation samples are shown in (a).

Table 5. Accuracy comparisons of snow cover extraction using five single SVM classifiers (before ensemble learning) and optimized SVM classifiers (after ensemble learning) of the five images on the extended area.

Image	Status	Precision	Recall	*F*-score
T1	Before	0.72	1.00	0.84
	After	0.91	0.99	0.95
T2	Before	0.76	1.00	0.87
	After	0.92	0.94	0.93
T3	Before	0.73	1.00	0.85
	After	0.90	0.99	0.94
T4	Before	0.75	1.00	0.86
	After	0.96	0.99	0.97
T5	Before	0.75	0.95	0.84
	After	0.99	0.95	0.97

phenomenon of multitemporal images is exploited to generate the diversity of classifiers trained on multiple images, which is essential to construct an efficient ensemble. Hence, the dataset-shift problem is an advantage rather than a disadvantage in the proposed multitemporal ensemble learning framework. As an ensemble is constructed based on the multiple classifiers, the performance of the classifiers can be improved through ensemble learning. Specifically, ensemble learning can simultaneously optimize the parameters of SVM classifiers for multitemporal images.

According to the experimental results, the accuracy of snow cover extraction from multitemporal GF-1 images is high using only one set of training samples. Considering the reasonable accuracy level and the simplicity of the proposed framework, the method is efficient to extract snow cover from multitemporal images. Specifically, the proposed framework mainly consists of two sequential steps: joint feature selection and joint parameter optimization. The effectiveness of these two steps are comprehensively evaluated. After the joint feature selection, the distinguish ability of snow and snow-free is enlarged and the feature spaces are compressed while the diversity of multiple images is retained. After the joint parameter optimization, the accuracy of snow cover extraction for multitemporal images is increased significantly. It is noted that the number of training samples (50 pixels per class) is extremely smaller than that of the validation samples (3000 pixels per class). This means that the proposed framework is robust when the training sample size is small. In addition, although only five images are used in the experiment, the proposed framework can be conveniently applied to multitemporal images with more phases.

This study mainly contributes to the construction of the multitemporal ensemble learning framework for snow cover extraction. Using the proposed framework, the time-series snow cover in multitemporal images can be extracted synchronously with high accuracy only by selecting training samples once. Though some strategies were proposed to solve the dataset-shift problem, e.g. active learning (Tuia, Pasolli, and Emery 2011) and multitask learning (Leiva-Murillo, Gomez-Chova, and Camps-Valls 2013), they cannot complete the image classification synchronously by using one training sample set. Thus, the accuracies of snow extraction based on the proposed framework are compared with that based on the classifiers with uniform parameters, and there is no additional comparison with other methods in the experiment.

It is well known that collecting adequate training samples is time-consuming and laborious, especially in mountain areas. The semi-supervised method, which combines labelled and unlabelled data to boost the performance of a classifier (Zhu, Ghahramani, and Lafferty 2003), is promising to address this problem. The combination of ensemble learning and semi-supervised learning has been proved beneficial to each other (Li and Zhou 2007). Consequently, to make the extraction of multitemporal snow cover more effective, the combination of the proposed framework and semi-supervised algorithm is suggested as a promising strategy in the future. It can facilitate the operational use of high-resolution satellites to produce snow cover products with both high-spatial and -temporal resolutions.

In addition, how to discriminate snow and cloud is a classic issue for snow cover extraction. The reflectance of cloud is similar to snow in visible and near-infrared bands, but usually higher than snow in shortwave-infrared band. Hence, when the remote sensing image with the shortwave-infrared band is utilized for extracting snow cover, the discrimination between snow and cloud would be easy. Unfortunately, most high-spatial-resolution remote sensing images lack the shortwave-infrared band, which makes it a challenging problem to discriminate snow and cloud. However, the satellite images generally provide quality control information, i.e., the fraction of cloud is known. Thus, the uncertainty is under control though the cloud effects snow cover extraction. Since the focus of this study is to extend the classic ensemble learning to multitemporal ensemble learning, the effect of the cloud for remote sensing images without shortwave-infrared band is not analysed quantitatively and remains to be explored in the future.

6. Conclusions

This study proposes a multitemporal ensemble learning framework for snow cover extraction from both high-spatial and -temporal resolution satellite images in mountain areas. The time-series snow cover in multitemporal images can be extracted synchronously with high accuracy only by selecting training samples once. The original ensemble learning model is extended to multitemporal images to improve the performance of multitemporal snow cover extraction. Specifically, a joint feature selection method and a joint parameter optimization method are proposed based on the multitemporal ensemble learning model. The experiments show that the performances of multitemporal ensemble classifiers are superior to that of single classifiers, which confirms the effectiveness of the proposed framework.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China [Grant Nos. 41871235, 41671344, and 41601366], the Natural Science Foundation of Jiangsu Province [Grant No. BK20160623], and the Fundamental Research Funds for the Central Universities [Grant No. 020914380068].

ORCID

Pengfeng Xiao http://orcid.org/0000-0003-2739-3302

References

Belgiu, M., and L. Drăguţ. 2016. "Random Forest in Remote Sensing: A Review of Applications and Future Directions." *ISPRS Journal of Photogrammetry and Remote Sensing* 114: 24–31. doi:10.1016/j.isprsjprs.2016.01.011.

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

Chang, -C.-C., C.-W. Hsu, and C.-J. Lin. 2000. "The Analysis of Decomposition Methods for Support Vector Machines." *IEEE Transactions on Neural Networks* 11: 1003–1008. doi:10.1109/72.857780.

Cooley, T., G. P. Anderson, G. W. Felde, M. L. Hoke, A. J. Ratkowski, J. H. Chetwynd, J. A. Gardner, et al., 2002. "FLAASH, a MODTRAN4-based Atmospheric Correction Algorithm, Its Application and Validation". IEEE International Geoscience and Remote Sensing Symposium, Toronto, Ontario, Canada, 3, pp. 1414–1418.

Crawford, C. J., S. M. Manson, M. E. Bauer, and D. K. Hall. 2013. "Multitemporal Snow Cover Mapping in Mountainous Terrain for Landsat Climate Data Record Development." *Remote Sensing of Environment* 135: 224–233. doi:10.1016/j.rse.2013.04.004.

Czyzowska-Wisniewski, E. H., W. J. D. van Leeuwen, K. K. Hirschboeck, S. E. Marsh, and W. T. Wisniewski. 2015. "Fractional Snow Cover Estimation in Complex Alpine-forested Environments Using an Artificial Neural Network." *Remote Sensing of Environment* 156: 403–417. doi:10.1016/j.rse.2014.09.026.

Dankers, R., and S. M. D. Jong. 2004. "Monitoring Snow-cover Dynamics in Northern Fennoscandia with SPOT VEGETATION Images." *International Journal of Remote Sensing* 25: 2933–2949. doi:10.1080/01431160310001618374.

Dietterich, T. G. 2002. "Ensemble Learning." In *The Handbook of Brain Theory and Neural Networks*, edited by M. Arbib, 405–408. Cambridge, MA: MIT Press.

Dozier, J. 1989. "Spectral Signature of Alpine Snow Cover from the Landsat Thematic Mapper." *Remote Sensing of Environment* 28: 9–22. doi:10.1016/0034-4257(89)90101-6.

Engelbrecht, A. P. 2006. *Fundamentals of Computational Swarm Intelligence*. Hoboken, NJ: John Wiley & Sons.

Hall, D. K., G. A. Riggs, V. V. Salomonson, N. E. DiGirolamo, and K. J. Bayr. 2002. "MODIS Snow-cover Products." *Remote Sensing of Environment* 83: 181–194. doi:10.1016/S0034-4257(02)00095-0.

Hall, M. A. 1999. *Correlation-based Feature Selection for Machine Learning*. PhD Thesis, University of Waikato, Hamilton, New Zealand.

Hinkler, J., J. B. Ørbæk, and B. U. Hansen. 2003. "Detection of Spatial, Temporal, and Spectral Surface Changes in the Ny-ålesund Area 79° N, Svalbard, Using a Low Cost Multispectral Camera in Combination with Spectroradiometer Measurements." *Physics and Chemistry of the Earth, Parts A/B/C* 28: 1229–1239. doi:10.1016/j.pce.2003.08.059.

Hinkler, J., S. B. Pedersen, M. Rasch, and B. U. Hansen. 2002. "Automatic Snow Cover Monitoring at High Temporal and Spatial Resolution, Using Images Taken by a Standard Digital Camera." *International Journal of Remote Sensing* 23: 4669–4682. doi:10.1080/01431160110113881.

Huang, C.-L., and C.-J. Wang. 2006. "A GA-based Feature Selection and Parameters Optimization for Support Vector Machines." *Expert Systems with Applications* 31: 231–240. doi:10.1016/j.eswa.2005.09.024.

Hüsler, F., T. Jonas, M. Riffler, J. P. Musial, and S. Wunderle. 2014. "A Satellite-based Snow Cover Climatology (1985–2011) for the European Alps Derived from AVHRR Data." *The Cryosphere* 8: 73–90. doi:10.5194/tc-8-73-2014.

Jiang, L., P. Xiao, X. Feng, N. Ye, G. He, and X. Zhang. 2015. "Calculation of Snow Reflectance from GF-1 Satellite Image in Rugged Mountain Areas." *Journal of Nanjing University Natural Science* 51: 944–954.

Jiménez, F., E. Marzano, G. Sánchez, G. Sciavicco, and N. Vitacolonna, 2015. "Attribute Selection via Multi-Objective Evolutionary Computation Applied to Multi-Skill Contact Center Data Classification". 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, pp. 488–495. doi:10.1109/SSCI.2015.78

Keerthi, S. S., S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy. 2001. "Improvements to Platt's SMO Algorithm for SVM Classifier Design." *Neural Computation* 13: 637–649. doi:10.1162/089976601300014493.

Kuncheva, L. I., and C. J. Whitaker. 2003. "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy." *Machine Learning* 51: 181–207. doi:10.1023/A:1022859003006.

Leiva-Murillo, J. M., L. Gomez-Chova, and G. Camps-Valls. 2013. "Multitask Remote Sensing Data Classification." *IEEE Transactions on Geoscience and Remote Sensing* 51: 151–161. doi:10.1109/TGRS.2012.2200043.

Li, M., and Z. H. Zhou. 2007. "Improve Computer-Aided Diagnosis With Machine Learning Techniques Using Undiagnosed Samples." *IEEE Transactions on Systems, Man, and Cybernetics - Part A* 37: 1088–1098. doi:10.1109/TSMCA.2007.904745.

Mityók, Z. K., D. K. Bolton, N. C. Coops, E. E. Berman, and S. Senger. 2018. "Snow Cover Mapped Daily at 30 Meters Resolution Using a Fusion of Multi-temporal MODIS NDSI Data and Landsat Surface Reflectance." *Canadian Journal of Remote Sensing* 44: 413–434. doi:10.1080/07038992.2018.1538775.

Moreno-Torres, J. G., T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera. 2012. "A Unifying View on Dataset Shift in Classification." *Pattern Recognition* 45: 521–530. doi:10.1016/j.patcog.2011.06.019.

Negi, H. S., A. V. Kulkarni, and B. S. Semwal. 2009. "Estimation of Snow Cover Distribution in Beas Basin, Indian Himalaya Using Satellite Data and Ground Measurements." *Journal of Earth System Science* 118: 525. doi:10.1007/s12040-009-0039-0.

Piao, Y., M. Piao, C. H. Jin, H. S. Shon, J.-M. Chung, B. Hwang, and K. H. Ryu. 2015. "A New Ensemble Method with Feature Space Partitioning for High-Dimensional Data Classification." *Mathematical Problems in Engineering* 2015: 1–12. doi:10.1155/2015/590678.

Platt, J. 1999. "Fast Training of Support Vector Machines Using Sequential Minimal Optimization." In *Advances in Kernel Methods - Support Vector Learning*, edited by B. Schölkopf, C. J. C. Burges, and A. J. Smola, 185–208. Cambridge, MA, USA: MIT Press.

Rittger, K., T. H. Painter, and J. Dozier. 2013. "Assessment of Methods for Mapping Snow Cover from MODIS." *Advances in Water Resources* 51: 367–380. doi:10.1016/j.advwatres.2012.03.002.

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. "An Assessment of the Effectiveness of a Random Forest Classifier for Land-cover Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93–104. doi:10.1016/j.isprsjprs.2011.11.002.

Rosenthal, W., and J. Dozier. 1996. "Automated Mapping of Montane Snow Cover at Subpixel Resolution from the Landsat Thematic Mapper." *Water Resources Research* 32: 115–130. doi:10.1029/95WR02718.

Sagi, O., and L. Rokach. 2018. "Ensemble Learning: A Survey." *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery* 8: e1249. doi: 10.1002/widm.1249.

Samat, A., P. Du, S. Liu, J. Li, and L. Cheng. 2014. "E2LMs: Ensemble Extreme Learning Machines for Hyperspectral Image Classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7: 1060–1069. doi:10.1109/JSTARS.4609443.

Sirguey, P., R. Mathieu, and Y. Arnaud. 2009. "Subpixel Monitoring of the Seasonal Snow Cover with MODIS at 250 M Spatial Resolution in the Southern Alps of New Zealand: Methodology and Accuracy Assessment." *Remote Sensing of Environment* 113: 160–181. doi:10.1016/j.rse.2008.09.008.

Stein, G., B. Chen, A. S. Wu, and K. A. Hua, 2005. "Decision Tree Classifier for Network Intrusion Detection with GA-based Feature Selection". Proceedings of the 43rd Annual Southeast Regional Conference - Volume 2, ACM-SE 43. ACM, New York, NY, USA, pp. 136–141.

Tsoumakas, G., I. Katakis, and I. Vlahavas. 2011. "Random k-Labelsets for Multilabel Classification." *IEEE Transactions on Knowledge and Data Engineering* 23: 1079–1089. doi:10.1109/TKDE.2010.164.

Tuia, D., E. Pasolli, and W. J. Emery. 2011. "Using Active Learning to Adapt Remote Sensing Image Classifiers." *Remote Sensing of Environment* 115: 2232–2242. doi:10.1016/j.rse.2011.04.022.

Weiss, K., T. M. Khoshgoftaar, and D. Wang. 2016. "A Survey of Transfer Learning." *Journal of Big Data* 3: 9. doi:10.1186/s40537-016-0043-6.

Zhao, J., X. Xie, X. Xu, and S. Sun. 2017. "Multi-view Learning Overview: Recent Progress and New Challenges." *Information Fusion* 38: 43–54. doi:10.1016/j.inffus.2017.02.007.

Zheng, X., Y. Yuan, and X. Lu. 2017. "Dimensionality Reduction by Spatial–Spectral Preservation in Selected Bands." *IEEE Transactions on Geoscience and Remote Sensing* 55: 5185–5197. doi:10.1109/TGRS.2017.2703598.

Zhu, L., P. Xiao, X. Feng, X. Zhang, Y. Huang, and C. Li. 2016. "A Co-training, Mutual Learning Approach Towards Mapping Snow Cover from Multi-temporal High-spatial Resolution Satellite Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 122: 179–191. doi:10.1016/j.isprsjprs.2016.11.003.

Zhu, L., P. Xiao, X. Feng, X. Zhang, Z. Wang, and L. Jiang. 2014. "Support Vector Machine-based Decision Tree for Snow Cover Extraction in Mountain Areas Using High Spatial Resolution Remote Sensing Image." *Journal of Applied Remote Sensing* 8: 084698. doi:10.1117/1.JRS.8.084698.

Zhu, X., Z. Ghahramani, and J. Lafferty, 2003. "Semi-supervised Learning Using Gaussian Fields and Harmonic Functions". Proceedings of the 20th International Conference on Machine Learning, ICML'03, Washington, DC, USA: AAAI Press, pp. 912–919.

Zhu, Z., S. Wang, and C. E. Woodcock. 2015. "Improvement and Expansion of the Fmask Algorithm: Cloud, Cloud Shadow, and Snow Detection for Landsats 4–7, 8, and Sentinel 2 Images." *Remote Sensing of Environment* 159: 269–277. doi:10.1016/j.rse.2014.12.014.