# The edge-preservation multi-classifier relearning framework for the classification of high-resolution remotely sensed imagery

Xiaopeng Han [a], Xin Huang [a,b,*], Jiayi Li [b], Yansheng Li [b], Michael Ying Yang [c], Jianya Gong [b]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China
[b] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, PR China
[c] Department of Earth Observation Science (EOS), University of Twente, Enschede, Netherlands

## ABSTRACT

In recent years, the availability of high-resolution imagery has enabled more detailed observation of the Earth. However, it is imperative to simultaneously achieve accurate interpretation and preserve the spatial details for the classification of such high-resolution data. To this aim, we propose the edge-preservation multi-classifier relearning framework (EMRF). This multi-classifier framework is made up of support vector machine (SVM), random forest (RF), and sparse multinomial logistic regression via variable splitting and augmented Lagrangian (LORSAL) classifiers, considering their complementary characteristics. To better characterize complex scenes of remote sensing images, relearning based on landscape metrics is proposed, which iteratively quantizes both the landscape composition and spatial configuration by the use of the initial classification results. In addition, a novel tri-training strategy is proposed to solve the over-smoothing effect of relearning by means of automatic selection of training samples with low classification certainties, which always distribute in or near the edge areas. Finally, EMRF flexibly combines the strengths of relearning and tri-training via the classification certainties calculated by the probabilistic output of the respective classifiers. It should be noted that, in order to achieve an unbiased evaluation, we assessed the classification accuracy of the proposed framework using both edge and non-edge test samples. The experimental results obtained with four multispectral high-resolution images confirm the efficacy of the proposed framework, in terms of both edge and non-edge accuracy.

© 2018 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

## 1. Introduction

Recently, in pace with the rapid development of space imaging techniques, the spatial resolution of remotely sensed imagery has become increasingly high. On the one hand, high-resolution remotely sensed imagery has led to an increased availability of spatial details and structural information of geospatial objects, which can be attributed to the improved observation capacity. However, on the other hand, the higher spatial resolution does not naturally mean more accurate interpretation of remote sensing data, because the increase of the intra-class variance and decrease of the inter-class variance in the spectral feature space result in mis-classification between spectrally similar classes (Alshehhi et al.,

2017; Ma et al., 2015, 2017; Myint et al., 2011). Per-pixel classification using spectral information alone is generally subject to the salt-and-pepper effect. Taking into account the aforementioned two aspects, there are two important requirements for the classification of high-resolution imagery: (1) enhancement of the class separability; and (2) fine delineation of the detailed and structural information of geospatial objects.

It is widely acknowledged that incorporating geometrical and spatial information can reduce classification uncertainty. Specifically, a large number of studies have addressed spectral-spatial joint feature calculation and classification. The commonly used spatial features include the gray-level co-occurrence matrix (GLCM) (Huang et al., 2014a; Pesaresi et al., 2009; Pu and Landry, 2012), wavelet transform (WT) (Cheng et al., 2015; Myint, 2004; Prabhakar and Geetha, 2017), the pixel shape index (PSI) (Zhang et al., 2006), and morphological profiles (Huang et al., 2016; Pesaresi and Benediktsson, 2002). The spatial feature calculation strategies can be viewed as a form of preprocessing prior to classification, which improve the class separability through the addition

* Corresponding author at: State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China; and School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, PR China.

*E-mail addresses:* huang_whu@163.com, xhuang@whu.edu.cn (X. Huang).

of spatial features. However, on the other hand, classification post-processing (CPP) methods, which improve the classification accuracy by refining the initial result in a more succinct way, have not received much attention. The traditional CPP optimization techniques such as the majority filter (Stuckens et al., 2000; Tan et al., 2015), Markov random fields (MRFs) (Rodriguez-Cuenca et al., 2012; Song et al., 2017), and object-based voting (Huang and Zhang, 2013) have a limited ability to improve the classification accuracy, due to the fact that the class separability in the feature space is not enhanced. Recently, relearning based on the primitive co-occurrence matrix (PCM) (called relearning-PCM) was developed for learning a supervised model from the spatial features extracted from the initial classification (Huang et al., 2014c). Geiss and Taubenbock (2015) then extended the relearning framework into an object-based image analysis framework. In this framework, multiple class-related features are derived from a triplet obtained by multiscale segmentation, and are then fed into the classifier to refine the classification model.

It has been demonstrated that the aforementioned relearning methods have the potential to achieve comparable or better classification results than the state-of-the-art spectral-spatial classification algorithms. Although relearning methods are effective at smoothing the classification results, it is unavoidable that detailed structures and edges are gradually blurred in the iterative process, i.e., the over-smoothing effect in edge regions. In fact, for very high resolution (VHR) data, geographical objects are homogenous clusters of pixels, but in the edge areas between these objects, the many mixed pixels that occur due to the finite ground spatial resolution can result in inefficient shape recognition and misclassification (Gamba et al., 2007). Moreover, spatial features refer to the contextual correlation of neighboring pixels, and more neighborhood information about other land-cover classes is taken into account when calculating spatial features for edge pixels. The spatial feature pattern of edge pixels of an object will differ from that of a homogenous area, which may lead to edge issues, i.e., low classification certainty for edge areas. Consequently, detailed structures and edges can be over-smoothed.

In order to preserve the edge details, Solaiman et al. (1995) introduced edge features into an information fusion method for the CPP of Landsat Thematic Mapper (TM) images. In Smits and Dellepiane (1997) and Zhao et al. (2015), an MRF model with an adaptive neighborhood was proposed, in which edge pixels, detected through edge extraction, benefit the classification process, without blurring the border details for synthetic aperture radar (SAR) data. Gamba et al. (2007) improved urban area mapping by incorporating boundary information into a spatially aware classifier for VHR images. Specifically, the boundary and non-boundary pixels are separately classified using different mapping techniques, according to their respective natures, and the two mapping results are integrated through a decision fusion process. In Rodriguez-Cuenca et al. (2012), a spatial contextual post-classification method based on the imposed directional information was developed for preserving linear objects in multispectral imagery. By summarizing the existing literature, however, it should be recognized that all the above-mentioned methods address the edge issues by taking advantage of an edge detector. As a result, their effectiveness is highly reliant on the edge-detection quality. In addition, few studies have simultaneously focused on increasing class separability and preserving edge details.

Based on the above analysis, we propose the edge-preservation multi-classifier relearning framework (EMRF), aiming to simultaneously improve the classification accuracy and preserve detailed edges and structures. In EMRF, the key points are highlighted from the following aspects.

(1) In order to enhance the separability between spectrally similar classes, and then improve the classification result, relearning by iterative learning is adopted in EMRF. It should be noted that relearning-PCM describes the contextual occurrence of the class labels of pixels in a window or neighborhood by characterizing the spatial content of the labeling space at the pixel level. With regard to object-based relearning (OBR), spatial-hierarchical features are derived using a triplet of hierarchical segmentation and, therefore, this method depicts the spatial correlation of class labels at the object level. In the proposed approach, landscape metrics are embedded into the relearning model, considering that landscape features can effectively quantify the spatial configuration of the labeling space at the class level (McGarigal et al., 2002).

In this context, we deploy relearning features via landscape metrics, which are derived from the initial classification results (referred to as relearning-landscape in the following text), aiming to iteratively quantify both the landscape composition and spatial configuration and refine the classification model. In addition, the traditional relearning based on a single classifier is extended to a multi-classifier system as this can exploit the strengths of the individual classifiers and obtain an enhanced performance (Woźniak et al., 2014). Relearning-landscape has the task of not only smoothing the salt-and-pepper effect, but also further enhancing the class separability.

(2) In order to alleviate the edge issues of relearning, a tri-training method is exploited, which works in a semi-supervised manner (Zhou and Li, 2005). In a multi-classifier tri-training system, some of the classifiers will perform poorly, with low classification certainty, while others will yield complementary and reliable results (Woźniak et al., 2014). Under this circumstance, unlabeled samples are selected for the classifiers with low certainty, since these samples are more informative and useful in refining the classifier (Foody and Mathur, 2006). With the aid of these selected training samples, the classification model can be iteratively optimized, and it is more likely that edge pixels will be accurately identified (Tuia et al., 2011). In order to guarantee the reliability of the unlabeled samples, the following conditions should be met: (1) all three classifiers give consistent decisions; and (2) at least one reliable classifier (i.e., with high classification certainty) exists.

(3) It should be mentioned that tri-training suffers from salt-and-pepper noise in homogenous regions because incorrect labels may be added to the training set during the semi-supervised learning (Tan et al., 2016). Therefore, the proposed EMRF method aims at increasing the class separability by the relearning-landscape technique, while at the same time preserving edge details through the tri-training. Specifically, the classification certainty is used to discriminate between edge and non-edge regions in the image, based on the consideration that the certainty values of non-edge areas are generally higher than those of edge regions. The relearning-landscape module is then used to classify non-edge pixels, considering its ability to reduce the salt-and-pepper effect, and the tri-training module is used to identify the class labels of edge pixels, considering its better ability to achieve reliable results for inter-region borders.

In order to evaluate the proposed framework, four multispectral high-resolution data sets were used. When performing the accuracy assessment, both edge and non-edge test samples were

individually used for testing the classification performance for edge and non-edge regions.

The rest of this paper is organized as follows. Section 2 describes the proposed EMRF method in detail. Section 3 introduces the multisource data sets and parameter settings. Section 4 presents the experimental results. A further discussion about the proposed framework is provided in Section 5. Finally, we draw our conclusions in Section 6.

## 2. Methodology

### 2.1. Relearning-landscape

Relearning has been proven to be an effective CPP method (Geiss and Taubenbock, 2015; Huang et al., 2014c). A supervised classification model is iteratively learned from the spatial features which are derived from the initial classification map, and the classification map is gradually optimized and updated according to the feedback provided by the relearned features.

In this study, to better characterize the complex scenes of remote sensing images, we deployed landscape features as the relearning features for the classification of remotely sensed imagery. Landscape metrics can effectively quantify the spatial structures in terms of both composition and configuration, such as area, edge, shape, and aggregation (Fang et al., 2016; Oort et al., 2004). Eight commonly used landscape metrics were investigated in this study (Table 1). We refer the reader to Li et al. (2011b) and McGarigal et al. (2002) for further details about these landscape metrics.

The flowchart of the proposed relearning-landscape module is shown in Fig. 1. Its calculation is described in the following steps.

**Step1:** *Initialization.* Remotely sensed imagery is fed into a classifier, resulting in the initial classification map.
**Step2:** *Landscape feature calculation.* To calculate the landscape features of a pixel located at $x$ in the classified map, we extract the contextual label information in a window (of size $= w$), whose central pixel is $x$. The landscape feature with metric $m$ of class $i$ for pixel $x$ can then be expressed as $h_i(x, w, m)$ and, therefore, all metrics for class $i$ are denoted as:

$$h_i(x, w) = [h_i(x, w, 1), \ldots, h_i(x, w, m), \ldots, h_i(x, w, M)] \qquad (1)$$

where $M$ is the number of landscape metrics. Examples of landscape features extracted from QuickBird images of a local region of Wuhan, China, are provided in Fig. 2. Subsequently, the landscape features for all the land-cover classes $h(x, w)$ can be written as:

$$h(x, w) = [h_1(x, w), \ldots, h_i(x, w), \ldots, h_n(x, w)] \qquad (2)$$

where $n$ is the number of land-cover classes.

**Step3:** *Relearning.* The landscape features are then used as the input for relearning, and they are iteratively updated according to the current classification result, until the relearning stops. At the same time, the classification model can be gradually optimized according to the feedback provided by the landscape features.

Relearning-landscape can take into account the arrangement and contextual information of the land-cover classes from the labeling space, and can thus reduce the salt-and-pepper noise in the classification map. However, this procedure may blur the classification results in edge regions. This phenomenon can be attributed to the fact that the spatial pattern in an edge area differs from that in a homogenous area, since neighborhood information with respect to other land-cover classes is taken into account when extracting contextual information for edge pixels. In this situation, spatial details and edges can be over-smoothed.

### 2.2. Tri-training

Tri-training is an effective semi-supervised learning algorithm (Tan et al., 2016; Zhou and Li, 2005). Three classifiers are initially learned from the original training set, and are then iteratively refined using unlabeled examples in the tri-training process.

In this paper, we propose a novel tri-training method for alleviating edge issues by means of automatic selection of informative training samples, based on the following considerations. Firstly, the three tri-training classifiers can collaboratively select training samples with low classification certainties, which are mainly distributed near the edge areas, and hence are informative for the classification (Foody and Mathur, 2006; Mellor and Boukir, 2017). With the aid of these newly selected training samples, the classifiers are enhanced iteratively, and it is more likely that edge pixels are accurately identified (Tuia et al., 2011). It is important that the classifiers considered in the tri-training module should be diverse, and their performance should be complementary. The flowchart of the tri-training module is described as follows.

**Table 1**
Landscape pattern metrics used for the relearning.

| Landscape metrics | Calculation | Description | Performance |
|---|---|---|---|
| Mean patch size (MPS) | $\frac{\sum_{i=1}^{n} a_i}{n}$ | $a_i$ is the area ($m^2$) of patch $i$<br>$n$ is the number of patches for class $i$ | The relative size of the patches in the landscape |
| Standard deviation of area (AREA_SD) | $\mathrm{Std}\left(\bigcup_{i=1}^{n} a_i\right)$ | $a_i$ is the area ($m^2$) of patch $i$ | Area distribution of different land-cover classes |
| Largest patch index (LPI) | $\frac{max_{i=1}^{n}(a_i)}{A}$ | $a_i$ is the area ($m^2$) of patch $i$<br>A is the total landscape area ($m^2$) | The percentage of total landscape area comprised by the largest land-cover patch, highlighting the dominant type in an urban scene |
| Edge density (ED) | $\frac{E}{A}$ | $E$ is the total length of edges in the landscape, and A is the total landscape area ($m^2$) | Total edges of a land-cover class relative to the total landscape area, quantifying the landscape structure from the edge aspect |
| Mean shape index (SHAPE_MN) | $\frac{\sum_{i=1}^{n} \frac{p_i}{2\sqrt{\pi a_i}}}{n}$ | $p_i$ is the perimeter of land-cover patch $i$, $a_i$ is the area of the land-cover patch, and $n$ is the number of patches within the landscape | Average measure of the shape complexity of each land-cover class |
| Standard deviation of shape index (SHAPE_SD) | $\mathrm{Std}\left(\bigcup_{i=1}^{n} \frac{p_i}{2\sqrt{\pi a_i}}\right)$ | $p_i$ is the perimeter of land-cover patch $i$, $a_i$ is the area of the land-cover patch, and $n$ is the number of patches within the landscape | Distribution characteristics of shape complexity |
| Number of patches (NP) | $n_i$ | $n_i$ is the number of patches for class $i$ | Spatial fragmentation of land-cover patches |
| Splitting index (SPLIT) | $\frac{A^2}{\sum_{i=1}^{n} a_i^2}$ | $a_i$ is the area ($m^2$) of patch $i$<br>A is the total landscape area ($m^2$) | Spatial fragmentation of land-cover patches, but with different sensitivities to NP |
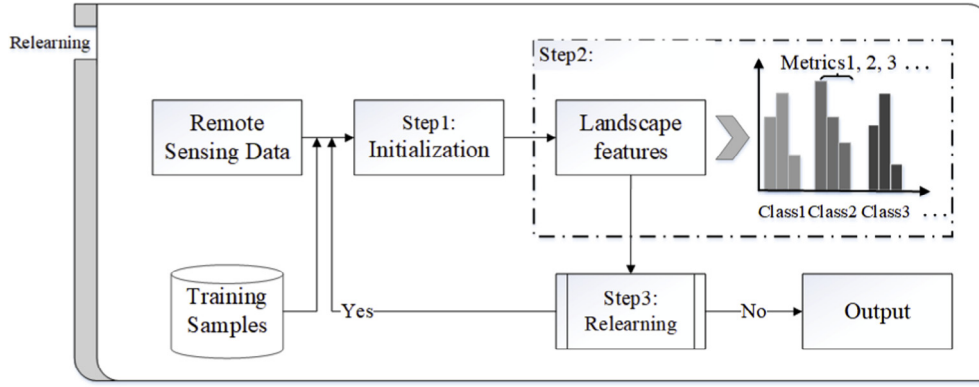
**Fig. 1.** Flowchart of the relearning-landscape module.

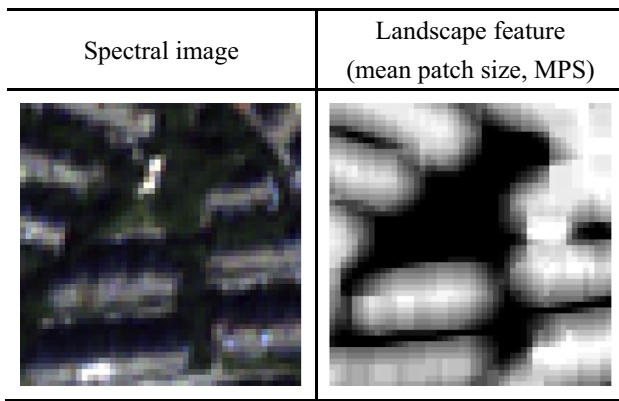| Spectral image | Landscape feature (mean patch size, MPS) |
|---|---|
|  |  |

**Fig. 2.** Demonstration of landscape features (mean patch size, MPS) of buildings, with the window size = 9.

**Step 1:** *Multi-classifier system classification.* The three classifiers (denoted as C1, C2, and C3) are initially trained. The spectral information is fed into the respective classifiers, resulting in the crisp (class label) and soft (probability) outputs for each classifier. Classification certainty maps are then derived from the probabilistic output of the classifiers, respectively (Grimm et al., 2008; Wang et al., 2017; Zhang and Seto, 2011):

$$S(x) = \sum_{k=1}^{K-1} [\hat{p}_k(x) - \hat{p}_{k+1}(x)] \cdot \frac{1}{k} \qquad (3)$$

where $S(x)$ is the classification certainty for pixel $x$, $\hat{p}_1(x),\ldots,\hat{p}_k(x),\ldots,\hat{p}_K(x)$ represent the multi-class probabilistic outputs in descending order, and $K$ is the number of information classes. A larger value of $S(x)$ foreshadows a more reliable classification of pixel $x$. After this step, the label results and certainty maps of the multiple classifiers are used for the subsequent analysis.

**Step 2:** *Automatic training sample selection.* The main objective of this step is to select informative unlabeled samples to improve the discriminative ability of the corresponding classification model. For any classifier, candidate samples are first selected if the other two classifiers agree on the labeling of these samples. Then, from the candidate samples, we choose pixels with certainty values lower than a threshold $Tri_{min}$, and at the same time with one of the other two certainty values larger than $Tri_{max}$. The chosen samples are then added to the training sets of the current classifier. More details of this step are shown in Fig. 3. In this way, two thresholds are used for the sample selection in tri-training: $Tri_{max}$ is used to guarantee that there

is at least one reliable classifier among the three classifiers, while $Tri_{min}$ is used to determine which classifier requires the newly selected training samples.

Please note that the correctness of the newly selected samples is guaranteed by the conditions that not only are consistent decisions on labels made by the three classifiers, but at least one reliable classifier in the multi-classifier system also exists.

**Step 3:** *Redundancy reduction of training samples.* The resulting samples from **step 2** may be spatially clustered and highly correlated. In order to reduce the redundancy and decrease the computational complexity, the connected components are labeled and only the central pixels in each component are retained to construct the new training sets. Subsequently, the training samples are iteratively updated until tri-training stops.

It should be mentioned that, although tri-training can be used to mitigate the over-smoothing effect, it is subject to salt-and-pepper noise in homogenous regions of objects due to the introduction of incorrect labels into the training sets (Tan et al., 2016; Zhou and Li, 2005).

### 2.3. The proposed edge-preservation multi-classifier relearning framework (EMRF)

In order to simultaneously increase the class separability and preserve edge details, EMRF, which is composed of the relearning-landscape and tri-training modules, is proposed. The basic idea of EMRF is to adopt relearning-landscape in non-edge pixels, aiming to smooth the classification in the homogeneous regions and, on the other hand, to impose the tri-training module in the edge pixels, aiming to preserve edges and details. The proposed EMRF is presented in Fig. 4, and is described as follows.

**Step 1:** *Multi-classifier system initialization.* The main objective of this step is to obtain the initial classification results by the use of the multiple classifiers. To this aim, spectral information is fed into the multi-classifier system, resulting in the class label and certainty map for each classifier. The reliable pixels ($x_r$) are defined as the ones that all the classifiers give the same label to, and the residual pixels are defined as unreliable ($x_{un}$). The labels of the reliable pixels can be determined by majority voting (MV):

$$C(x_r) = argmax_{k=\{1,\ldots,K\}} V_x(k)$$

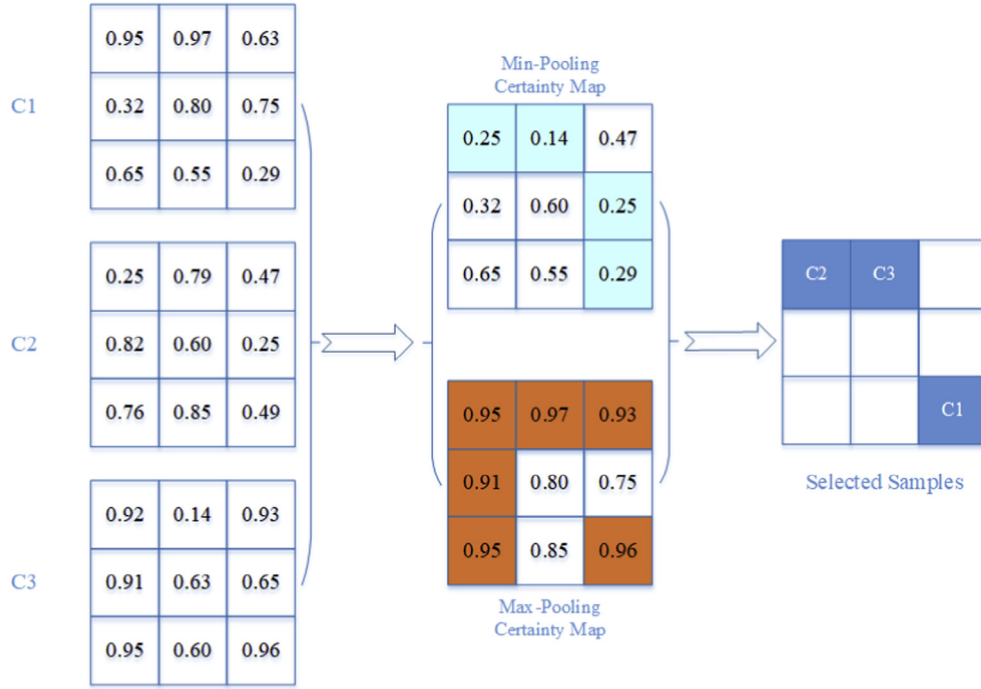$$\text{with } V_x(k) = \sum_{f=1}^{F} I(C^f(x) = k), \qquad (4)$$

**Fig. 3.** Demonstration of the sample selection in *step 2* of the tri-training module: $Tri_{max}$ is set to 0.90 and $Tri_{min}$ is set to 0.30. The leftmost column denotes the certainty maps for the C1, C2, and C3 classifiers, respectively. The middle column represents the pooling results for each pixel in the image. Max (min)-pooling is used to select the maximum (minimum) from certainty values of the three classifiers for each pixel. Specifically, samples marked in cyan have certainty values less than $Tri_{min}$ (0.30), while the certainty values of pixels marked in orange are larger than $Tri_{max}$ (0.90). The last column represents the samples which are selected as the final training sets for the C1, C2, and C3 classifiers, respectively. For example, the sample labeled 'C1' is added to the new training set of the C1 classifier. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
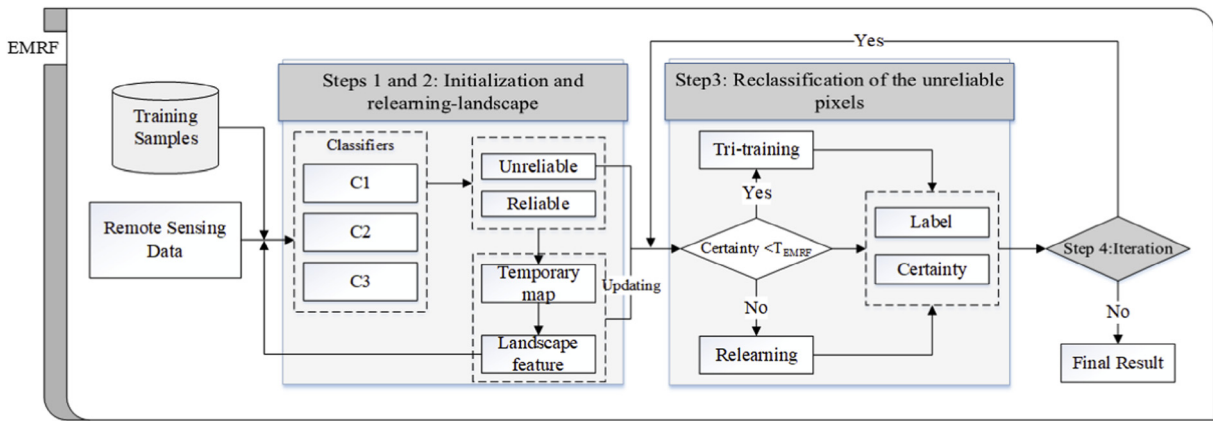


**Fig. 4.** The proposed edge-preservation multi-classifier relearning framework (EMRF), in which C1, C2, and C3 represent the three classifiers.

where $C(x_r)$ is the class label of reliable pixel $x$, $V_x(k)$ is the number of votes that pixel $x$ receives for class $k$, $I(.)$ represents the indicator function, and $C^f(x)$ is the class label of classifier $f$. The unreliable pixels are classified according to the certainty measure (defined in Eq. (3)).

$$C(x_{un}) = C(x_{\hat{f}}) \text{ with } \hat{f} = argmax_{f=\{1,...,F\}} S_f(x) \qquad (5)$$

where $C(x_{un})$ is the class label of unreliable pixel $x$, $S_f(x)$ is the classification certainty of classifier $f$ for pixel $x$, and $\hat{f}$ is the optimal classifier that has the largest certainty measure. In this way, the multiple classifiers are fused by minimizing the classification uncertainty (Huang and Zhang, 2013). An initial classification map is obtained, from which the landscape features can be derived.

**Step 2:** *Relearning-landscape.* Spectral information concatenated with the landscape features is fed into the multi-classifier system for reclassification. **Step 1** is then repeated and enhanced by considering the generated landscape metrics. Actually, this process can be viewed as multi-classifier relearning-landscape. In the meantime, new training samples can be generated for each classifier by implementing the tri-training method in this procedure (as described in Section 2.2).

**Step 3:** *Reclassification of the unreliable pixels.* Only unreliable pixels generated in **Step 2** are considered for further refinement, in order to raise the efficiency of the proposed method. Specifically, a moving window is centered at each unreliable pixel, and the landscape features are iteratively updated for this pixel, based on the current classification result. In this way, the

procedure in **Step 2** is iteratively implemented by updating the relearning-landscape and tri-training modules. In particular, these unreliable pixels are divided into relatively high and low certainty values (in terms of a threshold $T_{EMRF}$), and are identified according to the following rules.

**Rule1:** The pixels with high certainty are often found in the non-edge areas, and hence are classified by the updated relearning model, since relearning can effectively identify the non-edge areas.

**Rule 2:** The pixels with low certainty are classified through the tri-training classification, which is good at identifying edge pixels by selecting informative samples near the boundary regions (Foody and Mathur, 2006; Tan et al., 2016; Zhou and Li, 2005).

In summary, in this step, the threshold $T_{EMRF}$ is used to determine the two groups of pixels that will be further processed by the relearning-landscape and tri-training modules, respectively. A smaller value signifies that only a small number of unreliable pixels are fed into the tri-training module, which can smooth the classification result but blur the details and edges. In turn, a large value means that more details and edges can be preserved, but salt-and-pepper noise may exist.

**Step 4:** *Iteration.* The classification result in **Step 2** can be iteratively updated by the output in **Step 3**, until the loop termination.

It should be underlined that, in the EMRF method, the relearning-landscape and tri-training modules work in a collaborative manner. To be specific, relearning-landscape can increase class separability between land-cover classes, and help tri-training to acquire more reliable and informative training samples. With the newly selected training samples, the classification model is substantially improved and, therefore, a more reliable classification map can be generated and subsequently used for updating the landscape features in the relearning-landscape module. Therefore, the proposed EMRF benefits from the collaboration of relearning-landscape and tri-training.

## 3. Data sets and experimental setup

### 3.1. Data sets

To validate the effectiveness of the proposed framework, experiments were conducted on four multispectral remote sensing images: GeoEye-1 Wuhan (GE-1), QuickBird Wuhan (QB), WorldView-2 Hainan (WV-2), and ZY-3 Wuhan (ZY-3). In these data sets, seven land-cover classes are considered—buildings, roads, trees, grass, water, soil, and shadow—since they are the basic elements in urban areas (Huang et al., 2014b; Luo and Zhang, 2014). The characteristics of the four data sets are listed below:
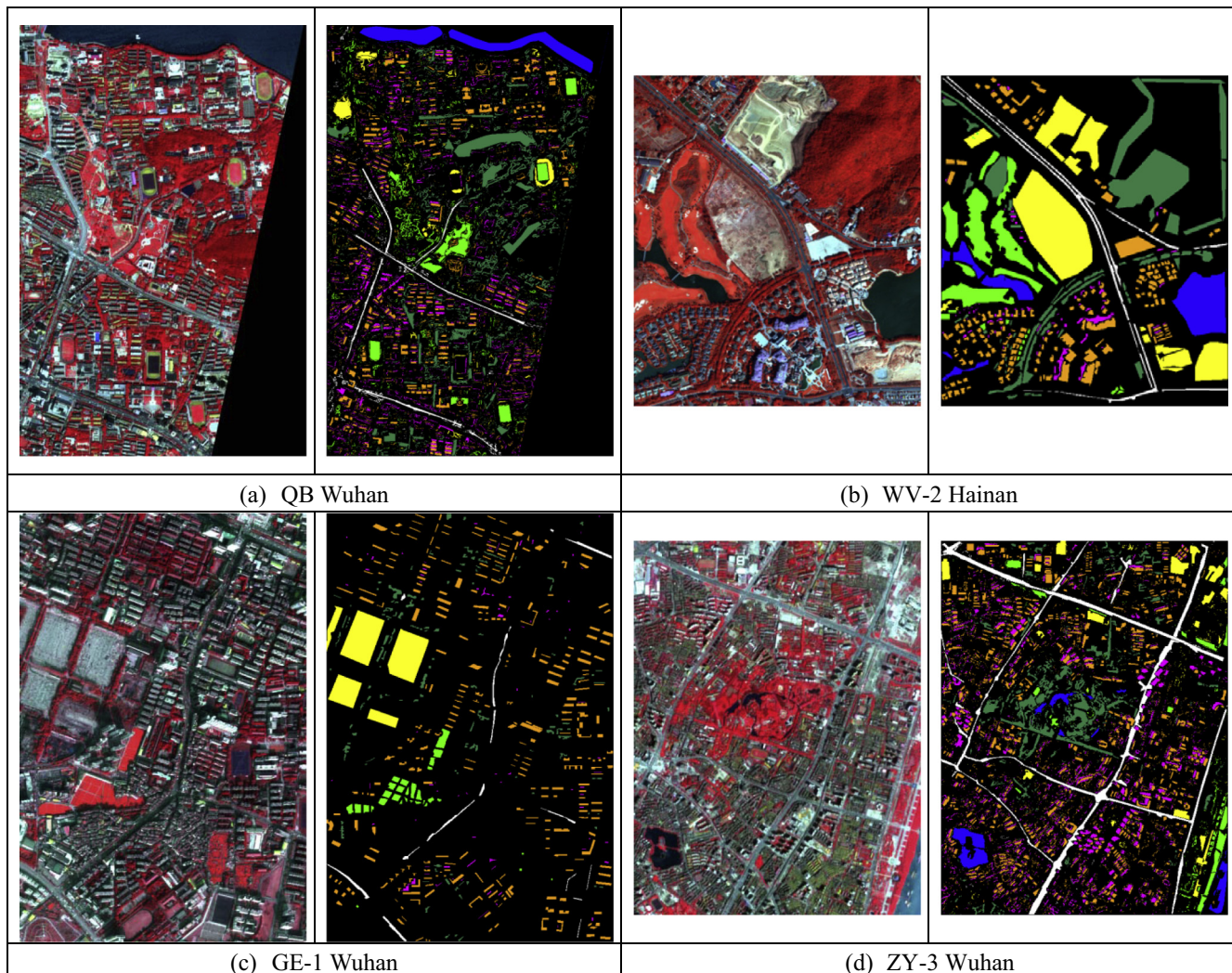


|                    |                    |
| :----------------: | :----------------: |
| (a)  QB Wuhan      | (b)  WV-2 Hainan   |
| (c)  GE-1 Wuhan    | (d)  ZY-3 Wuhan    |

**Fig. 5.** The test data sets and their reference maps: (a) QB Wuhan; (b) WV-2 Hainan; (c) GE-1 Wuhan; and (d) ZY-3 Wuhan.

(1) QB: This data set contains 1123 × 748 pixels, with four spectral bands at a 2.4-m spatial resolution (Fig. 5(a)). This image covers a campus scene, including regular buildings with heterogeneous roofs, forests, meadows, etc.

(2) WV-2: The second data set is WorldView-2 high spatial resolution (HSR) data (eight multispectral bands with a 2-m spatial resolution). This image covers a suburban area in the Hainan province of China, with the size of 600 × 520 pixels (Fig. 5(b)).

(3) GE-1: The third image is made up of 908 × 607 pixels, with four spectral bands and a 2-m spatial resolution, showing a typical urban landscape with dense residential areas, a lot of bare land for construction, and sparse vegetation (Fig. 5 (c)). It should be mentioned that there is no water in this data set, resulting in six land-cover classes.

(4) ZY-3: The last data set was also acquired over the city of Wuhan by ZY-3, which is China's first civilian high-resolution mapping satellite. This image contains 651 × 499 pixels, with a spatial resolution of 5.8 m and four spectral bands (Fig. 5(d)).

### 3.2. Reference maps

The reference maps shown in Fig. 5 were delineated manually according to the field investigation and our prior knowledge of the study areas. In order to separately assess the classification performance for edge pixels, we divided the reference map into edge and non-edge samples (see Fig. 6 and Table 2) and, therefore, the edge and non-edge classification accuracies of the proposed framework could be calculated. Specifically, the identification of
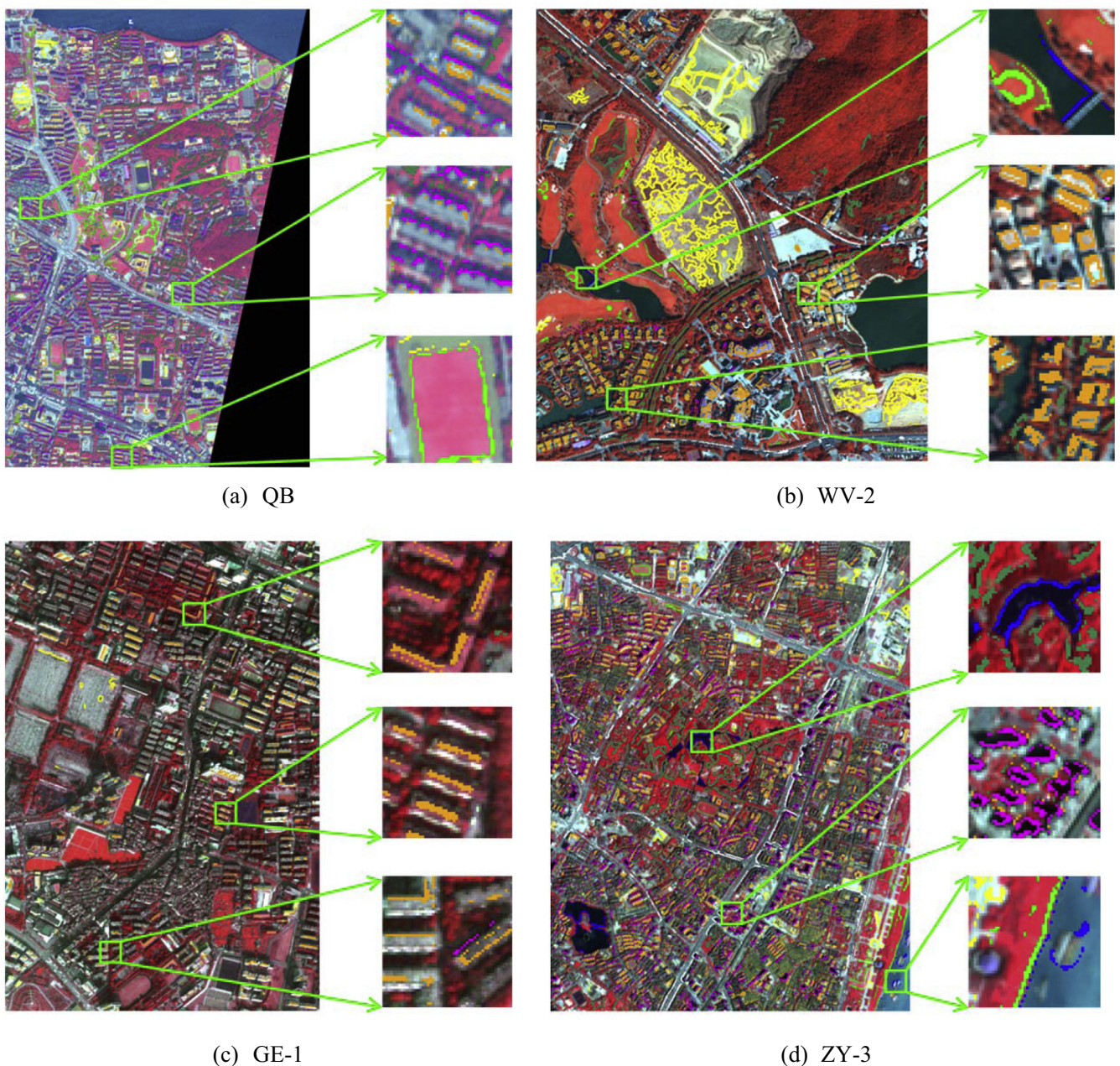


(a) QB

(b) WV-2

(c) GE-1

(d) ZY-3

**Fig. 6.** Edge samples used in this study: (a) QB; (b) WV-2; (c) GE-1; and (d) ZY-3.

**Table 2**
Number of reference samples (in pixels) for the four high-resolution data sets.

| Class | QB | | WV-2 | | GE-1 | | ZY-3 | |
|---|---|---|---|---|---|---|---|---|
| | Non-edge | Edge | Non-edge | Edge | Non-edge | Edge | Non-edge | Edge |
| Buildings | 9914 | 19,027 | 4628 | 7000 | 12,363 | 7761 | 6631 | 15,797 |
| Roads | 3623 | 4605 | 2774 | 2632 | 2794 | 443 | 10,309 | 3732 |
| Trees | 15,175 | 26,355 | 11,753 | 2383 | 4469 | 951 | 4502 | 7924 |
| Grass | 8338 | 11,795 | 6947 | 520 | 4036 | 112 | 2498 | 1798 |
| Water | 16,664 | 165 | 11,084 | 175 | – | – | 3583 | 721 |
| Soil | 2940 | 4439 | 13,248 | 8991 | 17,753 | 546 | 3249 | 1978 |
| Shadow | 3451 | 17,804 | 1167 | 310 | 1187 | 193 | 2644 | 11,082 |

edge pixels was achieved by the use of a standard Canny edge detector, with a standard deviation of $\sqrt{2}$ (Bao et al., 2005). The identified edges were widened by one pixel to form buffer areas. In this way, the test samples located in the buffer areas are called "edge samples", while the rest are defined as "non-edge samples" in the following.

### 3.3. Experimental parameters

The parameter settings used in the experiments are listed below, and the parameter sensitivity analysis is discussed in Section 5.3.

#### 3.3.1. Relearning-landscape

- Window size: In order to simultaneously capture the details and characterize the neighborhood extent (e.g., the spatial pattern and arrangement of the land-cover classes), a window size of 9 × 9 pixels was used.

#### 3.3.2. Tri-training

- Sample selection thresholds: $Tri_{min}$ and $Tri_{max}$ were set as 0.3 and 0.9, respectively, considering the tradeoff between computational burden and the performance of the classification model.

#### 3.3.3. EMRF

- Certainty threshold (as described in **step 3** of Section 2.3): was set as 0.8, aiming at smoothing the salt-and-pepper effect and at the same time preserving the edge details.

#### 3.3.4. Classification

- Classifier: It is important that the classifiers in EMRF should be diverse, and their performance should be complementary. Three classifiers were used to implement the proposed framework: support vector machine (SVM), random forest (RF), and sparse multinomial logistic regression via variable splitting and augmented Lagrangian (LORSAL). A description of these classifiers follows.
  (1) Support vector machine (SVM). SVM is a supervised non-parametric statistical learning technique, which is not constrained to prior assumptions on the distribution of the input data. Due to its ability to deal with large input spaces and produce sparse solutions, SVM has been widely used for the classification of remotely sensed imagery (Fauvel et al., 2008; Melgani and Bruzzone, 2004; Mountrakis et al., 2011). The parameters of SVM were set as kernel = radial basis function (RBF), penalty coefficient = 100, and RBF bandwidth = $1/n$ (where $n$ is the dimension of the input features) (Huang et al., 2014b).
  (2) Random forest (RF). RF is a classifier constructed from an ensemble of classification and regression trees (CART), which uses the majority vote of its constituent terminal

nodes to predict the class of a given observation. RF can handle a high-dimensional feature space with less computation, and it is insensitive to noise in training samples (Belgiu and Drăguţ, 2016). For the RF classifier, 200 trees were constructed, considering the tradeoff between computational burden and classification accuracy (Huang et al., 2016). A random subset of $\sqrt{n}$ features was used for RF at each node, where $n$ is the number of features (Wade et al., 2016).
  (3) Sparse multinomial logistic regression via variable splitting and augmented Lagrangian (LORSAL). As a discriminative classifier, multinomial logistic regression (MLR) directly models the class posterior densities instead of the joint probability distributions. To interpret high-dimensional data sets, the LORSAL algorithm has been proposed to replace the difficult non-smooth convex problem of MLR. The LORSAL-based sparse classifier (called the LORSAL classifier) has been proven to be effective for the classification of remotely sensed imagery, even with a limited number of training samples (Li et al., 2011a). With regard to the parameters in the LORSAL method, the empirical parameter settings were used (Li et al., 2015).
- Training: 50 training samples per class are randomly selected from the reference maps, and the rest are used for validation (Schindler, 2012).
- Accuracy assessment: the overall accuracy (OA) was computed from the confusion matrix for the quantitative assessment.

## 4. Results

The experimental results obtained with the QB, WV-2, GE-1, and ZY-3 data sets are presented in Figs. 7 and 8. In the experiments, individual use of the multi-classifier relearning-landscape and tri-training modules was considered for a comparative analysis. The general comments regarding the results are summarized as follows:

(1) Relearning-landscape obtains better results than tri-training in terms of non-edge accuracy, which indicates that landscape features have the potential to enhance the class separability (Fig. 7). Furthermore, this phenomenon can also be confirmed by Fig. 8, where relearning-landscape achieves higher overall accuracies than tri-training in the WV-2, GE-1, and ZY-3 experiments. However, its edge accuracy is relatively low due to the over-smoothing effect. For instance, after five iterations, its edge accuracies are 84.0%, 85.9%, 89.3%, and 80.8%, for QB, WV-2, GE-1, and ZY-3, respectively, which are lower than those obtained by tri-training, which are 89.0%, 93.7%, 93.2%, and 86.9%, respectively.
(2) Concerning the edge regions, tri-training outperforms relearning-landscape in all the cases. However, it can also be seen that tri-training has the lowest non-edge accuracy, due to the introduction of incorrect labels into the training sets (Tan et al., 2016; Zhou and Li, 2005), as previously mentioned. For instance, after five iterations, its non-edge accu-
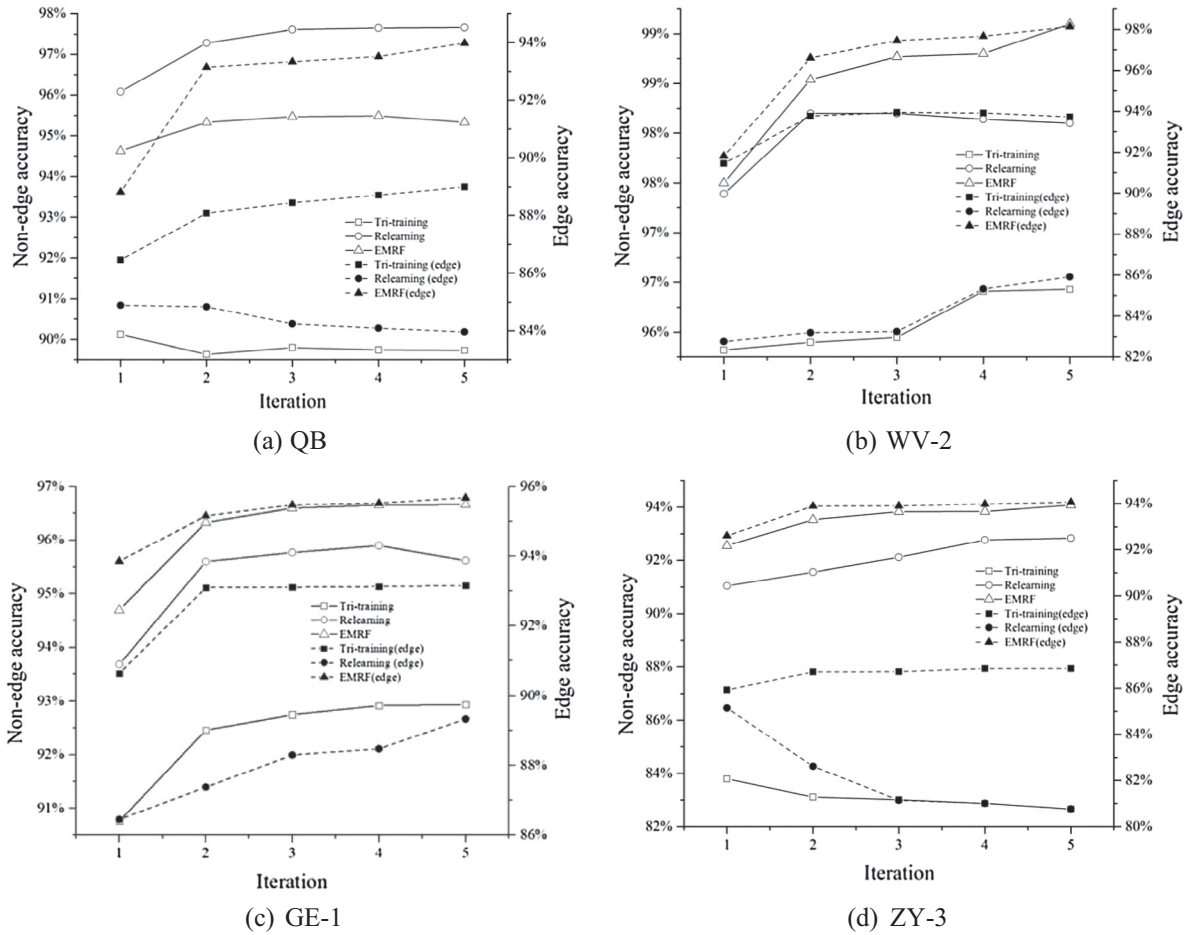
**Fig. 7.** Edge and non-edge accuracies of relearning-landscape, tri-training, and EMRF for: (a) QB; (b) WV-2; (c) GE-1; and (d) ZY-3. Note that the solid lines represent the non-edge accuracy curves, with the vertical y-axis on the left, while the dashed lines represent the edge accuracy, with the vertical y-axis on the right. The x-axis represents the iteration number.

racies are 89.7%, 92.9%, and 82.7% for QB, GE-1, and ZY-3, respectively, which are lower than those obtained by relearning, which are 97.6%, 95.6%, and 92.8%, respectively (Fig. 7).

(3) EMRF achieves the highest non-edge and edge accuracies in nearly all the cases. This can be attributed to the fact that EMRF takes both class separability enhancement and edge-detail preservation into account, by courtesy of the collaboration between relearning and tri-training. Moreover, in terms of the overall accuracy based on both edge and non-edge samples, EMRF obtains the most accurate results in all the experiments (Fig. 8). Specifically, after five iterations, the OA values are 94.6%, 98.8%, 96.5%, and 94.0% for QB, WV-2, GE-1, and ZY-3, respectively (Tables 3–6). Compared to relearning-landscape and tri-training, the accuracy improvements achieved by EMRF are 2.0–8.8%, which confirms the efficacy of the proposed EMRF.

(4) To allow a visual inspection, the classification maps obtained with the different data sets are displayed in Fig. 9. With respect to the spectral-based classification, misclassifications between the spectrally similar classes such as roads, buildings, and soil are obvious, and salt-and-pepper noise can be clearly observed (Fig. 9(a)). The classification map of relearning-landscape appears clear, showing the efficiency of reducing the salt-and-pepper noise (Fig. 9(b)). However, it should be mentioned that the efficiency of relearning-landscape can be attributed to the fact that this

method is able to learn the intrinsic spatial configuration from the raw classification, and it can provide sufficient discriminative information for the spectrally similar classes. Nevertheless, the detailed structures and edges of the classification map obtained with relearning-landscape are blurred, due to the over-smoothing effect. Edges and details are preserved in the classification map obtained with the tri-training method, but homogenous regions are subject to salt-and-pepper noise (Fig. 9(c)). The objects in the classification map obtained with EMRF show homogeneous surfaces, and the boundaries between adjacent objects are clear (Fig. 9(d)).

## 5. Discussion

This section includes the landscape feature analysis, a comparison study, and the parameter sensitivity analysis, followed by an evaluation of the proposed EMRF method.

### 5.1. Landscape feature analysis

Eight commonly used landscape metrics were investigated in the relearning-landscape module. Each landscape metric was individually calculated for relearning and classification, aiming to investigate their role in the classification. The general results with the different landscape features are presented in Fig. 10. The accuracies are consistently high in all cases. For example, the OA of
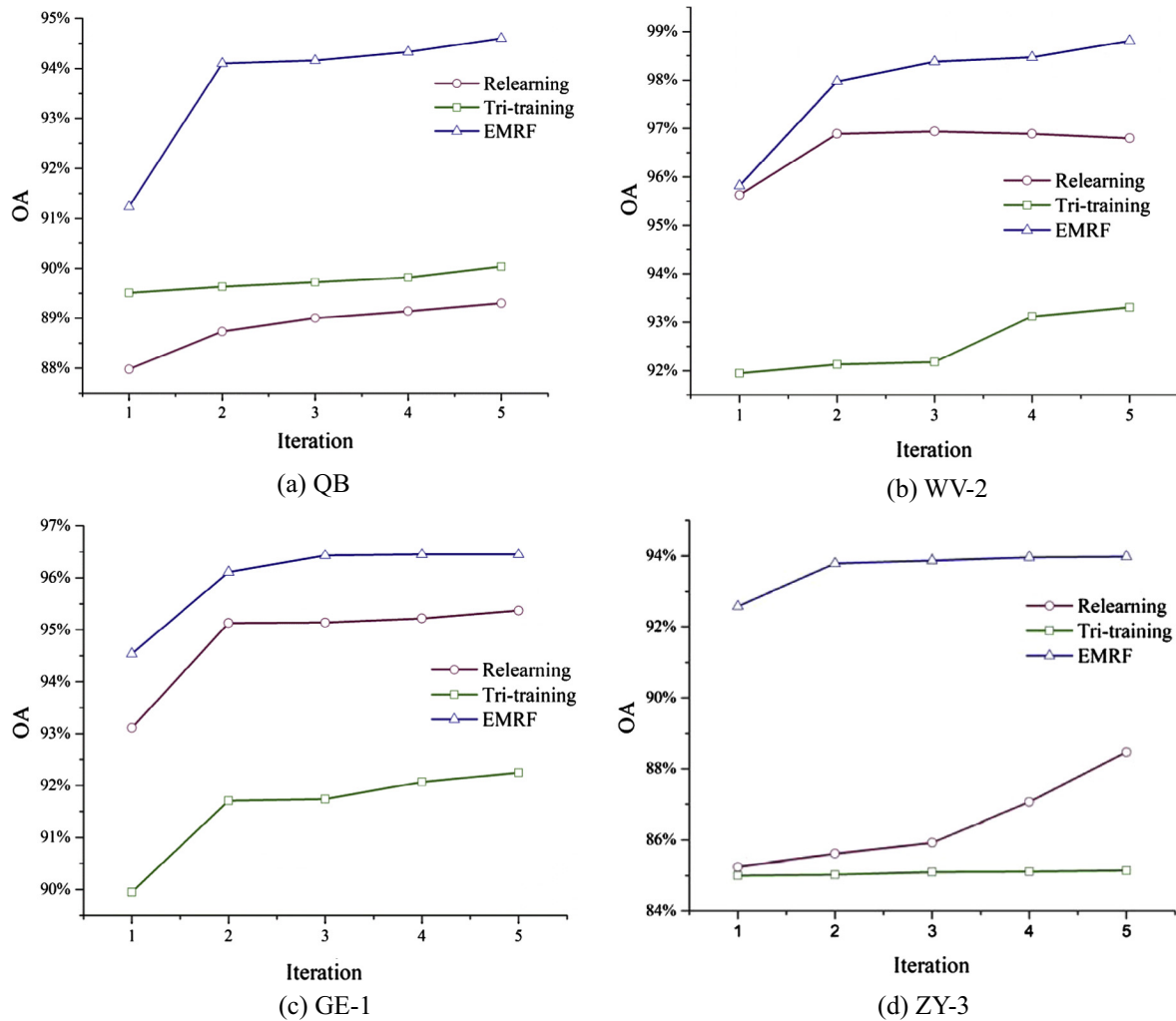
**Fig. 8.** Overall accuracies of relearning-landscape, tri-training, and EMRF for: (a) QB; (b) WV-2; (c) GE-1; and (d) ZY-3, based on both edge and non-edge samples.

**Table 3**
Confusion matrix obtained by the proposed EMRF method (after five iterations) for the QB data set (UA = user's accuracy, PA = producer's accuracy, and OA = overall accuracy).

| Classified data | Reference data | | | | | | | Total | UA (%) |
|---|---|---|---|---|---|---|---|---|---|
| | Buildings | Roads | Trees | Grass | Water | Soil | Shadow | | |
| Buildings | 24,677 | 328 | 96 | 474 | 5 | 52 | 339 | 25,971 | 95.0 |
| Roads | 1681 | 7826 | 31 | 156 | 0 | 9 | 130 | 9833 | 79.6 |
| Trees | 11 | 1 | 40,422 | 278 | 0 | 135 | 278 | 41,125 | 98.3 |
| Grass | 3 | 1 | 722 | 18,961 | 0 | 25 | 0 | 19,712 | 96.2 |
| Water | 0 | 5 | 0 | 0 | 16,774 | 0 | 1 | 16,780 | 100.0 |
| Soil | 2388 | 12 | 0 | 214 | 0 | 7108 | 56 | 9778 | 72.7 |
| Shadow | 131 | 5 | 209 | 0 | 0 | 0 | 20,401 | 20,746 | 98.3 |
| Total | 28,891 | 8178 | 41,480 | 20,083 | 16,779 | 7329 | 21,205 | | |
| PA(%) | 85.4 | 95.7 | 97.5 | 94.4 | 100.0 | 97.0 | 96.2 | | |
| OA = 94.6% | | | | | | | | | |

relearning with the largest patch index (LPI) is above 95.7% in the WV-2 data set. Only 3–4 iterations are needed for relearning-landscape to achieve a stable status. It can be seen that edge density (ED), LPI, and mean patch size (MPS) show excellent performances in terms of accuracy among the different landscape metrics. This indicates that the three metrics can effectively characterize different land-cover classes in the urban scenes. Specifically, both LPI and MPS describe the spatial structure from area aspects (see Table 1), while ED reflects the distribution of the land cover classes from edge aspect. In this study, the three metrics

were combined to deploy the relearning features, aiming to describe the different aspects of the landscape structures.

On the other hand, however, it should be mentioned that MPS, LPI, and ED show different performances for different image scenes. For instance, the highest accuracy is achieved by ED in the QB, GE-1, and ZY-3 datasets (Fig. 10(a), (c), and (d)), which exhibit typical urban characteristics, with dense buildings, sparse vegetation, and bare land. Considering that there are more edges and details in these urban scenes, ED can effectively quantify the landscape structures. However, the WV-2 data set represents a

**Table 4**
Confusion matrix obtained by the proposed EMRF method (after five iterations) for the WV-2 data set (UA = user's accuracy, PA = producer's accuracy, and OA = overall accuracy).

| Classified data | Reference data | | | | | | | Total | UA (%) |
|---|---|---|---|---|---|---|---|---|---|
| | Buildings | Roads | Soil | Grass | Shadow | Trees | Water | | |
| Buildings | *10,916* | 80 | 10 | 1 | 5 | 1 | 0 | 11,013 | 99.1 |
| Roads | 482 | *5274* | 0 | 4 | 0 | 1 | 0 | 5761 | 91.6 |
| Soil | 31 | 1 | *22,179* | 1 | 0 | 0 | 0 | 22,212 | 99.9 |
| Grass | 4 | 0 | 0 | *7411* | 0 | 7 | 0 | 7422 | 99.9 |
| Shadow | 138 | 0 | 0 | 0 | *1419* | 93 | 0 | 1650 | 86.0 |
| Trees | 7 | 1 | 0 | 0 | 3 | *13,984* | 0 | 13,995 | 99.9 |
| Water | 0 | 0 | 0 | 0 | 0 | 0 | *11,209* | 11,209 | 100.0 |
| Total | 11,578 | 5356 | 22,189 | 7417 | 1427 | 14,086 | 11,209 | | |
| PA (%) | 94.3 | 98.5 | 100.0 | 99.9 | 99.4 | 99.3 | 100.0 | | |
| OA = 98.8% | | | | | | | | | |

**Table 5**
Confusion matrix obtained by the proposed EMRF method (after five iterations) for the GE-1 data set (UA = user's accuracy, PA = producer's accuracy, and OA = overall accuracy).

| Classified data | Reference data | | | | | | Total | UA (%) |
|---|---|---|---|---|---|---|---|---|
| | Roads | Grass | Buildings | Soil | Shadow | Trees | | |
| Roads | *2914* | 0 | 1363 | 7 | 18 | 0 | 4302 | 67.7 |
| Grass | 0 | *4096* | 0 | 0 | 0 | 0 | 4096 | 100.0 |
| Buildings | 220 | 2 | *18,536* | 7 | 12 | 0 | 18,777 | 98.7 |
| Soil | 0 | 0 | 71 | *18,235* | 0 | 0 | 18,306 | 99.6 |
| Shadow | 53 | 0 | 93 | 0 | *1300* | 0 | 1446 | 89.9 |
| Trees | 0 | 0 | 11 | 0 | 0 | *5370* | 5381 | 99.8 |
| Total | 3187 | 4098 | 20,074 | 18,249 | 1330 | 5370 | | |
| PA (%) | 91.4 | 100.0 | 92.3 | 99.9 | 97.7 | 100.0 | | |
| OA = 96.5% | | | | | | | | |

**Table 6**
Confusion matrix obtained by the proposed EMRF method (after five iterations) for the ZY-3 data set (UA = user's accuracy, PA = producer's accuracy, and OA = overall accuracy).

| Classified data | Reference data | | | | | | | Total | UA (%) |
|---|---|---|---|---|---|---|---|---|---|
| | Roads | Grass | Soil | Buildings | Trees | Shadow | Water | | |
| Roads | *11,940* | 0 | 57 | 464 | 2 | 7 | 0 | 12,470 | 95.8 |
| Grass | 123 | *4118* | 42 | 168 | 223 | 1 | 23 | 4698 | 87.7 |
| Soil | 119 | 6 | *4940* | 292 | 0 | 0 | 0 | 5357 | 92.2 |
| Buildings | 1791 | 28 | 138 | *21,203* | 103 | 243 | 12 | 23,518 | 90.2 |
| Trees | 0 | 94 | 0 | 14 | *11,729* | 13 | 17 | 11,867 | 98.8 |
| Shadow | 18 | 0 | 0 | 237 | 292 | *13,412* | 29 | 13,988 | 95.9 |
| Water | 0 | 0 | 0 | 0 | 27 | 0 | *4173* | 4200 | 99.4 |
| Total | 13,991 | 4246 | 5177 | 22,378 | 12,376 | 13,676 | 4254 | | |
| PA (%) | 85.3 | 97.0 | 95.4 | 94.8 | 94.8 | 98.1 | 98.1 | | |
| ZY-3: OA = 94.0% | | | | | | | | | |

suburban landscape in Hainan Island, with a lot of grassland, small buildings, water bodies, and a golf course, and this data set possesses less-fragmented spatial patterns. LPI can quantify the percentage of total landscape area comprised by the largest land-cover class, indicating the dominant land-cover patch and fragmentation of the landscape (Fang et al., 2016; Han et al., 2017). This can partly explain why LPI obtains the highest accuracy in the WV-2 experiment (Fig. 10(b)).

### 5.2. Comparisons

#### 5.2.1. Comparisons between relearning-landscape and other relearning methods

In order to verify the effectiveness of relearning-landscape, its performance was compared with conventional relearning based on the primitive co-occurrence matrix (relearning-PCM) (Huang et al., 2014c) and object-based relearning (OBR) (Geiss and Taubenbock, 2015).

The experimental results show that relearning-landscape obtains a better performance than relearning-PCM and OBR, which

shows that landscape metrics are more appropriate for representing the structures and arrangement in the relearning procedure (Fig. 11).

#### 5.2.2. Additional experiments with the proposed EMRF method

Additional experiments were conducted in order to further verify the effectiveness of the proposed EMRF method. To this aim, the initial classification was implemented with a set of state-of-the-art algorithms: the spectral-spatial approach based on the gray-level co-occurrence matrix (GLCM) (Pesaresi et al., 2009), spectral-spatial classification using differential morphological profiles (DMP) (Benediktsson et al., 2003), and multi-index learning (MIL) (Huang et al., 2014b). Subsequently, the initial classification results were used as the input of the proposed EMRF for further refinement. When the initial classification was conducted with spectral-spatial classification based on the GLCM, it is referred to as "EMRF-GLCM".

The experimental results are reported in Table 7, where it can be clearly observed that the proposed EMRF method can significantly raise the classification accuracies for the different input
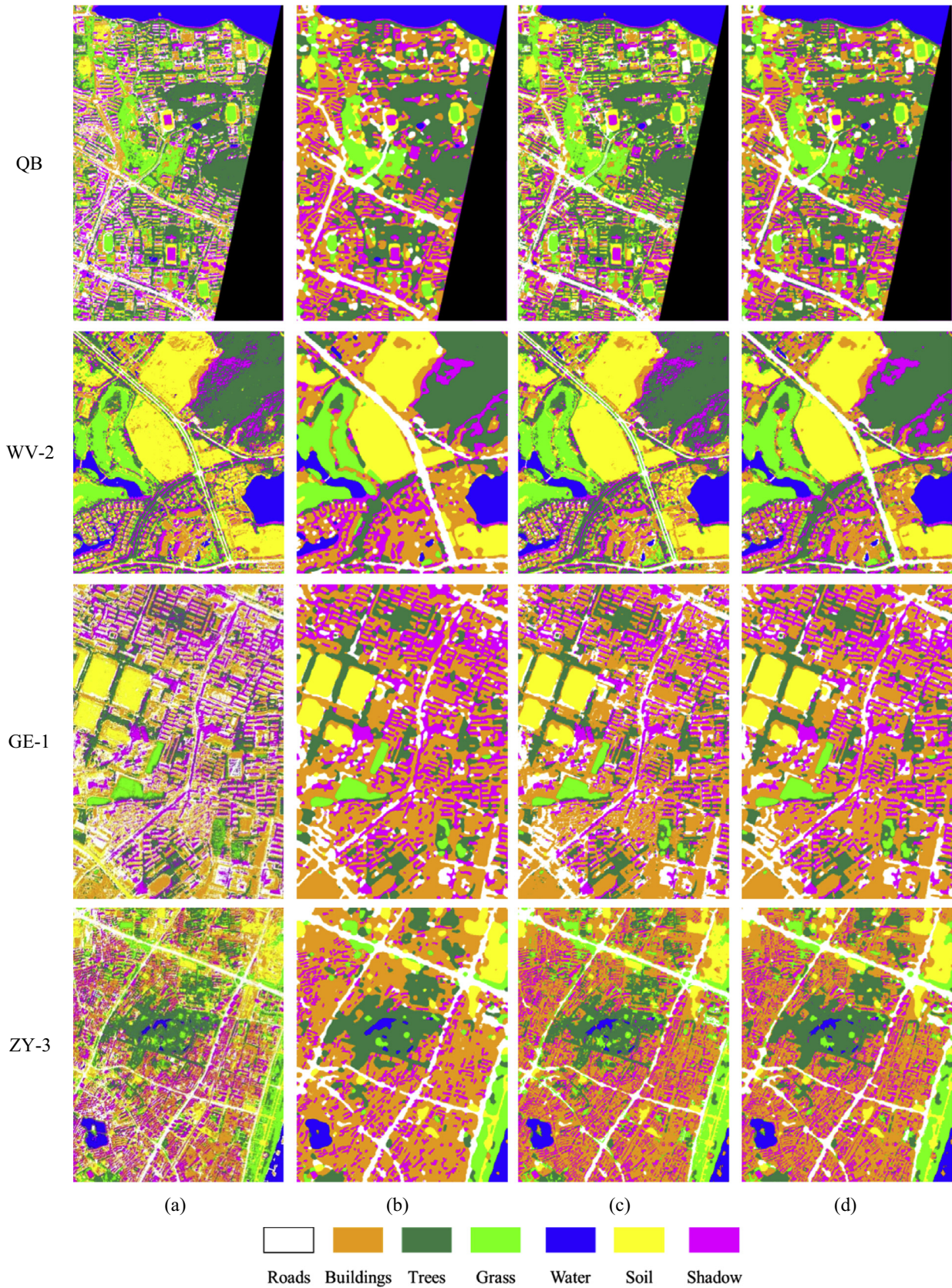
**Fig. 9.** Classification maps obtained with the different data sets by: (a) raw spectral-based classification; (b) relearning-landscape; (c) tri-training; and (d) EMRF (after five iterations).
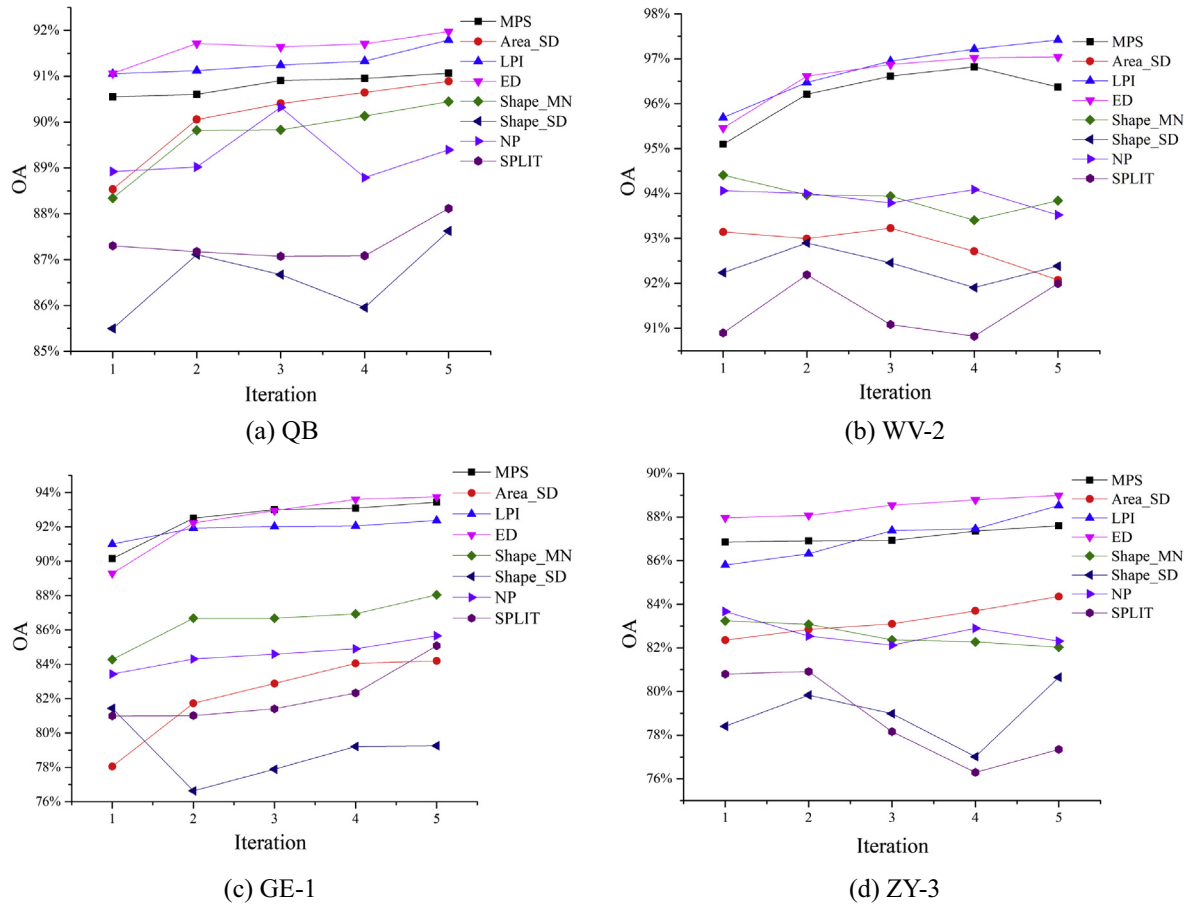
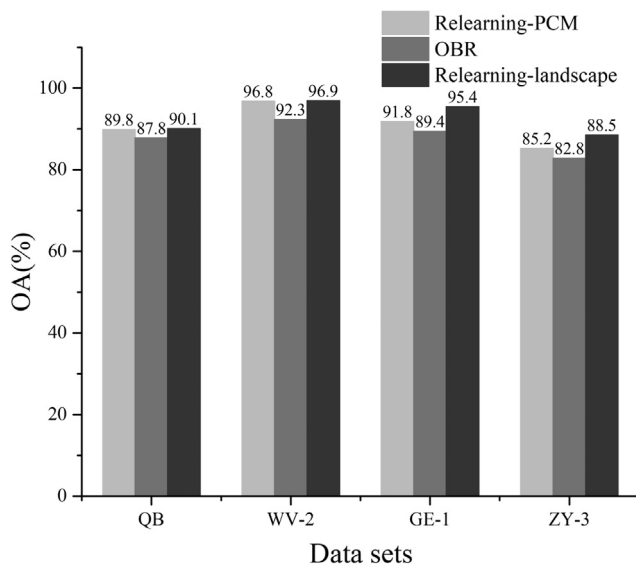Fig. 10. Analysis of relearning with different landscape features: (a) QB; (b) WV-2; (c) GE-1; and (d) ZY-3.



Fig. 11. Comparison between different relearning algorithms: the primitive co-occurrence matrix (relearning-PCM), object-based relearning (OBR), and the proposed relearning-landscape.

features. Specifically, the accuracy improvements achieved by EMRF are 14.6%, 7.9%, and 7.2% when compared with the GLCM, DMP, and MIL methods in the WV-2 experiment.

## 5.3. Parameter analysis

### 5.3.1. Effects of different window sizes on relearning-landscape

The window size is an important parameter for calculating the landscape metrics. A small size is inadequate to characterize the neighborhood extent, while a large size may fail to preserve the details. Therefore, we took the QB data set as an example, and investigated the relationship between window size and classification performance. The results are shown in Fig. 12, with three landscape metrics (MPS, LPI and ED) (see Table 1) used in the experiments. It can be observed that a window of $9 \times 9$ pixels is appropriate for describing the landscape features, depending on the spatial resolution and the characteristics of the information classes in the images.

### 5.3.2. Sample selection threshold for tri-training

The parameter sensitivity analysis for tri-training with the QB data set is given in Fig. 13. It can be observed that tri-training methods with $Tri_{max}$ in a range of 0.6–1.0 and $Tri_{min}$ in a range of 0–0.5 can lead to similar and stable performances. For instance, the maximum OA is 89.7% when $Tri_{min}$ and $Tri_{max}$ are set as 0.1 and 0.9, respectively, and the minimum OA is 89.4% when $Tri_{min}$ and $Tri_{max}$ are set as 0.5 and 0.6, respectively.

Actually, according to **step 2** in the tri-training module, we choose candidate samples with certainty values lower than a threshold $Tri_{min}$, and at the same time with one of the other two certainty values larger than $Tri_{max}$. When $Tri_{min}$ is high or $Tri_{max}$ is low, there will be more candidate samples to be selected by tri-training. This undoubtedly leads to a higher computational cost.

**Table 7**
Classification accuracies (%) of several of the state-of-the-art spectral-spatial classification algorithms, as well as their refinements using the proposed EMRF method. For instance, "EMRF-GLCM" represents GLCM-based spectral-spatial classification refined by EMRF.

| Data set | GLCM | EMRF-GLCM | DMP | EMRF-DMP | MIL | EMRF-MIL |
|---|---|---|---|---|---|---|
| QB | 91.9 | 96.3 | 91.1 | 96.8 | 92.1 | 97.0 |
| WV-2 | 83.5 | 98.1 | 91.3 | 99.2 | 91.9 | 99.1 |
| GE-1 | 81.5 | 95.6 | 86.9 | 95.4 | 90.4 | 96.8 |
| ZY-3 | 83.4 | 92.9 | 81.9 | 93.2 | 82.7 | 93.7 |
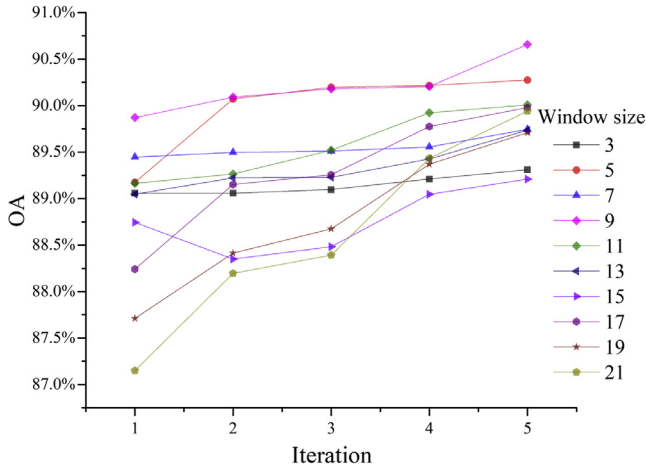


**Fig. 12.** Analysis of the effect of different window sizes on the classification accuracy (OA), with the horizontal and vertical axes being the iteration number and the obtained OA values with different window sizes, respectively.
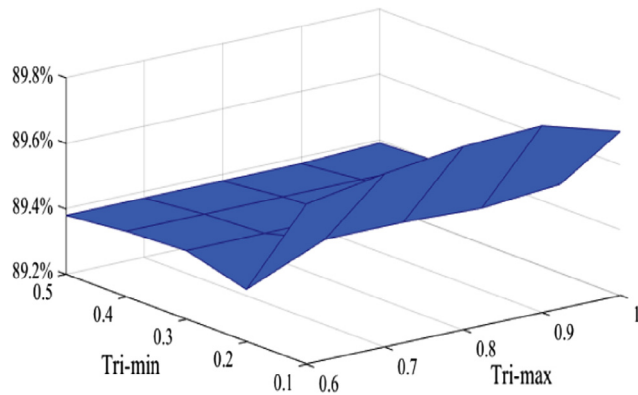


**Fig. 13.** Analysis of the parameter sensitivity with the QB data set for the tri-training module.

In turn, if $Tri_{min}$ is low or $Tri_{max}$ is high, there will be insufficient samples to improve the classification model. Therefore, in this study, $Tri_{min}$ and $Tri_{max}$ were set as 0.3 and 0.9, respectively, considering the tradeoff between computational burden and performance of the classification model.

### 5.3.3. Certainty threshold for EMRF

This subsection describes how the certainty threshold for EMRF influences the classification performance. The certainty threshold $T_{EMRF}$ is adopted to determine the proportion of pixels that will be further processed by the relearning-landscape or tri-training module (as described in **step 3** of EMRF). A smaller value signifies that a small number of pixels are fed into the tri-training model, while more pixels are classified by relearning.

Fig. 14 shows the relationship between edge and non-edge accuracies with the certainty threshold $T_{EMRF}$. With the increase
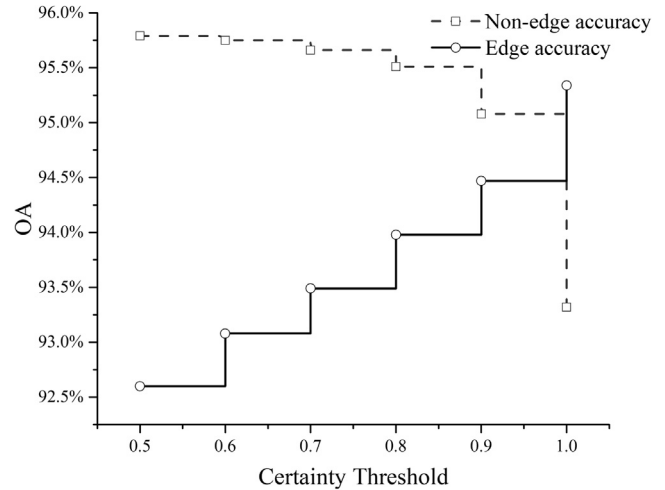


**Fig. 14.** Relationship between edge and non-edge accuracies with the certainty threshold $T_{EMRF}$ (as described in **step 3** of Section 2.3).

of $T_{EMRF}$, it can be seen that the edge accuracies increase, while the non-edge accuracies first remain stable, and are then with a slight decline. This phenomenon can be attributed to the fact that, with the increase of threshold $T_{EMRF}$, more edge pixels are classified by tri-training, resulting in a substantial increase of edge accuracies. On the other hand, non-edge pixels are classified by relearning-landscape, which guarantees relatively high and stable non-edge accuracies. However, when threshold $T_{EMRF}$ continues to increase (e.g., above 0.8 in this experiment), the non-edge accuracies become lower, since more pixels, including non-edge ones, are processed by tri-training, which can lead to salt-and-pepper noise in the classification. In order to obtain homogenous object surfaces and, at the same time, preserve edges and details in the
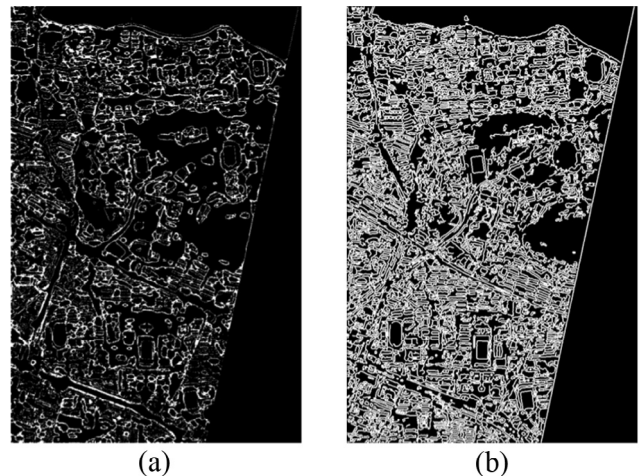


**Fig. 15.** (a) Difference between the classification results obtained with and without the tri-training module in EMRF, as indicated by the foreground pixels. (b) Edge areas in the image.

classification map, the certainty threshold $T_{EMRF}$ was set to 0.8 in this study.

### 5.4. Evaluation of the proposed EMRF method

#### 5.4.1. Role of the tri-training module in EMRF

The role of the tri-training module in the EMRF method was analyzed using the QB data set. We compared the differences between the classification results obtained with and without the tri-training module in EMRF. In the case of EMRF without tri-training, unreliable pixels are iteratively reclassified by the relearning-landscape module in each round. Striking differences between the classification maps of EMRF with and without the tri-training module can be observed in Fig. 15(a). For the compar-

ison, the boundary buffer zone obtained by the Canny edge detector is shown in Fig. 15(b). Through an overlay analysis, it can be found that 57.4% of the differences in Fig. 15(a) are within the edge zones. It is therefore shown that the tri-training module is mainly used for the interpretation of pixels located in edge regions, and this method has the potential to achieve reliable classification results for edge pixels (Fig. 7).

#### 5.4.2. Certainty analysis of EMRF

In order to analyze the mechanism of the proposed EMRF, a classification certainty analysis was conducted. Here, we take the SVM classifier of EMRF as an example, and certainty maps of the SVM classifier in terms of relearning-landscape, tri-training, and EMRF, are shown in Fig. 16. The first noteworthy
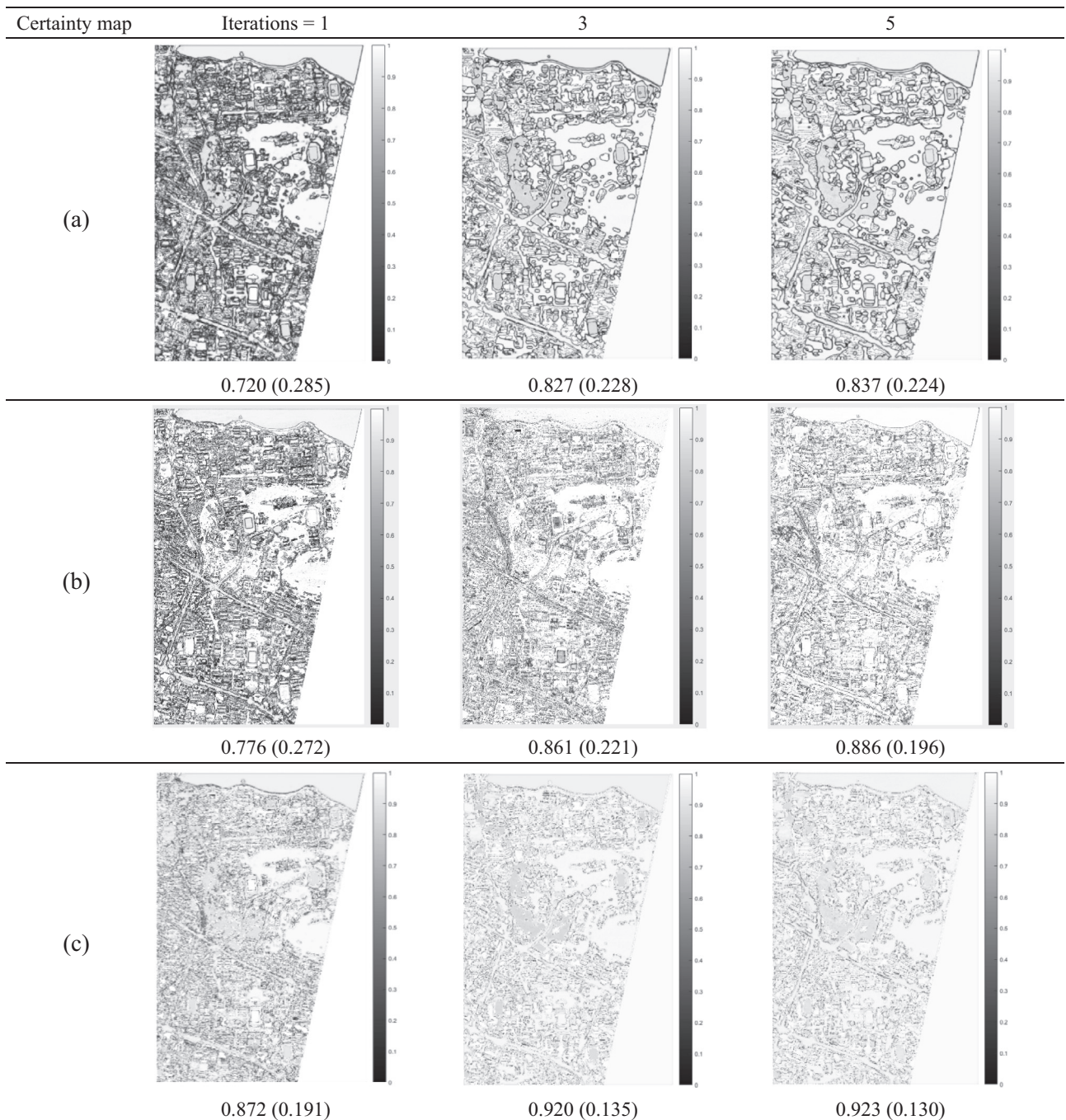


**Fig. 16.** Certainty maps of: (a) relearning; (b) tri-training; and (c) EMRF, after 1, 3, and 5 iterations. The mean value (standard deviation) is marked at the bottom of each certainty map.

observation from the results is that the certainty is substantially increased from iterations 1 to 5, regardless of the method used. For example, the certainty value of EMRF significantly increases from 0.872 to 0.923, with a lower standard deviation (from 0.191 to 0.130). This is understandable since the information about the spatial structure inherent in remotely sensed data is gradually delineated by the landscape features. Moreover, in each round, the highest certainty is achieved by EMRF. For instance, after the third iteration, the certainty values are 0.827, 0.861, and 0.920, with the standard deviations being 0.228, 0.221, and 0.135, for relearning, tri-training, and EMRF, respectively. Based on Figs. 7 and 16, it can be inferred that the increase in the certainty results in the increase in the classification accuracy.

## 6. Conclusions

This study was inspired by the fact that, as the spatial resolution of remotely sensed imagery is becoming increasingly high, it is difficult to simultaneously improve the class separability in the feature space and preserve edge details. In this context, the objective of this paper was to propose the edge-preservation multi-classifier relearning framework (EMRF) for the classification of high-resolution remotely sensed imagery.

In EMRF, relearning based on landscape features (relearning-landscape) is proposed to enhance the discriminative ability for land-cover classes, which is more appropriate for depicting the complex characteristics of remote sensing images. In order to mitigate the over-smoothing effect caused by the spatial features of relearning, a novel tri-training method is adopted, where unlabeled samples are exploited. EMRF flexibly combines the respective strengths of the relearning-landscape and tri-training modules, by taking advantage of their collaborative nature. To be specific, relearning-landscape can increase the class separability, and hence help tri-training acquire more informative and reliable training samples. With the aid of these newly selected training samples, the classification model can be substantially improved and, therefore, a more reliable classification map can be generated and subsequently used for updating the landscape features of relearning-landscape. In order to achieve an unbiased accuracy assessment, both edge and non-edge test samples were separately used for testing the classification performance. The experimental results obtained with four multispectral high-resolution data sets demonstrate that EMRF is not only able to significantly increase the classification accuracy and enhance the discriminative ability, but it can also preserve edges and details.

## Acknowledgements

## References

Alshehhi, R., Marpu, P.R., Wei, L.W., Mura, M.D., 2017. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 130, 139–149.

Bao, P., Zhang, L., Wu, X., 2005. Canny edge detection enhancement by scale multiplication. IEEE Trans. Pattern Anal. Mach. Intell. 27, 1485–1490.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Benediktsson, J.A., Pesaresi, M., Amason, K., 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. IEEE Trans. Geosci. Remote Sens. 41, 1940–1949.

Cheng, J., Liu, H., Liu, T., Wang, F., Li, H., 2015. Remote sensing image fusion via wavelet transform and sparse representation. ISPRS J. Photogramm. Remote Sens. 104, 158–173.

Fang, C., Li, G., Wang, S., 2016. Changing and differentiated urban landscape in China: spatiotemporal patterns and driving forces. Environ. Sci. Technol. 50, 2217.

Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R., 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans. Geosci. Remote Sens. 46, 3804–3814.

Foody, G.M., Mathur, A., 2006. The use of small training sets containing mixed pixels for accurate hard image classification: training on mixed spectral responses for classification by a SVM. Remote Sens. Environ. 103, 179–189.

Gamba, P., Dell'Acqua, F., Lisini, G., Trianni, G., 2007. Improved VHR urban area mapping exploiting object boundaries. IEEE Trans. Geosci. Remote Sens. 45, 2676–2682.

Geiss, C., Taubenbock, H., 2015. Object-based postclassification relearning. IEEE Geosci. Remote Sens. Lett. 12, 2336–2340.

Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X., Briggs, J.M., 2008. Global change and the ecology of cities. Science 319, 756–760.

Han, X., Huang, X., Liang, H., Ma, S., Gong, J., 2017. Analysis of the relationships between environmental noise and urban morphology. Environ. Pollut. 233, 755–763.

Huang, X., Han, X., Zhang, L., Gong, J., Liao, W., Benediktsson, J.A., 2016. Generalized differential morphological profiles for remote sensing image classification. IEEE J. Sel. Top Appl. Earth Obs Remote Sens. 9, 1736–1751.

Huang, X., Liu, X., Zhang, L., 2014a. A multichannel gray level co-occurrence matrix for multi/hyperspectral image texture representation. Remote Sens. 6, 8424–8445.

Huang, X., Lu, Q., Zhang, L., 2014b. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. ISPRS J. Photogramm. Remote Sens. 90, 36–48.

Huang, X., Lu, Q., Zhang, L., Plaza, A., 2014c. New postprocessing methods for remote sensing image classification: a systematic study. IEEE Trans. Geosci. Remote Sens. 52, 7140–7159.

Huang, X., Zhang, L., 2013. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 51, 257–272.

Li, J., Bioucas-Dias, J.M., Plaza, A., 2011a. Hyperspectral image segmentation using a new Bayesian approach with active learning. IEEE Trans. Geosci. Remote Sens. 49, 3947–3960.

Li, J., Huang, X., Gamba, P., Bioucas-Dias, J.M., Zhang, L., Benediktsson, J.A., Plaza, A., 2015. Multiple feature learning for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 53, 1592–1606.

Li, J., Song, C., Cao, L., Zhu, F., Meng, X., Wu, J., 2011b. Impacts of landscape structure on surface urban heat islands: a case study of Shanghai, China. Remote Sens. Environ. 115, 3249–3263.

Luo, B., Zhang, L., 2014. Robust autodual morphological profiles for the classification of high-resolution satellite images. IEEE Trans. Geosci. Remote Sens. 52, 1451–1462.

Ma, L., Cheng, L., Li, M., Liu, Y., Ma, X., 2015. Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery. ISPRS J. Photogramm. Remote Sens. 102, 14–27.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017. A review of supervised object-based land-cover image classification. ISPRS J. Photogramm. Remote Sens. 130.

McGarigal, K., Cushman, S.A., Neel, M.C., Ene, E., 2002. FRAGSTATS: spatial pattern analysis program for categorical maps.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1778–1790.

Mellor, A., Boukir, S., 2017. Exploring diversity in ensemble classification: applications in large area land cover mapping. ISPRS J. Photogramm. Remote Sens. 129, 151–161.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247–259.

Myint, S.W., 2004. Wavelets for urban spatial feature discrimination: comparisons with fractal, spatial autocorrelation, and spatial co-occurrence approaches. Photogramm. Eng. Remote Sens. 70, 803–812.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens. Environ. 115, 1145–1161.

Oort, P.A.J.v., Bregt, A.K., Bruin, S.d., Wit, A.J.W.d., Stein, A., 2004. Spatial variability in classification accuracy of agricultural crops in the Dutch national land-cover database. Int. J. Geograph. Inform. Sci. 18, 611–626.

Pesaresi, M., Benediktsson, J.A., 2002. A new approach for the morphological segmentation of high-resolution satellite imagery. IEEE Trans. Geosci. Remote Sens. 39, 309–320.

Pesaresi, M., Gerhardinger, A., Kayitakire, F., 2009. A robust built-up area presence index by anisotropic rotation-invariant textural measure. IEEE J. Sel. Top. Appl. Earth Obs Remote Sens. 1, 180–192.

Prabhakar, T.V.N., Geetha, P., 2017. Two-dimensional empirical wavelet transform based supervised hyperspectral image classification. ISPRS J. Photogramm. Remote Sens. 133, 37–45.

Pu, R., Landry, S., 2012. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. Remote Sens. Environ. 124, 516–533.

Rodriguez-Cuenca, B., Malpica, J.A., Alonso, M.C., 2012. A spatial contextual postclassification method for preserving linear objects in multispectral imagery. IEEE Trans. Geosci. Remote Sens. 51, 174–183.

Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. IEEE Trans. Geosci. Remote Sens. 50, 4534–4545.

Smits, P.C., Dellepiane, S.G., 1997. Synthetic aperture radar image segmentation by a detail preserving Markov random field approach. IEEE Trans. Geosci. Remote Sens. 35, 844–857.

Solaiman, B., Koffi, R.K., Mouchot, M.C., Hillion, A., 1995. An information fusion method for multispectral image classification postprocessing. IEEE Trans. Geosci. Remote Sens. 36, 395–406.

Song, W., Li, M., Zhang, P., Wu, Y., Jia, L., An, L., 2017. Unsupervised PolSAR image classification and segmentation using dirichlet process mixture model and markov random fields with similarity measure. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. pp. 1–13.

Stuckens, J., Coppin, P.R., Bauer, M.E., 2000. Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sens. Environ. 71, 282–296.

Tan, K., Hu, J., Li, J., Du, P., 2015. A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination. ISPRS J. Photogramm. Remote Sens. 105, 19–29.

Tan, K., Zhu, J., Du, Q., Wu, L., Du, P., 2016. A novel tri-training technique for semi-supervised classification of hyperspectral images based on diversity measurement. Remote Sens. 8, 749.

Tuia, D., Volpi, M., Copa, L., Kanevski, M., Munoz-Mari, J., 2011. A survey of active learning algorithms for supervised remote sensing image classification. IEEE J. Sel. Top. Signal Process. 5, 606–617.

Wade, B.S.C., Joshi, S.H., Gutman, B.A., Thompson, P.M., 2016. Machine learning on high dimensional shape data from subcortical brain surfaces: a comparison of feature selection and classification methods. Pattern Recogn. 9352, 36–43.

Wang, X.Z., Wang, R., Xu, C., 2017. Discovering the relationship between generalization and uncertainty by incorporating complexity of classification. IEEE Trans. Cybernet. pp. 1–13.

Woźniak, M., Graña, M., Corchado, E., 2014. A survey of multiple classifier systems as hybrid systems. Inform. Fusion 16, 3–17.

Zhang, L., Huang, X., Huang, B., Li, P., 2006. A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 44, 2950–2961.

Zhang, Q., Seto, K.C., 2011. Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. Remote Sens. Environ. 115, 2320–2329.

Zhao, J., Zhong, Y., Zhang, L., 2015. Detail-preserving smoothing classifier based on conditional random fields for high spatial resolution remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 53, 2440–2452.

Zhou, Z.H., Li, M., 2005. Tri-training: exploiting unlabeled data using three classifiers. IEEE Trans. Knowl. Data Eng. 17, 1529–1541.