





Article

# A Hybrid Data Balancing Method for Classification of Imbalanced Training Data within Google Earth Engine: Case Studies from Mountainous Regions

Amin Naboureh <sup>1,2</sup>, Ainong Li <sup>1,\*</sup>, Jinhua Bian <sup>1</sup>, Guangbin Lei <sup>1</sup> and Meisam Amani <sup>3</sup>

<sup>1</sup> Research Center for Digital Mountain and Remote Sensing Application, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610041, China; amin.nabore@mails.ucas.ac.cn (A.N.); bianjinhua@imde.ac.cn (J.B.); leiguangbin@imde.ac.cn (G.L.)

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> Wood Environment & Infrastructure Solutions, Ottawa, ON K2E 7K3, Canada; meisam.amani@woodplc.com

\* Correspondence: ainongli@imde.ac.cn; Tel.: +86-2885220048

Received: 18 September 2020; Accepted: 9 October 2020; Published: 11 October 2020



**Abstract:** Distribution of Land Cover (LC) classes is mostly imbalanced with some majority LC classes dominating against minority classes in mountainous areas. Although standard Machine Learning (ML) classifiers can achieve high accuracies for majority classes, they largely fail to provide reasonable accuracies for minority classes. This is mainly due to the class imbalance problem. In this study, a hybrid data balancing method, called the Partial Random Over-Sampling and Random Under-Sampling (PROSRUS), was proposed to resolve the class imbalance issue. Unlike most data balancing techniques which seek to fully balance datasets, PROSRUS uses a partial balancing approach with hundreds of fractions for majority and minority classes to balance datasets. For this, time-series of Landsat-8 and SRTM topographic data along with various spectral indices and topographic data were used over three mountainous sites within the Google Earth Engine (GEE) cloud platform. It was observed that PROSRUS had better performance than several other balancing methods and increased the accuracy of minority classes without a reduction in overall classification accuracy. Furthermore, adopting complementary information, particularly topographic data, considerably increased the accuracy of minority classes in mountainous areas. Finally, the obtained results from PROSRUS indicated that every imbalanced dataset requires a specific fraction(s) for addressing the class imbalance problem, because different datasets contain various characteristics.

**Keywords:** class imbalance problem; Google Earth Engine; land cover mapping; mountainous regions; time-series of Landsat

## 1. Introduction

Mountains, covering a quarter of earth's land surface, are globally distributed from the Tropics to the poles and from maritime to continental environments [1]. Obtaining up-to-date and accurate information of Mountain Land Cover (MLC) types is important for various applications, including global warming and environmental changes [2–4]. Moreover, MLC data is a vital part of the assessment and management of natural hazards studies (e.g., landslides and wildfires) [5–7]. Considering the large extent and limited accessibility of mountainous areas, Remote Sensing (RS) datasets are well-suited for mapping MLC classes. This is mainly related to the global coverage, the availability of various spatial and spectral resolutions, and frequent observations from RS systems [8–10].

The RS community has been so far examined various datasets and methodologies to meet users' requirements for generating accurate MLC maps [11–14]. The advent of state-of-the-art Machine Learning (ML) techniques has particularly helped the RS community to improve the accuracy of MLC

classifications [4,15]. However, the class imbalance problem, which mainly happens during the training of classification algorithms, is a common issue in almost all ML classifiers [16]. In most supervised classifiers, the class imbalance problem occurs when one or some of the classes have fewer samples than the others [17]. This is mainly because the numbers of samples for each MLC class depend on various factors, such as area and accessibility. For example, some classes may only cover a small portion of the study area, while the others cover larger regions, which causes inconsistency in the numbers of samples for different classes [18,19].

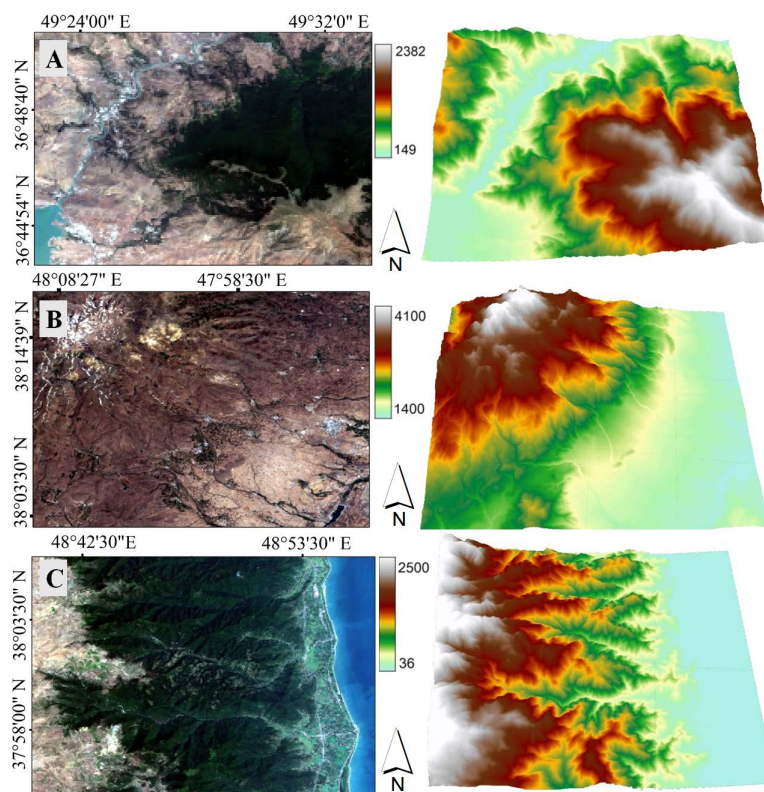
To date, several methods have been proposed to address the class imbalance problem, which can be generally classified into three approaches: (1) applying specific classification methods by focusing on the learning of minority classes [16], (2) assigning higher weights on minority classes by adjusting classifiers (e.g., cost-sensitive methods) [20], and (3) rebalancing training datasets (e.g., oversampling and under-sampling techniques) [21,22]. Among these methods, rebalancing training datasets has been received more attention mainly because these techniques are simple to implement and yield reasonable accuracies [23]. Rebalancing methods can be added to existing classifiers, and do not change the base classifiers [21]. Although rebalancing techniques are well-documented in the literature, they have multiple limitations related to the absence of the choice 'fit for all data sets' method, removal of some useful information, potential overfitting, and generating noisy samples [16,24]. Regarding the balancing rules, although it was argued that fully rebalancing original data might lead to a decrease in Overall Accuracy (OA) [25], partial balancing of datasets has been rarely considered by the RS community. Additionally, the role of different balancing ratios (fractions) to balance imbalanced datasets has been ignored in most data balancing studies [26–28]. However, this is important because datasets are different in terms of imbalance ratio, number of classes, and number of samples per class [25]. Therefore, investigating the impacts of different balancing fractions on balancing datasets is of great importance.

MLC classification can be relatively challenging due to a series of factors, such as high spatial heterogeneity, rugged terrain, and the cloud contamination in optical satellite imagery over the mountainous areas [4,9]. It has been proven that obtaining high classification accuracy over complex landscapes, such as mountainous regions, was challenging with the sole use of spectral bands [9,29]. Thus, examining the role of different complementary information (e.g., spectral indices and topographic data) to achieve more accurate MLC maps, particularly for minority classes, is necessary. Therefore, this study also discussed the role of different complementary datasets, including spectral indices and topographic data, on the accuracies of minority classes in mountainous areas. Additionally, MLC classification becomes more challenging over large mountainous areas because hundreds to thousands of satellite imagery might need to be processed and classified in a cost and time-efficient approach. To resolve these issues, different cloud computing platforms have been developed, one of the most commonly-used of which is Google Earth Engine (GEE) [30–32]. GEE is a cloud-based geospatial data analyzing server, which ensures that users can access and process massive troves of RS datasets. GEE allows experts and researchers to employ rich RS datasets to study local, regional, and global applications [32,33]. GEE not only allows users to access long time-series data but also substantially decreases the computational time [34,35]. Given the aforementioned background, the role of different balancing fractions was investigated in this study by proposing a new hybrid balancing technique, called the Partial Random Over-Sampling and Random Under-Sampling (PROSRUS). PROSRUS integrated by Random Forest (RF) based on time series of Landsat-8 OLI and Google Earth Engine (GEE) platform to handle the class imbalance problem in MLC classification.

## 2. Study Areas

In this study, three experiment sites (Figure 1) with different areas, elevation ranges, climate conditions, spatial distributions of MLC types, and number of samples were selected to comprehensively evaluate the robustness and performance of the proposed method. Site-1 (Figure 1A, Lon = 49°21'54"–49°34'09"E, Lat = 36°43'36"–36°51'10"N) covers an area of approximately 270 km<sup>2</sup>

in Gilan province, Iran, with an elevation ranging from 150 to 2382 m. This site covers a wide range of MLC types, dominated by Forest and Bare land. Site-2 (Figure 1B, Lon = 47°44′07″–48°08′27″E, Lat = 38°02′35″–38°17′28″N) covers an area of approximately 462 km<sup>2</sup>. This site belongs to Ardabil province, Iran, and is a part of Savalan Mountain with an elevation varying between 1400 and 4100 m. Bare land and Cultivated land are two dominant MLC types in this experiment site. Site-3 (Figure 1C, Lon = 48°39′46″–48°57′24″E, Lat = 37°54′14″–38°06′59″N) covers an area of 620 km<sup>2</sup> with an elevation ranging from 36 to 2500 m. Site-1 and Site-3 belong to Gilan province, Iran, and are parts of the Alborz chain. Most area in Site-3 is covered by Forest. Based on the Köppen-Geiger climate classification [36], Site-1 has arid, steppe, cold (BSk) climate; Site-2 has cold, dry summer, cold summer (Dsc) climate; and Site-3 has temperate, dry summer, hot summer (Csa) climate.

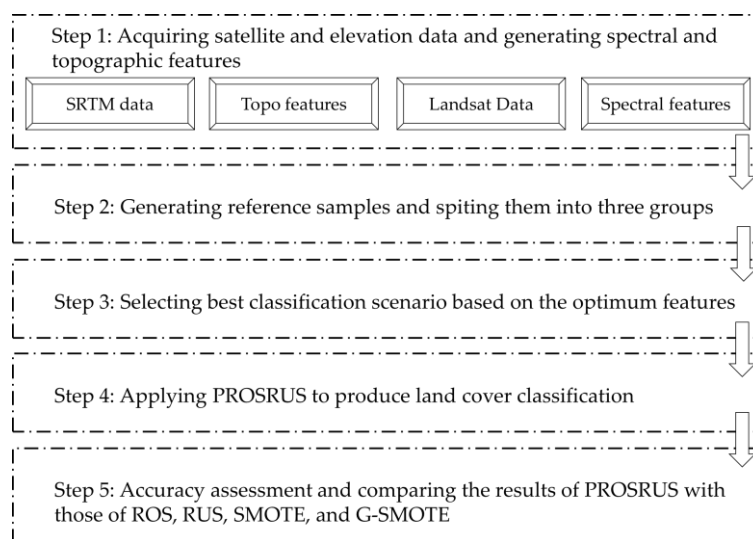


**Figure 1.** Experiment sites: (A) Site-1, (B) Site-2, and (C) Site-3. Left and right columns show the true color composites of Landsat-8 (R: Band-4, G: Band-3, B: Band-2) and the SRTM elevation maps, respectively

### 3. Method

#### 3.1. Overall Workflow

As illustrated in Figure 2, the overall workflow of this study, which were implemented in GEE consists of five main steps: (1) acquiring time-series of Landsat-8 OLI and SRTM imagery and generating complementary data (i.e., spectral indices and topographic products) within the GEE platform, (2) Generating reference samples and splitting them into three groups based on the number of samples (i.e., majority classes, middle classes, and minority classes), (3) Selecting the best spectral and topographic features for LC classification and assessing the effects of various features on the accuracy of minority classes using RF classifier, (4) Applying the PROSRUS method using 200 different fractions, and (5) Accuracy assessment of PROSRUS and evaluating its accuracy compared to those of the Random Over-Sampling (ROS) [24], Random Under-Sampling (RUS) [37], Synthetic Minority Over-sampling Technique (SMOTE) [21], and Geometric SMOTE (G-SMOTE) [38] techniques.



**Figure 2.** Overall workflow of the study (GEE: Google Earth Engine, PROSRUS: Partial Random Over-Sampling and Random Under-Sampling).

### 3.2. Acquiring Landsat and Elevation Data, and Generating Complementary Data within the GEE Platform

The time-series of Landsat-8 surface reflectance Tier 1 products (ID: LANDSAT/LC08/C01/T1\_SR) with less than 10% cloud coverage between May and October 2019 were used in this study. A total of 9, 13, and 15 Landsat-8 scenes were processed for the Site-1, Site-2, and Site-3, respectively (refer to Appendix A for more information). From the available spectral bands of Landsat-8 image, six bands (i.e., Bands 2-7) were used in this study. A median function, which can remove noisy, very dark, and very bright pixels [39], was applied to produce a single Landsat-8 mosaic image for each experiment site. Several spectral indices, including Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Soil-Adjusted Vegetation Index (SAVI), and Normalized Difference Built-up Index (NDBI) (see Table 1) were also generated from Landsat-8 imagery to investigate the effect of the spectral indices on the overall classification accuracy and the accuracies of the minority classes. NDVI, which helps us for generating an image showing the relative biomass, has been applying broadly in LC mapping [40]. It has been proven that NDBI along with NDVI are effective indices for identifying urban built-up area and discriminating them from other land cover types (e.g., trees and grassland) [41]. The NDWI allows scholars to recognize water bodies from other objects such as soil and terrestrial vegetation features [42]. SAVI can help us to discriminate soil-vegetation systems [43]. Furthermore, the Shuttle Radar Topography Mission (SRTM) data, which is available in the GEE platform (ID: USGS/SRTMGL1\_003), was applied to generate complementary topographic information, including elevation, slope, and aspect. The effects of these elevation products were also investigated on the accuracy of classification.

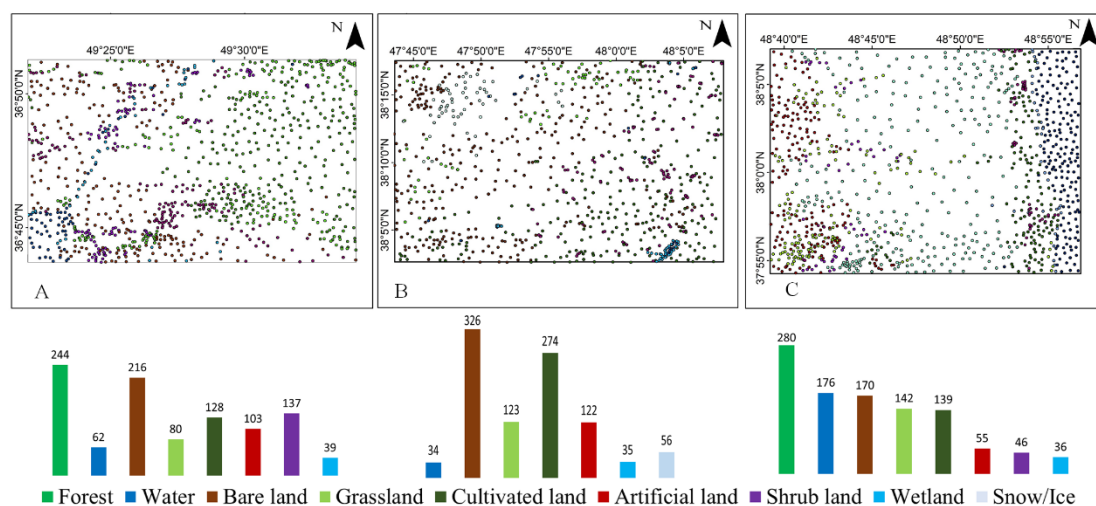
**Table 1.** List of generated spectral indices and their corresponding formulas (NIR = near infrared, SWIR = short wave infrared, L = 0.428).

Name	Formula	Reference
NDVI	$(\text{NIR} - \text{Red})/(\text{NIR} + \text{Red})$	[44]
NDWI	$(\text{Green} - \text{NIR})/(\text{Green} + \text{NIR})$	[45]
SAVI	$((\text{NIR} - \text{Red})/(\text{NIR} + \text{Red} + \text{L})) \times (1 + \text{L})$	[43]
NDBI	$(\text{SWIR} - \text{NIR})/(\text{SWIR} + \text{NIR})$	[41]

### 3.3. Generating Reference Samples and Spitting Them into Majority, Middle, and Minority Groups

Collection of in-situ samples in mountainous areas is often labor-intensive and expensive. However, generating reliable reference datasets is a basic requirement for accurate supervised Land Cover (LC) classification. Therefore, the reference samples over the experiment sites were generated using accurate visual interpretation of very high spatial resolution images of Google Earth (Figure 3). The specifications of the LC Classification System, developed by the Food and Agriculture Organization of the United Nations [46], were considered in the generating reference samples. It includes nine LC types, including Forest, Grassland, Shrub land, Cultivated land, Artificial land, Water bodies, Wetland, Permanent snow/ice, and Bare land. Based on the distributions of MLC types, 1089, 970, and 1044 samples were, respectively, generated for Site-1, Site-2, and Site-3. It should be noted that MLC classes covering larger areas relatively received more samples. Finally, the generated reference samples were randomly divided into two groups training and validation (50% and 50%).

To split the generated reference samples into three different groups (i.e., Majority, Middle, and Minority), first, the Highest Number of Samples (HNS) among different classes in each experiment site (i.e., Forest class with 244 samples in Site-1, Bare land Class with 326 samples in Site-2, and Forest class with 280 samples in Site-3) were selected. Then, the class(es) with samples between 70% and 100% of HNS was grouped as the Majority Class; the class(es) with samples between 35% and 70% of the HNS was grouped as the Middle Class; and the class(es) with samples between 0% and 35% of the HNS was grouped as the Minority Class.



**Figure 3.** Left and right columns, respectively, show the number of the generated reference samples for each land cover class and their spatial distributions over each experiment site ((A) Site-1, (B) Site-2, and (C) Site-3).

### 3.4. Selecting Best Classification Scenario Based on the Optimum Features

Among the available ML algorithms, RF has been drawing considerable attention in LC mapping [8,47]. This is mainly due to its high performance, availability in different computing environments, and its low sensitivity to noisy data [48,49]. RF combines multiple decision trees to classify the input data [50,51]. Moreover, it takes and resamples the input dataset several times to avoid the overfitting problem [5,50]. To achieve the most accurate RF model, two main parameters should be accurately optimized: (1) the number of trees in the forest (ntree); (2) the number of variables available for splitting at each tree node (mtry). In this study, after multiple trial and errors to find the optimum values of these parameters, the ntree and mtry were set to 500 and the square root of the total number of input features, respectively.

Four well known spectral indices, including NDVI, NDWI, SAVI, and NDBI, along with topographic products features, including elevation, slope, and aspect were used to identify best classification scenario. The most optimum spectral and topographic features were selected based

on the results of RF classifications applied to four following scenarios. Additionally, the effects of the complementary datasets (i.e., spectral indices and topographic features) on the accuracy of MLC mapping, especially those of the minority classes, were investigated.

- Scenario 1: Time-series of Landsat images + original imbalanced data.
- Scenario 2: Time-series of Landsat images + spectral indices + original imbalanced data.
- Scenario 3: Time-series of Landsat images + topographic features + original imbalanced data.
- Scenario 4: Time-series of Landsat images + topographic features + spectral indices+ original imbalanced data.

After comparing the results of the four scenarios and selecting the optimal input features (i.e., scenario with the best result), the proposed PROSRUS method was implemented to address the class imbalance problem.

### 3.5. Applying PROSRUS Method

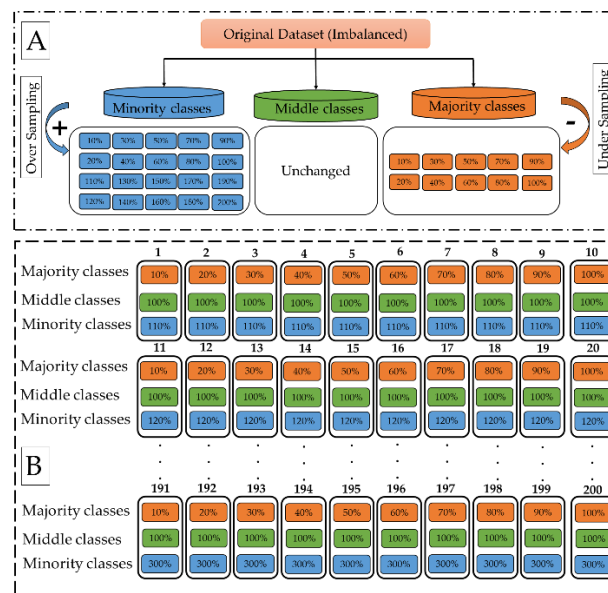
In this study, a hybrid data balancing method, called PROSRUS, was proposed. The PROSRUS method combines two well-known data-level balancing methods, including ROS [24] and RUS [37]. ROS, as a straightforward oversampling technique, randomly duplicates samples from minority class(es) to balance the distribution of classes. Fully balancing of an original imbalanced dataset using this method could cause overfitting of the classifier because of the duplication [52]. On the other hand, RUS randomly deletes samples from the majority class(es) to adjust the data distribution. The main shortcoming of a fully balancing dataset using RUS is that it may miss valuable information [23].

The proposed hybrid method not only takes the advantages of both ROS and RUS, but also limits their disadvantages by examining 200 different fractions in the balancing scheme. More specifically, as shown in Figure 4, original data were initially divided into three following groups based on the number of samples of different LC classes: Group-1 (minority classes), Group-2 (middle classes), and Group-3 (majority classes). Subsequently, after multiple trial and errors, 200 different fractions (it is possible to define any other preferred fractions) are employed for balancing LC classes to extract the optimal fraction(s) among them. In this way, as a partial balancing approach, ROS was used for oversampling samples in Group-1, and RUS was applied for under-sampling in Group-3, while samples of Group-2 were unchanged. For example, in fraction-1, only 10% of samples from Group-3 (90% of samples removed using RUS), 100% of Group-2 (unchanged), and 110% of Group-1 (10% new samples added using ROS) were contributed to the balancing process. The code for applying PROSRUS in the GEE platform is available in the Supplementary Material.

### 3.6. Accuracy Assessment and Comparison

The accuracy of obtained MLC maps using the proposed PROSRUS method were evaluated using the OA, User's Accuracy (UA), and Producer's Accuracy (PA) measures. Since OA is affected by majority classes rather than the minority ones [25], the Geometric Mean (G-Mean) index was also applied for accuracy assessment. G-Mean is particularly suitable for the evaluation of a classification with a class imbalance problem with more focusing on the accuracy of minority classes [53]. Accordingly, the G-Mean of PA (GM-PA) and G-Mean of UA (GM-UA) were also calculated.

The results of PROSRUS were also compared with those of the four well-known balancing techniques, including ROS, RUS, SMOTE, and G-SMOTE. To this end, RF along each of these data balancing techniques were applied to the optimum features (i.e., best scenario discussed in Section 3.4). For comparison purposes, the methods were named as RF-PROSRUS, RF-ROS, RF-RUS, RF-SMOTE, and RF-G-SMOTE.



**Figure 4.** (A) The schematic representation of separating the original imbalanced data into different classes. (B) An overview of the defined fractions for rebalancing the original reference dataset.

#### 4. Results

After grouping the LC classes based on the number of samples over each experiment site (Table 2), the impacts of different complementary information and different balancing techniques in MLC classes were investigated as follows:

**Table 2.** Grouping the land cover classes based on the number of samples over each experiment site.

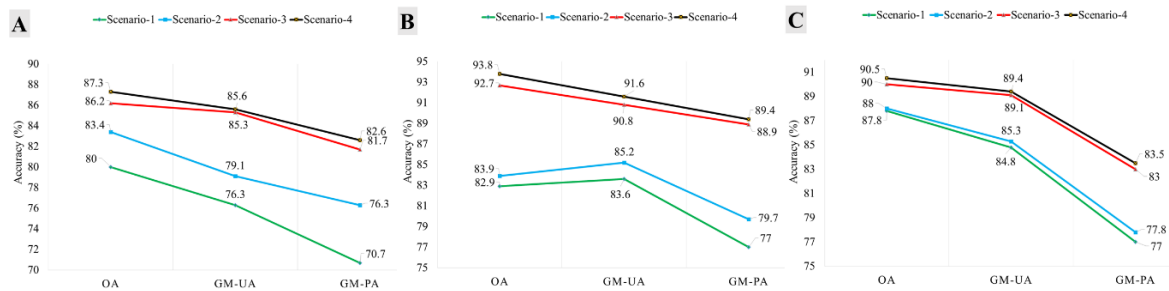
Site	Group-1 (Minority Classes)	Group-2 (Middle Classes)	Group-3 (Majority Classes)
1	Wetland, Water bodies, Grassland	Shrubland, Cultivated land, Artificial land	Forest, Bare land
2	Water bodies, Snow, Wetland	Artificial land, Grassland	Cultivated land, Bare land
3	Artificial land, Shrubland, Wetland	Water bodies, Bare land, Grassland, Cultivated land	Forest

##### 4.1. Optimum Classification Scenario

The effects of different complementary information, such as spectral indices (see Table 1) and topographic data (elevation, slope, and aspect) on the accuracy of minority classes in mountainous areas were investigated using four different classification scenarios explained in Section 3.2. As it is clear from Figure 5, including complementary information considerably improved the accuracy of MLC classification, particularly minority classes. Scenario-4 (time-series of Landsat images + original imbalanced data + topographic features + spectral indices) resulted in the highest accuracy. The OAs, GM-UAs, and GM-PAs of this classification scenario, respectively, ranged between 87.3%–93.8%, 85.6%–91.6%, and 82.6%–89.4% over the three experiment sites. As shown in Figure 5, all three overall accuracy assessment metrics (i.e., OA, GM-UA, and GM-PA), generally had the highest values using Scenario-4. For example, in Site-1, OA, GM-UA, and GM-PA, respectively, increased from 80% to 87.3%, 76.3% to 85.6%, and 70.7% to 82.6% compared to when only spectral bands of Landsat-8 were used (i.e., Scenario-1).

Although both topographic features (Scenario-3) and spectral indices (Scenario-2) improved all three accuracy assessment metrics for simple RF (Scenario-1), topographic data had higher impacts than spectral indices on improving MLC classification results (see Figure 5). The OAs, GM-UAs, and GM-PAs of Scenario-3, respectively, ranged between 86.2%–92.7%, 85.3%–90.8%, and 81.7%–88.9%

over the three experiment sites. Moreover, the OAs, GM-UAs, and GM-PAs of Scenario-2, respectively, ranged between 83.4%–88%, 79.1%–85.3%, and 76.3%–79.7% over the three experiment sites.



**Figure 5.** Mountain Land Cover (MLC) classification results based on the Random Forest (RF) algorithm and imbalanced samples over Site-1 (A), Site-2 (B), and Site-3 (C) using different classification scenarios (see Section 3.2 for the explanations of the scenarios).

Regarding different MLC types, minority classes showed stronger responses to including topographic and spectral features (Table 3). For example, regarding UA values, the highest improvement compared to Scenario-1, were observed in two (out of three) experiment sites for the minority classes: Grassland class in Site-1 (18.9%), and Wetland class in Site-3 (17.8%). According to the PA values, the highest improvement also achieved by minority classes as follows: Wetland class in Site-1 (31.2%), Wetland class in Site-2 (18.7%), and Artificial land class in Site-3 (28.5%). The results indicated that including complementary information to the classification procedure was necessary to improve not only the overall classification accuracy but also the individual class accuracies, especially those of the minority MLC types.

**Table 3.** Effects of the Scenario-4 on User’s Accuracy (UA) and Producer’s Accuracy (PA) values over three experiment sites. The increased and decreased in the accuracies are indicated by + and – signs, respectively (refer to Appendix B for more detailed information).

Sites	Evaluation Metrics (per Class)	LC Classes								
		Artificial Land	Bare Land	Cultivated Land	Forest	Grassland	Shrub Land	Water Bodies	Wetland	Snow
Site-1	UA (%)	+1.3	+3.8	+17.2	0	+18.9	+14.5	0	+14.6	none
	PA (%)	+5.9	−1.9	+28.9	+6	+5.4	+12.2	0	+31.2%	none
Site-2	UA (%)	+8.1	+11.8	+9.9	none	+22.4	none	0	+3.3	0
	PA (%)	+9.8	+5.9	+11.8	none	−0.2	none	0	+18.7	+9.6
Site-3	UA (%)	+4.9	+8.1	+3.2	0	+3.6	+7.7	0	+17.8	none
	PA (%)	+28.5	+2.6	+9.1	0	+6.1	+4.6	0	+6	none

#### 4.2. Comparison of Balancing Techniques

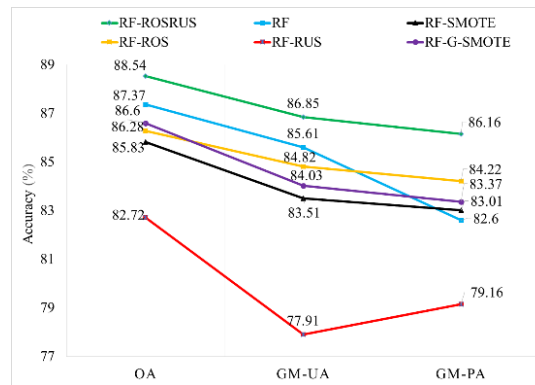
The proposed method along with four balancing techniques (i.e., ROS, RUS, SMOTE, and G-SMOTE) were applied over three experiment sites to study the impact of different balancing techniques on the accuracy of MLC classification. The results are these investigations are discussed in the following.

##### 4.2.1. Site-1

In Site-1, the proposed PROSRUS with the fraction numbers of 190 showed the best performance (Figure 6). This fraction used 210%, 100%, and 100% of Group-1 (minority classes), Group-2 (middle classes), and Group-3 (majority classes), respectively. As is clear from Figure 6, in comparison to Scenario-4 with imbalanced samples, it respectively improved GM-PA, GM-UA, and OA values by approximately 3.5%, 1.2%, and 1.2%. This proved the high potential of the proposed balancing method to provide high accuracies for both majority and minority classes. RF-G-SMOTE yielded the

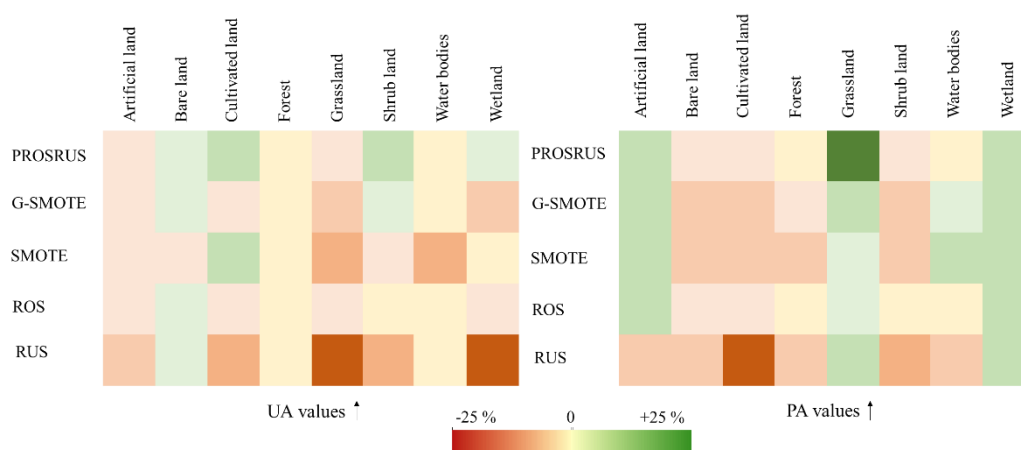


second-best result by providing OA = 86.6%, GM-UA = 84.03%, and GM-PA = 83.01%. Unlike the PROSRUS-190 and RF-G-SMOTE that increased all three overall metrics, the price for increasing the accuracy of minority classes was a reduction in the OA values for the other three resampling techniques (i.e., RF-SMOTE, RF-RUS, and RF-ROS). For example, although RF-ROS increased the value of GM-PA by approximately 1.6%, it reduced OA by approximately 1.1%. This amount for RF-SMOTE was even higher (i.e., a decrease of 1.5% in OA).



**Figure 6.** Accuracy assessment of the five resampling methods and the Scenario-4 classification method (i.e., RF with imbalanced samples) over Site-1.

Regarding individual MLC classes, the fraction numbers of 190 of PROSRUS improved the UA values of four classes (out of eight), including Wetland (1.2%), Bare land (2.8%), Cultivated land (3.9%), and Shrub land (4.6%) (see Figure 7). However, two classes of Artificial land (0.9%) and Grassland (1.9%) experienced downtrends. RF-G-SMOTE, as the second-best method, improved UA values of the Bare land class by 0.5% and the Shrub land class by 2.8%, while decreased UA values for four classes, including Artificial land (1.3%), Grassland (10%), Cultivated land (0.1%), and Wetland (7.5%). Regarding the PA values, the proposed method improved the values of three classes (Artificial land = 5.9%, Grassland = 18.5%, and Wetland = 6.3%). However, the PA values of the Bare land, Cultivated land, and Shrub land classes, respectively, decreased by 1.9%, 3.3%, and 1.3% using the proposed RF-PROSRUS method.

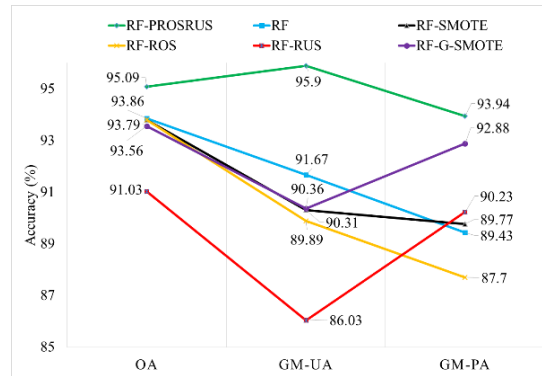


**Figure 7.** Effects (increase or decrease) of different balancing methods on the UA and PA values of various classes in Site-1. Refer to Appendix C for more information

#### 4.2.2. Site-2

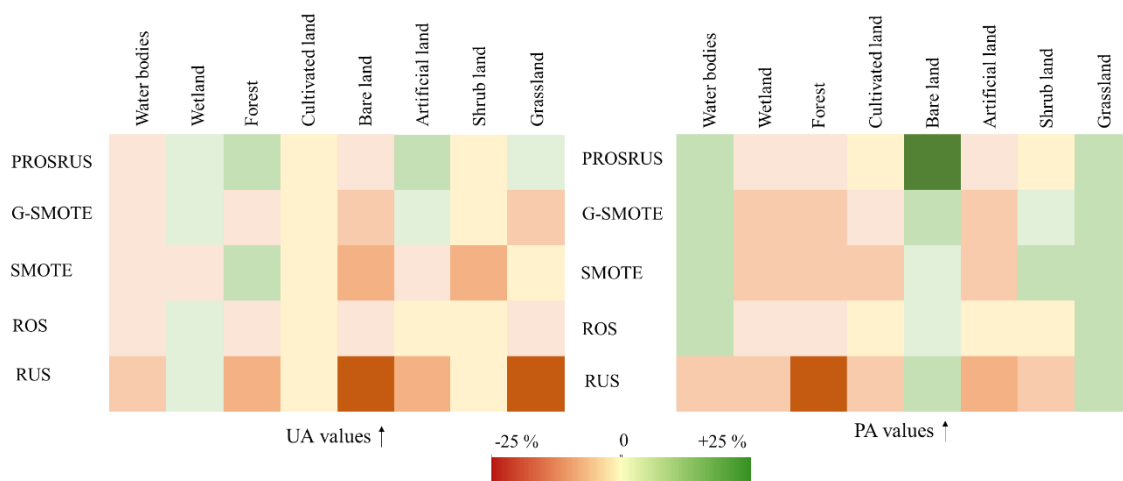
In Site-2, the fraction numbers of 26 using 130% of Group-1 (minority classes), 100% of Group-2 (middle classes), and 70% of Group-3 (majority classes), showed the best performance (Figure 8).

The proposed method, respectively, increased GM-PA, GM-UA, and OA values by approximately 4.5%, 4.2%, and 1.2% in comparison to Scenario-4. This confirmed the high potential of PROSRUS in dealing with the class imbalance problem. Similar to Site-1, RF-G-SMOTE showed the second-best results by providing OA = 93.86%, GM-UA = 91.67%, and GM-PA = 92.88%.



**Figure 8.** Accuracy assessment of the five resampling methods and the Scenario-4 classification method (i.e., RF with imbalanced samples) over Site-2.

PROSRUS-26, improved the UA values of four classes (out of seven) of Water bodies (14.3%), Wetland (7.5%), Cultivated land (0.6%), and Bare land (1.7%). However, it decreased the UA values of the Artificial land and Grassland classes by approximately 1.2% and 1.7%, respectively (Figure 9). The RF-G-SMOTE algorithm, which provided the second-best performance, increased UA values for the Water bodies (11.5%) and Bare land (1.4%) classes and decreased PA values for the Cultivated land (1%), Wetland (2%), and Artificial land (7.1%) classes. In the case of PA values, PROSRUS-26 increased the accuracies of three classes of Water bodies (7.2%), Artificial land (5.6%), and Wetland (18.8%), while decreased the accuracies of the Bare land class by 1.1%. Moreover, RF-G-SMOTE increased the accuracies of Water bodies (16.2%), Cultivated land (2.3%), and Artificial land (10.6%), while it decreased the accuracies of the other four classes (Snow = 4%, Bare land = 6.2%, Wetland = 2.7%, and Grassland = 8.8%).

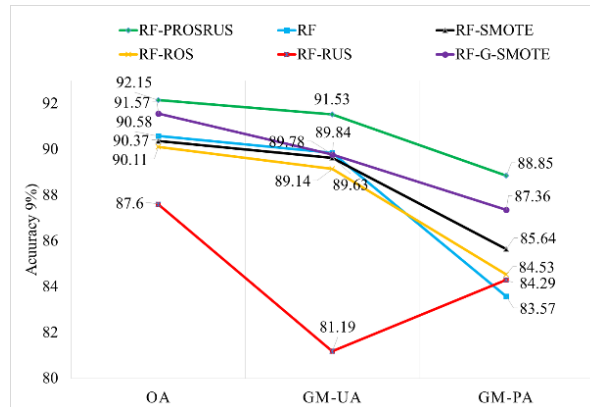


**Figure 9.** Effects (increase or decrease) of different balancing methods on the UA and PA values of various classes in Site-2. Refer to Appendix D for more information

### 4.2.3. Site-3

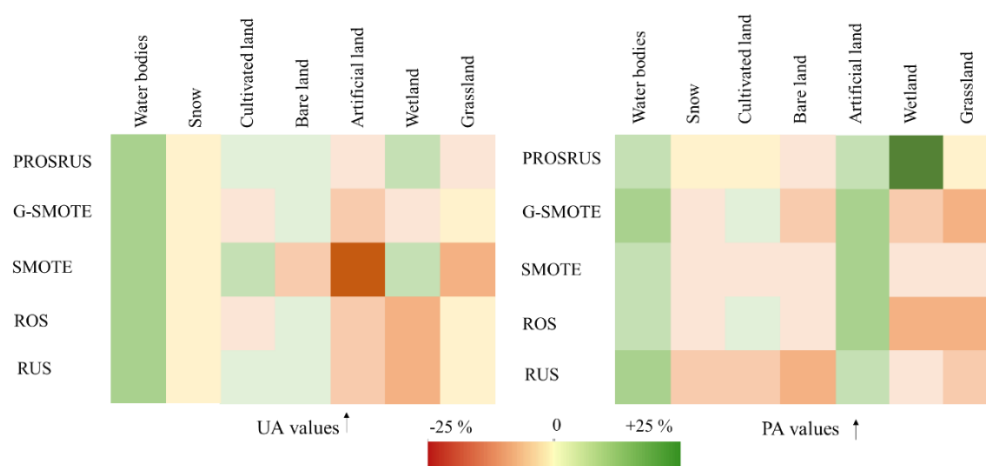
In Site-3, the fraction numbers of 74 using 180% of Group-1 (minority classes), 100% of Group-2 (middle classes), and 50% of Group-3 (majority classes) had the best performance in improving MLC

classification using the proposed method (Figure 10). Reaching to OA = 92.15, GM-UA = 91.53, and GM-PA = 88.85 in comparison to the Scenario-4, the proposed method increased these overall accuracy metrics by approximately 1.6%, 1.7%, and 5.3%, respectively. RF-G-SMOTE outperformed remaining three resampling methods and obtained the second-best place.



**Figure 10.** Accuracy assessment of the five resampling methods and the Scenario-4 classification method (i.e., RF with imbalanced samples) over Site-3.

Applying PROSRUS-74 led to an increase in the UA values for four classes (out of eight), including Grassland, Bare land, Forest, and Artificial land classes by 5%, 1.3%, 2.7%, and 6.8%, respectively. However, it decreased the accuracy of the Shrub land (−0.4) and Cultivated land (−2.2%) classes. RF-G-SMOTE, respectively, increased the UA values for the Forest, Cultivated land, Bare land, and Grassland classes approximately by 2%, 0.1%, 2.1%, and 3.38%. It however decreased the PA values of the other three classes. Regarding PA values, PROSRUS-74 improved the results of five classes compared to the Scenario-4 (Shrub land = 23%, Bare land = 2.4%, Cultivated land = 1.5%, Grassland = 1.5%, and Artificial land = 0.01%) (Figure 11). However, there was a reduction of 1.4% in the PA value of the Forest class using the proposed method.



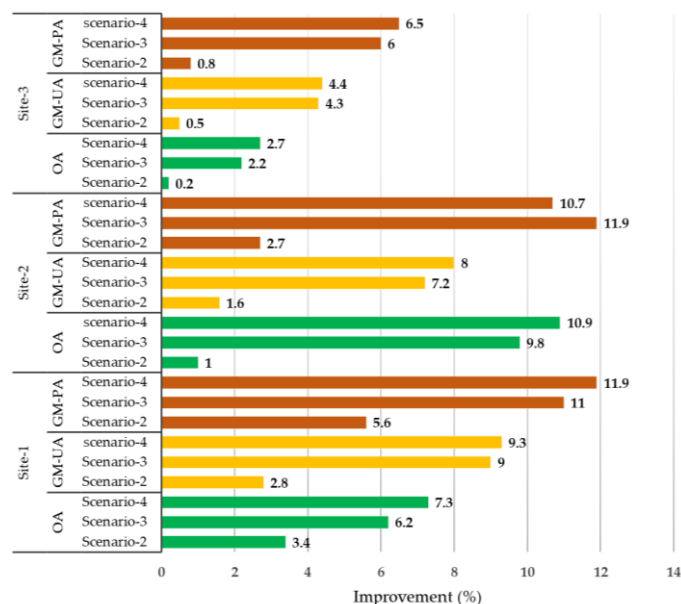
**Figure 11.** Effects (increase or decrease) of different balancing methods on the UA and PA values of various classes in Site-3. Refer to Appendix E for more information

### 5. Discussion

GEE has markedly improved the LC mapping studies by providing a huge number of geospatial datasets, in particular, the archive of Landsat data [54,55]. In this study, 37 Landsat-8 OLI scenes and SRTM data were used to study the potential of balancing methods on MLC classification. The GEE

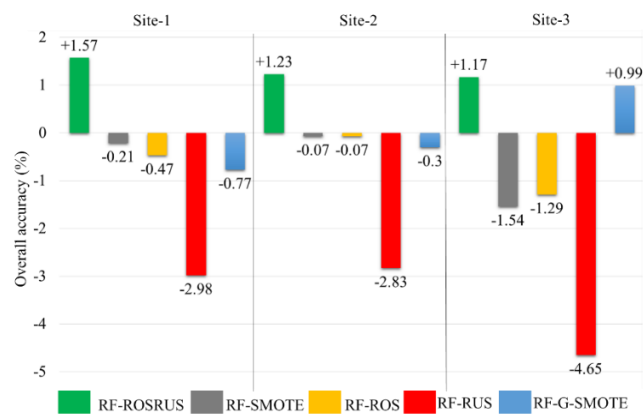
platform allowed us to have a faster and easier classification process because of several factors, such as providing atmospherically corrected time-series of Landsat data, high-performance computing capability, image-based functions, and integrated RF algorithm to the GEE API.

The experiments demonstrated the efficiency of adopting complementary information to improve the accuracy of MLC classification. We were able to increase the average OA, GM-UA, and GM-PA by 7%, 7.2%, and 10.2% using all spectral and topographic features (i.e., slope, elevation, aspect, NDVI, NDWI, NDBI, and SAVI), respectively. This can be explained by the fact that both topographic data and spectral indices provided important information, which in turn improved the MLC mapping accuracy [56,57]. By comparing the results of the four different scenarios over three experiment sites, it was observed that although Scenario-4 (i.e., integrating spectral indices and topographic data) showed the highest accuracies, the impact of topographic data was higher than the spectral indices in MLC classification (see Figure 12). This corresponded well to multiple studies, such as [29,58,59]. Based on Figure 5, among all three overall accuracy metrics, the GM-PA metric showed the highest improvement (Site-1 = 11.9%, Site-2 = 12.4%, and Site-3 = 6.5%) after adopting the complementary information. It can be concluded that including spectral and topographic features had bigger effects on the accuracy of minority classes.



**Figure 12.** Impacts of different Scenarios on MLC classification accuracy.

It was also observed that PROSRUS outperformed all other data balancing techniques, including ROS, RUS, SMOTE, and G-SMOTE. PROSRUS along with RF algorithm improved the average OA by approximately 1.3% considering all experiment sites (Figure 13). Higher improvements in the GM-UA and GM-PA values were even observed after adopting the proposed method (i.e., approximately by 1.8% and 4.6%, respectively). The reason might be attributed to two main factors as follows: (1) PROSRUS only duplicated samples from minority classes and did not generate artificial samples. Generating artificial samples by some balancing methods can sometimes lead to misclassification [24]; (2) PROSRUS partially balanced dataset to find the most optimal fraction(s) for addressing the class imbalance problem. This decreased the drawbacks of fully balancing datasets using ROS and RUS (e.g., overfitting for fully ROS [52] and losing critical information for RUS [60]).



**Figure 13.** Performance of different resampling methods in improving overall classification accuracy compared to the Scenario-4 classification method (i.e., RF with imbalanced samples).

Based on previous studies, improving the accuracy of minority classes usually leads to decrease in OA. For example, Waldner et al. [25] reported that “the price for increasing the accuracy of minority classes was a decrease in OA”. However, among all five resampling methods, PROSRUS was the only method that successfully improved the accuracy of minority classes without a reduction in OA in all experiment sites (i.e., Site-1 = 1.57%, Site-2 = 1.23%, and Site-3 = 1.17%). Our experiments also confirmed that G-SMOTE outperforms SMOTE in most cases, which was in agreement with [27], ROS had higher accuracies than RUS, which confirmed the findings of [61], and had lower accuracies than SMOTE, which was in the agreement with [16].

The experiments showed that a specific balancing ratio cannot provide optimal results in all datasets and settings. For example, fraction numbers of 190, 74, and 26 showed the best results among all applied 200 fractions over Site-1, Site-2, and Site-3, respectively. The reason that different datasets react differently to various fractions can be related to the issue that the imbalance ratio is different from a dataset to another one [25]. Therefore, it is necessary to investigate different fractions to achieve the most accurate MLC map.

## 6. Conclusions

In this study, a hybrid data balancing technique was proposed to address the class imbalance problem, which is a common problem in LC classification using ML algorithms. Additionally, the role of complementary information on MLC mapping was investigated. All the investigations were conducted over three different experiment sites using the time-series of Landsat-8 OLI within the GEE cloud computing platform. The study revealed the feasibility and reliability of improving the accuracy of LC classes in mountainous areas by adopting the RF classification algorithm, using both spectral and topographic features, and PROSRUS as a data balancing technique. The experiments also showed that topographic data including elevation, slope, and aspect had higher impacts than spectral indices in improving the accuracy of MLC maps. Moreover, it was illustrated that higher accuracies could be obtained for both minority and majority classes using an appropriate balancing ratio. Moreover, it was concluded that every dataset requires a specific balancing ratio to obtain the optimal result because the imbalance ratios and complexity levels are different for different datasets. In summary, since the performance of the proposed balancing method was substantially better than those of the RF with imbalanced data and four rebalancing techniques (i.e., ROS, RUS, SMOTE, and G-SMOTE), it was concluded that the integration of complementary information and PROSRUS method was a valid alternative practice that should be considered for LC classification in mountainous areas.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-4292/12/20/3301/s1>, S1: Scripts for investigating the role of different complementary information on the accuracies of MLC classes. S2: Scripts for implementing PROSRUS based on time-series of Landsat and the GEE platform.

**Author Contributions:** Conceptualization, A.N., A.L.; methodology, A.N.; software, A.N.; validation, A.L., J.B., G.L.; Data curation, A.N.; formal analysis, A.L., J.B., G.L., M.A.; writing—original draft preparation, A.N.; writing—review and editing, A.L., M.A.; supervision, A.L.; funding acquisition, A.L., J.B., G.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was jointly funded by the strategic priority research program of the Chinese Academy of Science (CAS) (XDA19030303), and the national natural science foundation project of china (41631180, 41701432, 41571373), the national key research and development program of China (NO. 2016YFA0600103, 2016YFC0500201-06), the CAS “light of west China” program, the youth innovation promotion association CAS (grant 2019365), and the CAS-TWAS president’s fellowship for international doctoral students.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Detailed information of the selected Landsat images (Title of all images: USGS Landsat 8 Surface Reflectance Tier 1).

ID	Site	WRS_Path	WRS_Row	Cloud Cover	Sensing Time
1	1	166	34	5.4	2019-06-27
2	1	166	34	3.92	2019-09-15
3	1	166	35	0.9	2019-06-11
4	1	166	35	0.13	2019-06-27
5	1	166	35	0.61	2019-07-13
6	1	166	35	2.61	2019-07-29
7	1	166	35	1.08	2019-08-14
8	1	166	35	3.81	2019-08-30
9	1	166	35	0.13	2019-09-15
10	2	167	33	2.21	2019-06-18
11	2	167	33	7.94	2019-07-04
12	2	167	33	2.78	2019-07-20
13	2	167	33	6.29	2019-08-05
14	2	167	33	0.72	2019-08-21
15	2	167	34	8.7	2019-05-01
16	2	167	34	6.75	2019-05-17
17	2	167	34	1.77	2019-06-02
18	2	167	34	1.96	2019-06-18
19	2	167	34	1.96	2019-07-04
20	2	167	34	1.74	2019-07-20
21	2	167	34	2.9	2019-08-05
22	2	167	34	1	2019-08-21
23	3	167	34	5.4	2019-06-27
24	3	166	34	3.92	2019-09-15
25	3	167	33	2.21	2019-06-18
26	3	167	33	7.94	2019-07-04
27	3	167	33	2.78	2019-07-20
28	3	167	33	6.29	2019-08-05
29	3	167	33	0.72	2019-08-21
30	3	167	34	8.7	2019-05-01
31	3	167	34	6.75	2019-05-17
32	3	167	34	1.77	2019-06-02
33	3	167	34	1.96	2019-06-18
34	3	167	34	1.96	2019-07-04
35	3	167	34	1.74	2019-07-20
36	3	167	34	2.9	2019-08-05
37	3	167	34	0.98	2019-08-21

## Appendix B

Table A2. The result of accuracy assessment using four different scenarios in the three experiment sites.

Experiment Sites	Scenarios	Evaluation Metrics (per Class)	LC Classes									Overall Metrics		
			Artificial Land	Bare Land	Cultivated Land	Forest	Grassland	Shrub Land	Water Bodies	Wetland	Snow	OA (%)	GM-UA (%)	GM-PA (%)
Site-1	1	UA (%)	83.7	79.6	56.3	97.8	61.1	72.5	100	66.6	-	80	76.3	70.7
		PA (%)	60.7	93.3	51.6	91.8	59.4	78.3	94.2	50	-			
	2	UA (%)	84	81.5	67.3	97.8	66.6	76.8	100	66.6	-	83.4	79.1	76.3
		PA (%)	72.5	92.3	55	98.5	64.8	85.1	91.4	62.5	-			
	3	UA (%)	79	78.5	72.5	97.8	82.7	89.3	100	86.6	-	86.2	85.3	81.7
		PA (%)	66.6	90.4	75	97.8	64.8	90.5	94.2	81.2	-			
	4	UA (%)	85	83.4	73.5	97.8	80	87	100	81.2	-	87.3	85.6	82.6
		PA (%)	66.6	91.4	83.3	97.8	64.8	90.5	91.4	81.2	-			
Site-2	1	UA (%)	88.8	84.7	79.4	-	71.7	-	80	83.3	100	82.9	83.6	77
		PA (%)	77.7	91.7	84.5	-	85.9	-	80	62.5	90.4			
	2	UA (%)	89	85.4	84.7	-	76	-	80	83.3	100	83.9	85.2	79.7
		PA (%)	79.1	89.9	88	-	62.5	-	80	68.7	95.2			
	3	UA (%)	96.9	95.3	90.6	-	94.1	-	76.6	84.3	100	92.7	90.8	88.9
		PA (%)	86.8	96.4	96.4	-	80.7	-	80	83.5	100			
	4	UA (%)	96.9	96.5	90.8	-	94.1	-	78.5	86.6	100	93.8	91.6	89.4
		PA (%)	87.5	97.6	97.8	-	85.7	-	78.5	81.2	100			
Site-3	1	UA (%)	84.4	80	85.3	94.4	73.5	76.9	98.8	82.2	-	87.8	84.8	77
		PA (%)	60.7	84.4	84.8	97.8	75.7	45.4	98.8	82.2	-			
	2	UA (%)	85	82.7	85.5	94.4	69.3	83.3	98.8	86.6	-	88	85.3	77.8
		PA (%)	60.7	84.7	89.4	98.5	80.3	45.4	98.8	82.2	-			
	3	UA (%)	89.3	87.2	85.1	94.4	76.1	84.6	98.8	100	-	90	89.1	83
		PA (%)	89.2	86.2	92.4	97.8	80.3	50	98.8	82.2	-			
	4	UA (%)	89.3	88.1	88.5	94.4	77.1	84.6	98.8	100	-	90.5	89.4	83.5
		PA (%)	89.2	87	93.9	97.8	81.8	50	98.8	88.2	-			

## Appendix C

Table A3. Accuracy assessment of the four data balancing methods and traditional RF (Scenario-4) over Site-1.

Method	Evaluation Metrics (per Class)	LC Classes								Overall Metrics		
		Artificial Land	Bare Land	Cultivated Land	Forest	Grassland	Shrub Land	Water Bodies	Wetland	OA (%)	GM-UA (%)	GM-PA (%)
RF-ROSRUS-190	UA (%)	84.1	86.2	77.4	97.8	78.1	91.6	100	82.3	88.54	86.85	86.16
	PA (%)	72.5	89.5	80	97.8	83.8	89.2	91.4	87.5			
RF-Scenario 4 (original data)	UA (%)	85	83.4	73.5	97.8	80	87	100	81.2	87.37	85.61	82.60
	PA (%)	66.6	91.4	83.3	97.8	64.8	90.5	91.4	81.2			
RF-SMOTE	UA (%)	80.3	79	80	97.8	72.9	85.1	94.2	81.2	85.83	83.51	83.01
	PA (%)	70.7	85.5	76.2	91.1	67.5	88.3	97	86.6			
RF-ROS	UA (%)	82.2	86	72.7	97.8	78.7	87	100	77.7	86.28	84.82	84.22
	PA (%)	72.5	87.6	80	97.8	70.2	90.5	91.4	87.5			
RF-RUS	UA (%)	81	85.1	63.5	97.8	62.2	80	100	63.6	82.72	77.91	79.16
	PA (%)	58.8	81.9	66.6	91.8	75.6	86.5	85.7	87.5			

## Appendix D

Table A4. Accuracy assessment of the four data balancing methods and traditional RF (Scenario-4) over Site-2.

Method	Evaluation Metrics (per Class)	LC Classes							Overall Metrics		
		Water Bodies	Snow	Cultivated Land	Bare Land	Artificial Land	Wetland	Grassland	OA (%)	GM-UA (%)	GM-PA (%)
RF-ROSRUS-26	UA (%)	100	100	91.4	98.2	95.7	94.1	92.3	95.09	95.9	93.94
	PA (%)	85.7	100	97.8	96.5	93.1	100	85.7			
RF-Scenario 4 (original data)	UA (%)	78.5	100	90.8	96.5	96.9	86.6	94.1	93.86	91.67	89.43
	PA (%)	78.5	100	97.8	97.6	87.5	81.2	85.7			
RF-SMOTE	UA (%)	88.8	100	97.3	89.8	73.3	91.4	84.2	93.79	90.31	89.77
	PA (%)	84.2	96	97	94.7	98.1	78.5	82			
RF-ROS	UA (%)	88.8	100	90.4	98	91.4	76.9	100	93.79	91.89	87.70
	PA (%)	84.2	96	100	94.7	98.1	71.4	74.3			
RF-RUS	UA (%)	90	100	93.2	97.8	91.3	76.9	100	91.03	86.03	90.23
	PA (%)	94.7	96	93.9	88.1	94.4	78.5	78.1			



## Appendix E

**Table A5.** Accuracy assessment of the four data balancing methods and traditional RF (Scenario-4) over Site-3.

Method	Evaluation Metrics (per Class)	LC Classes							Overall Metrics		
		Water Bodies	Snow	Cultivated Land	Bare Land	Artificial Land	Wetland	Grassland	OA (%)	GM-UA (%)	GM-PA (%)
RF-ROSRUS-26	UA (%)	100	100	91.4	98.2	95.7	94.1	92.3	95.09	95.9	93.94
	PA (%)	85.7	100	97.8	96.5	93.1	100	85.7			
RF-Scenario 4 (original data)	UA (%)	78.5	100	90.8	96.5	96.9	86.6	94.1	93.86	91.67	89.43
	PA (%)	78.5	100	97.8	97.6	87.5	81.2	85.7			
RF-SMOTE	UA (%)	88.8	100	97.3	89.8	73.3	91.4	84.2	93.79	90.31	89.77
	PA (%)	84.2	96	97	94.7	98.1	78.5	82			
RF-ROS	UA (%)	88.8	100	90.4	98	91.4	76.9	100	93.79	91.89	87.70
	PA (%)	84.2	96	100	94.7	98.1	71.4	74.3			
RF-RUS	UA (%)	90	100	93.2	97.8	91.3	76.9	100	91.03	86.03	90.23
	PA (%)	94.7	96	93.9	88.1	94.4	78.5	78.1			

## References

1. Friend, D.A. Mountain geography in 2002: The international year of mountains. *Geogr. Rev.* **2002**, *92*, iii–vi. [[CrossRef](#)]
2. Bian, J.; Li, A.; Lei, G.; Zhang, Z.; Nan, X. Global high-resolution mountain green cover index mapping based on landsat images and google earth engine. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 63–76. [[CrossRef](#)]
3. Chu, D. *Remote Sensing of Land Use and Land Cover in Mountain Region*; Springer: New York, NY, USA, 2020.
4. Adepoju, K.; Adelabu, S. Improved landsat-8 OLI and sentinel-2 MSI classification in mountainous terrain using machine learning on google earth engine. In Proceedings of the Biennial Conference of the Society of South African Geographers, Bloemfontein, South Africa, 1–5 October 2018; p. 5.
5. Ghorbanzadeh, O.; Valizadeh Kamran, K.; Blaschke, T.; Aryal, J.; Naboureh, A.; Einali, J.; Bian, J. Spatial prediction of wildfire susceptibility using field survey GPS data and machine learning approaches. *Fire* **2019**, *2*, 43. [[CrossRef](#)]
6. Moharrami, M.; Naboureh, A.; Gudiyangada Nachappa, T.; Ghorbanzadeh, O.; Guan, X.; Blaschke, T. National-scale landslide susceptibility mapping in Austria using fuzzy best-worst multi-criteria decision-making. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 393. [[CrossRef](#)]
7. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Aryal, J. Forest fire susceptibility and risk mapping using social/infrastructural vulnerability and environmental variables. *Fire* **2019**, *2*, 50. [[CrossRef](#)]
8. Amani, M.; Salehi, B.; Mahdavi, S.; Granger, J.E.; Brisco, B.; Hanson, A. Wetland classification using multi-source and multi-temporal optical remote sensing data in newfoundland and Labrador, Canada. *Can. J. Remote Sens.* **2017**, *43*, 360–373. [[CrossRef](#)]
9. Lei, G.; Li, A.; Bian, J.; Zhang, Z.; Jin, H.; Xi, N.; Wei, Z.; Wang, J.; Cao, X.; Tan, J. Land cover mapping in southwestern china using the HC-MMK approach. *Remote Sens.* **2016**, *8*, 305. [[CrossRef](#)]
10. Mahdavi, S.; Salehi, B.; Amani, M.; Granger, J.E.; Brisco, B.; Huang, W.; Hanson, A. Object-based classification of wetlands in Newfoundland and Labrador using multi-temporal PolSAR data. *Can. J. Remote Sens.* **2017**, *43*, 432–450. [[CrossRef](#)]
11. Rodríguez-Jeangros, N.; Hering, A.S.; Kaiser, T.; McCray, J.E. ScaMF–RM: A fused high-resolution land cover product of the Rocky Mountains. *Remote Sens.* **2017**, *9*, 1015. [[CrossRef](#)]
12. Kan, X.; Zhang, Y.; Zhu, L.; Xiao, L.; Wang, J.; Tian, W.; Tan, H. Snow cover mapping for mountainous areas by fusion of MODIS L1B and geographic data based on stacked denoising auto-encoders. *Comput. Mater. Contin.* **2018**, *57*, 49–68. [[CrossRef](#)]
13. Liu, C.; Huang, X.; Li, X.; Liang, T. MODIS fractional snow cover mapping using machine learning technology in a mountainous area. *Remote Sens.* **2020**, *12*, 962. [[CrossRef](#)]
14. Lei, G.; Li, A.; Bian, J.; Yan, H.; Zhang, L.; Zhang, Z.; Nan, X. OIC-MCE: A practical land cover mapping approach for limited samples based on multiple classifier ensemble and iterative classification. *Remote Sens.* **2020**, *12*, 987. [[CrossRef](#)]
15. Delalay, M.; Tiwari, V.; Ziegler, A.D.; Gopal, V.; Passy, P. Land-use and land-cover classification using sentinel-2 data and machine-learning algorithms: Operational method and its implementation for a mountainous area of Nepal. *J. Appl. Remote Sens.* **2019**, *13*, 014530. [[CrossRef](#)]
16. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2012**, *42*, 463–484. [[CrossRef](#)]
17. Mellor, A.; Boukir, S.; Haywood, A.; Jones, S. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 155–168. [[CrossRef](#)]
18. Azadbakht, M.; Fraser, C.S.; Khoshelham, K. Synergy of sampling techniques and ensemble classifiers for classification of urban environments using full-waveform LidAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 277–291. [[CrossRef](#)]

19. Feng, W.; Dauphin, G.; Huang, W.; Quan, Y.; Bao, W.; Wu, M.; Li, Q. Dynamic synthetic minority over-sampling technique-based rotation forest for the classification of imbalanced hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2159–2169. [[CrossRef](#)]
20. Liu, X.-Y.; Zhou, Z.-H. The influence of class imbalance on cost-sensitive learning: An empirical study. In Proceedings of the Sixth International Conference on Data Mining (ICDM'06), Hong Kong, China, 18–22 December 2006; IEEE: New York, NY, USA, 2006; pp. 970–974.
21. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
22. Khan, S.H.; Hayat, M.; Bennamoun, M.; Sohel, F.A.; Togneri, R. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 3573–3587.
23. Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [[CrossRef](#)]
24. Chawla, N.V. Data mining for imbalanced datasets: An overview. In *Data Mining and Knowledge Discovery Handbook*; Springer: New York, NY, USA, 2009; pp. 875–886.
25. Waldner, F.; Chen, Y.; Lawes, R.; Hochman, Z. Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sens. Environ.* **2019**, *233*, 111375. [[CrossRef](#)]
26. Feng, W.; Boukir, S.; Huang, W. Margin-based random forest for imbalanced land cover classification. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 June–2 August 2019; IEEE: New York, NY, USA, 2019; pp. 3085–3088.
27. Douzas, G.; Bacao, F.; Fonseca, J.; Khudinyan, M. Imbalanced learning in land cover classification: Improving minority classes' prediction accuracy using the geometric smote algorithm. *Remote Sens.* **2019**, *11*, 3040. [[CrossRef](#)]
28. Bogner, C.; Seo, B.; Rohner, D.; Reineking, B. Classification of rare land cover types: Distinguishing annual and perennial crops in an agricultural catchment in South Korea. *PLoS ONE* **2018**, *13*, e0190476. [[CrossRef](#)] [[PubMed](#)]
29. Hurskainen, P.; Adhikari, H.; Siljander, M.; Pellikka, P.; Hemp, A. Auxiliary datasets improve accuracy of object-based land use/land cover classification in heterogeneous savanna landscapes. *Remote Sens. Environ.* **2019**, *233*, 111354. [[CrossRef](#)]
30. Xie, S.; Liu, L.; Zhang, X.; Yang, J.; Chen, X.; Gao, Y. Automatic land-cover mapping using Landsat time-series data based on google earth engine. *Remote Sens.* **2019**, *11*, 3023. [[CrossRef](#)]
31. Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C.; Hobart, G.W. Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. *Remote Sens. Environ.* **2015**, *170*, 121–132. [[CrossRef](#)]
32. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
33. Eskandari, S.; Reza Jaafari, M.; Oliva, P.; Ghorbanzadeh, O.; Blaschke, T. Mapping land cover and tree canopy cover in Zagros forests of Iran: Application of sentinel-2, google earth, and field data. *Remote Sens.* **2020**, *12*, 1912. [[CrossRef](#)]
34. Amani, M.; Brisco, B.; Afshar, M.; Mirmazloumi, S.M.; Mahdavi, S.; Mirzadeh, S.M.J.; Huang, W.; Granger, J. A generalized supervised classification scheme to produce provincial wetland inventory maps: An application of google earth engine for big geo data processing. *Big Earth Data* **2019**, *3*, 378–394. [[CrossRef](#)]
35. Amani, M.; Mahdavi, S.; Afshar, M.; Brisco, B.; Huang, W.; Mohammad Javad Mirzadeh, S.; White, L.; Banks, S.; Montgomery, J.; Hopkinson, C. Canadian wetland inventory using google earth engine: The first map and preliminary results. *Remote Sens.* **2019**, *11*, 842. [[CrossRef](#)]
36. Raziei, T. Koppen-Geiger climate classification of Iran and investigation of its changes during 20th century. *J. Earth Space Phys.* **2017**, *43*, 419–439.
37. Batista, G.E.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [[CrossRef](#)]

38. Douzas, G.; Bacao, F. Geometric smote a geometrically enhanced drop-in replacement for smote. *Inf. Sci.* **2019**, *501*, 118–135. [[CrossRef](#)]
39. Ghorbanian, A.; Kakooei, M.; Amani, M.; Mahdavi, S.; Mohammadzadeh, A.; Hasanlou, M. Improved land cover map of Iran using sentinel imagery within google earth engine and a novel automatic workflow for land cover classification using migrated training samples. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 276–288. [[CrossRef](#)]
40. Naboureh, A.; Moghaddam, M.H.R.; Feizizadeh, B.; Blaschke, T. An integrated object-based image analysis and CA-Markov model approach for modeling land use/land cover trends in the Sarab plain. *Arab. J. Geosci.* **2017**, *10*, 259. [[CrossRef](#)]
41. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from tm imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [[CrossRef](#)]
42. Yang, X.; Zhao, S.; Qin, X.; Zhao, N.; Liang, L. Mapping of urban surface water bodies from sentinel-2 MSI imagery at 10 m resolution via NDWI-based image sharpening. *Remote Sens.* **2017**, *9*, 596. [[CrossRef](#)]
43. Huete, A. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
44. Rouse, J.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309.
45. McFeeters, S.K. The use of the normalized difference water index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [[CrossRef](#)]
46. Cord, A.; Conrad, C.; Schmidt, M.; Dech, S. Standardized FAO-LCCS land cover mapping in heterogeneous tree savannas of West Africa. *J. Arid Environ.* **2010**, *74*, 1083–1091. [[CrossRef](#)]
47. Ghimire, B.; Rogan, J.; Miller, J. Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sens. Lett.* **2010**, *1*, 45–54. [[CrossRef](#)]
48. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* **2016**, *187*, 156–168. [[CrossRef](#)]
49. Phiri, D.; Morgenroth, J.; Xu, C.; Hermosilla, T. Effects of pre-processing methods on Landsat oli-8 land cover classification using obia and random forests classifier. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 170–178. [[CrossRef](#)]
50. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
51. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
52. Santos, M.S.; Soares, J.P.; Abreu, P.H.; Araujo, H.; Santos, J. Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]. *IEEE Comput. Intell. Mag.* **2018**, *13*, 59–76. [[CrossRef](#)]
53. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [[CrossRef](#)]
54. Huang, H.; Chen, Y.; Clinton, N.; Wang, J.; Wang, X.; Liu, C.; Gong, P.; Yang, J.; Bai, Y.; Zheng, Y. Mapping major land cover dynamics in Beijing using all Landsat images in google earth engine. *Remote Sens. Environ.* **2017**, *202*, 166–176. [[CrossRef](#)]
55. Carrasco, L.; O’Neil, A.W.; Morton, R.D.; Rowland, C.S. Evaluating combinations of temporally aggregated sentinel-1, sentinel-2 and Landsat 8 for land cover mapping with google earth engine. *Remote Sens.* **2019**, *11*, 288. [[CrossRef](#)]
56. Gbodjo, Y.J.E.; Ienco, D.; Leroux, L. Toward spatio-spectral analysis of sentinel-2 time series data for land cover mapping. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 307–311. [[CrossRef](#)]
57. Stromann, O.; Nascetti, A.; Yousif, O.; Ban, Y. Dimensionality reduction and feature selection for object-based land cover classification based on sentinel-1 and sentinel-2 time series using google earth engine. *Remote Sens.* **2020**, *12*, 76. [[CrossRef](#)]
58. Tsai, Y.H.; Stow, D.; Chen, H.L.; Lewison, R.; An, L.; Shi, L. Mapping vegetation and land use types in Fanjingshan national nature reserve using google earth engine. *Remote Sens.* **2018**, *10*, 927. [[CrossRef](#)]
59. Zhu, Z.; Gallant, A.L.; Woodcock, C.E.; Pengra, B.; Olofsson, P.; Loveland, T.R.; Jin, S.; Dahal, D.; Yang, L.; Auch, R.F. Optimizing selection of training and auxiliary data for operational land cover classification for the l cmap initiative. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 206–221. [[CrossRef](#)]

60. Choi, J.M. A selective sampling method for imbalanced data learning on support vector machines. *Grad. Theses Diss.* **2010**. [[CrossRef](#)]
61. Johnson, J.M.; Khoshgoftaar, T.M. Deep learning and data sampling with imbalanced big data. In Proceedings of the 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), Los Angeles, CA, USA, 30 July–1 August 2019; IEEE: New York, NY, USA, 2019; pp. 175–183.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).