

Article

# Automatic Case-Based Reasoning Approach for Landslide Detection: Integration of Object-Oriented Image Analysis and a Genetic Algorithm

Jie Dou <sup>1,\*</sup>, Kuan-Tsung Chang <sup>2,\*</sup>, Shuisen Chen <sup>3</sup>, Ali P. Yunus <sup>1</sup>, Jin-King Liu <sup>4</sup>, Huan Xia <sup>5</sup> and Zhongfan Zhu <sup>6</sup>

<sup>1</sup> Department of Natural Environmental Studies, The University of Tokyo, Kashiwa 277-8568, Japan; E-Mail: yunusp@csis.u-tokyo.ac.jp

<sup>2</sup> Department of Civil Engineering and Environmental Informatics, Minghsin University of Science and Technology, Hsin-Chu 304, Taiwan

<sup>3</sup> Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou Institute of Geography, Guangzhou 510070, China; E-Mail: css@gdas.ac.cn

<sup>4</sup> LiDAR Technology Co., Zhubei, Hsin-Chu 302, Taiwan; E-Mail: jkliu@lidar.com.tw

<sup>5</sup> Guizhou University of Finance and Economics, Huaxi District, Guiyang 550025, China; E-Mail: xh240@126.com

<sup>6</sup> College of Water Sciences, Beijing Normal University, Xijiekouwai Street 19, Beijing 100875, China; E-Mail: zhuzhongfan1985@gmail.com

\* Authors to whom correspondence should be addressed; E-Mails: doujie@csis.u-tokyo.ac.jp or douj888@gmail.com (J.D.); ktchang1216@gmail.com (K.-T.C.); Tel.: +81-047-136-4306 (J.D.); +886-921-214-694 (K.-T.C.).

Academic Editors: Yu-Chang Chan and Prasad S. Thenkabail

Received: 13 November 2014 / Accepted: 3 April 2015 / Published: 13 April 2015

---

**Abstract:** This paper proposes an automatic method for detecting landslides by using an integrated approach comprising object-oriented image analysis (OOIA), a genetic algorithm (GA), and a case-based reasoning (CBR) technique. It consists of three main phases: (1) image processing and multi-image segmentation; (2) feature optimization; and (3) detecting landslides. The proposed approach was employed in a fast-growing urban region, the Pearl River Delta in South China. The results of detection were validated with the help of field surveys. The experimental results indicated that the proposed OOIA-GA-CBR (0.87) demonstrates higher classification performance than the stand-alone OOIA (0.75) method for detecting landslides. The area under curve (AUC) value was also higher than that of the

simple OOIA, indicating the high efficiency of the proposed landslide detection approach. The case library created using the integrated model can be reused for time-independent analysis, thus rendering our approach superior in comparison to other traditional methods, such as the maximum likelihood classifier. The results of this study thus facilitate fast generation of accurate landslide inventory maps, which will eventually extend our understanding of the evolution of landscapes shaped by landslide processes.

**Keywords:** genetic algorithm (GA); image classification; image segmentation; landslide detection; object-oriented image analysis (OOIA)

---

## 1. Introduction

Landslides are one of the most destructive natural hazards, and they often cause substantial damage to societies worldwide every year [1,2]. The intensity of landslides results in more substantial injuries and loss of life than any other type of natural disaster, including earthquakes, hurricanes, tsunamis, and floods [3]. This trend is likely to worsen in the future with the process of urbanization and economic development, deforestation, and the increased regional rainfall in landslide-prone areas caused by climate change [1]. Therefore, identifying landslide occurrences and their mitigation has become crucial and challenging in the field of hazard management and mitigation research [4–6]. Moreover, preparing landslide inventory maps is necessary to archive the extent of landslide phenomena in specific areas; to examine their spatial distribution and types, risk, vulnerability, recurrence, and statistical slope instability; and to investigate the evolution of landscapes controlled by landslide processes [4,7,8].

For a successful qualitative or quantitative landslide hazard evaluation, compiling a historical landslide-event inventory is particularly crucial for pre-disaster and post-disaster analysis [9,10]. Thus far, landslide inventory maps have largely been generated through visual interpretation of aerial photos or satellite images combined with extensive field surveys. However, such methods are labor-intensive and expensive and, therefore, inefficient for generating maps of large areas. Moreover, traditional map-generating techniques require prior knowledge about the involved hazard, and such techniques are highly subjective and have limited reproducibility [11]. By contrast, a semi-automated or automated classification approach can provide a scheme for addressing the aforementioned problems. Several studies have been conducted to detect and identify landslides [10,12]. Experiments based on the emerging technique of expert-based knowledge systems (EKSs) have been proposed; in these experiments, rules have been applied to classify and identify hazard prone areas. The rules are typically created from spectral, textural, and shape features [13]. This research has demonstrated that EKS performance is higher than that of traditional per-pixel approaches in classifying land cover types. The advantage of EKSs is in task specific knowledge; however, a limitation of implementing EKS methods is that identifying and defining rules for each separate problem is tedious and time consuming [14,15].

With the advancement in the spatial resolution of commercial satellite images (e.g., SPOT 5, ALOS, IKONOS, QuickBird, and Worldview1), traditional pixel-based methods have become inapt for identifying and characterizing landslides. Furthermore, post-classification processes in pixel-based classification are tedious. In contrast to the pixel-based method, the object-oriented image analysis

(OOIA) method aggregates pixels into a network of homogeneous objects corresponding to surface cover patches [16]. Landscape shape and textural features are prominent in high-resolution satellite images; however, the spectral data range is narrow. Nevertheless, OOIA has several advantages over the pixel-based method. For instance, OOIA combines shape and context information with spectral and textural information simultaneously, thus preventing the “salt and pepper” effect prevalent in the pixel-based classification method [17–19]. Additionally, OOIA provides a potentially automated method for detecting landslides and can consider the spectral, morphological, and contextual properties of landslides according to expert knowledge [10,20].

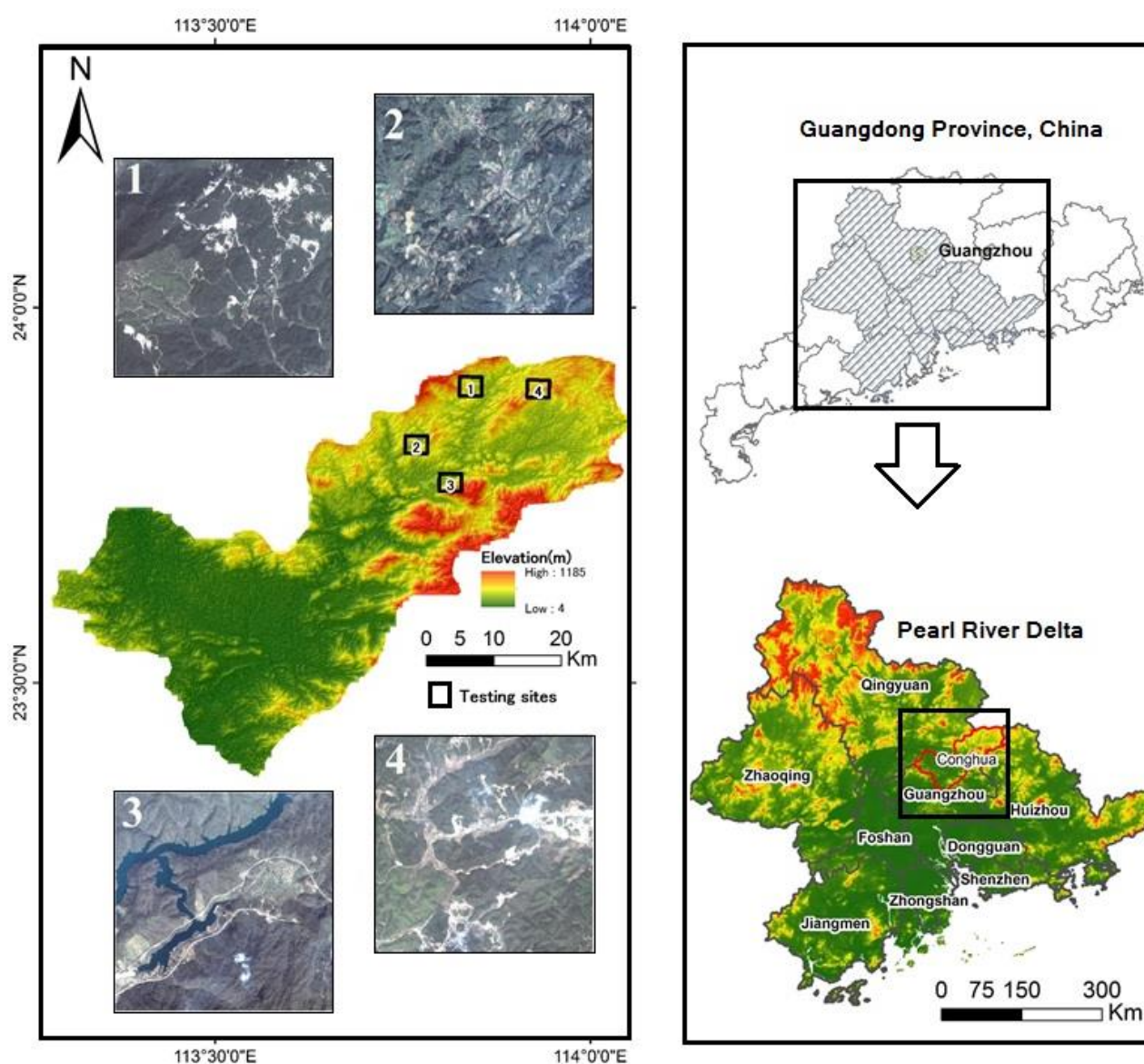
This paper proposes a relatively new approach for detecting landslides from high-spatial-resolution images by integrating three-phase object-oriented classification. The proposed approach is a combination of segmentation and feature optimization that involves using a genetic algorithm (GA). eCognition software was first used to segment QuickBird images and to extract spectral and textural features. The resulting objects were exported for further analysis. To reduce redundant data, the most appropriate features related to a landslide occurrence must be used; thus, a GA was applied. The GA method is considered to be a powerful tool for addressing the feature optimization problem because of its robustness. The GA method has been successfully implemented in several fields, such as feature selection in computational analysis, and classification of remote sensing images [21,22]. In this study, the optimization process was based on the optimal fitness value of landslide detection. The GA-driven feature optimization procedure offers several feature combinations for subsequent landslide detection. In the final phase, a relatively new approach was used for detecting landslides from high-spatial-resolution images based on case-based reasoning (CBR) techniques. CBR differs from EKSs, which require high levels of computing competence in cognitive tasks. Although EKSs are heuristic with judgments as well as with formal knowledge of established theories, they fail when applied to incomplete or inexact data because the systems accommodate more power than the user. Conversely, CBR is more similar to human perception, which uses knowledge derived from previous situations to solve new problems. CBR has been applied in various fields such as finance, marketing, and engineering [19,23–25]. In this study, CBR was used to detect landslides for mapping image objects. CBR approaches were generated based on the optimal feature combination to obtain the detection results. The experiment was conducted in a fast-growing urban area in the Pearl River Delta (PRD), China. According to a review of the literature, CBR integrated with OOIA, and GA has never been used to detect landslides because it is a relatively new technique.

## 2. Materials and Data Collection

### 2.1. Geography of the Study Area

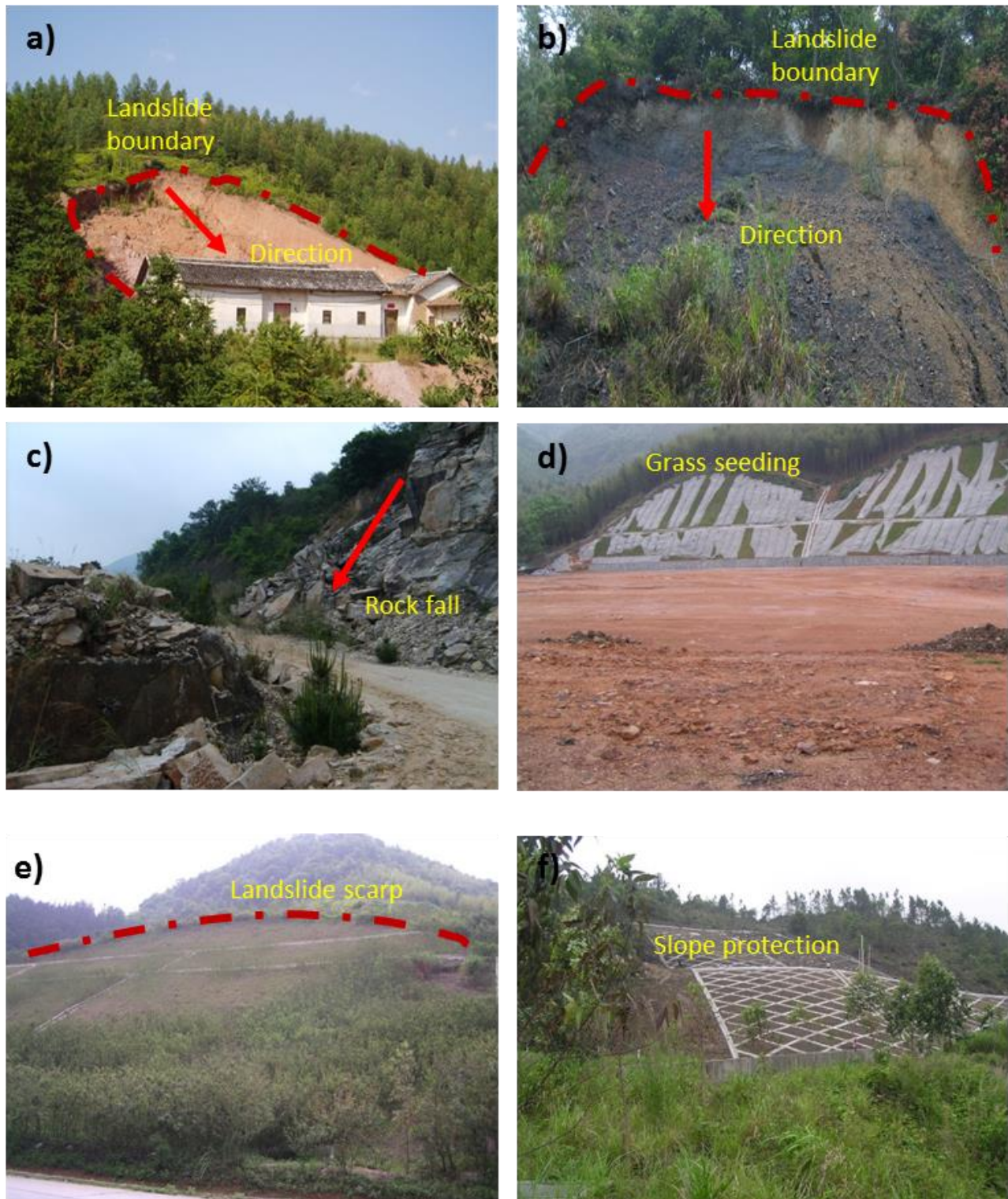
Conghua is located in a subtropical region between 113 °44'E to 113 °59'E and 23 °43'N to 23 °51'N in the center of Guangdong Province, South China (Figure 1). It is a district of Guangzhou City, which connects the PRD with the mountainous area in the north of Guangdong Province. Conghua district is known as “Guangzhou Garden” because of its vast forest coverage of approximately 67%, a total area of 1974.5 km<sup>2</sup>, across the territory of the Tropic of Cancer. The average temperature is approximately 21 °C. Nearly half of the study area is covered by mountains and highlands. In the past few years, the study area has experienced radical land-use/land-cover changes in response to the fast economic boom,

and it is one of the fastest developing regions in China. The rainfall in this area is abundant at approximately 2176 mm annually. Because of heavy rainfall, a complex geological environment, and intense human activities in this area, landslides have recently been occurring frequently (Figure 2) resulting in considerable economic losses from road damage and destruction of fundamental infrastructure [12]. In particular, the road section near National Highway 105 (NH 105) is frequently damaged by landslides, causing injury to people and severe destruction of buildings and property. Principal landslide types observed in Conghua are shown in Figure 2. The majority of them fall under the categories of debris slides, soil creep, and rock falls, according to Varnes [26]. Evidence of previous landslide scarps have been observed along NH 105 in the form of terraced off, but covered with little sharp boundary in the vegetation and concrete structures that were built for slope stabilization to arrest future failures (Figure 2d–f). Hence, timely identification of the spatial distribution of landslides and mitigating hazards is an urgent problem in this region. This information can be beneficial to urban planners and survey planning and management to prevent disasters and reduce economic losses.



**Figure 1.** The spatial location of the study area overlaying the elevation.





**Figure 2.** Examples of landslides in the study area. Principal landslide types observed in Conghua: (a) debris slide threatening a house; (b) soil creep; and (c) rock fall. Evidence of previous landslide scarps in the form of (d) terraced off, but covered with little, sharp boundary in the vegetation, and (e,f) concrete structures built for slope stabilization to arrest future failures.

## 2.2. Data Collection

High-resolution imagery is crucial for obtaining detailed landslide information. High-resolution Quickbird images of the study area with a spatial resolution of 0.6 m in three spectral bands in the visible wavelength, acquired on 10 October 2006 and 20 March 2007, were used to detect landslides by adopting image segmentation and feature selection methods. They enabled excellent recognition of the shape and location of objects. The main task was investigating landslides along the road because it is vulnerable to landslide damage. A SPOT 5 satellite image acquired on 7 December 2006 and DEM were used as ancillary data. Table 1 lists the specifications and details of data acquisition. We conducted field surveys in the same season as the image acquisition periods to support and validate our results. Figure 2 shows the field photographs of landslide locations.

**Table 1.** Spectral and spatial resolution of QuickBird and SPOT5 satellite image data.

Image Data	Pan ( $\mu\text{m}$ )	Red ( $\mu\text{m}$ )	Green ( $\mu\text{m}$ )	Blue ( $\mu\text{m}$ )	Near IR ( $\mu\text{m}$ )	Spatial Resolution	Date of Acquisition
QuickBird	0.45–0.9	0.63–0.69	0.52–0.6	0.45–0.52	0.78–0.9	Pan:0.61 m Ms:2.8 m	10 October 2006 and 20 March 2007
SPOT 5	0.48–0.71	0.61–0.68	0.5–0.59	Shortwave IR (1.58–1.75)	0.78–0.89	Pan 2.5 m Ms:10 m	7 December 2006

## 3. Methodology

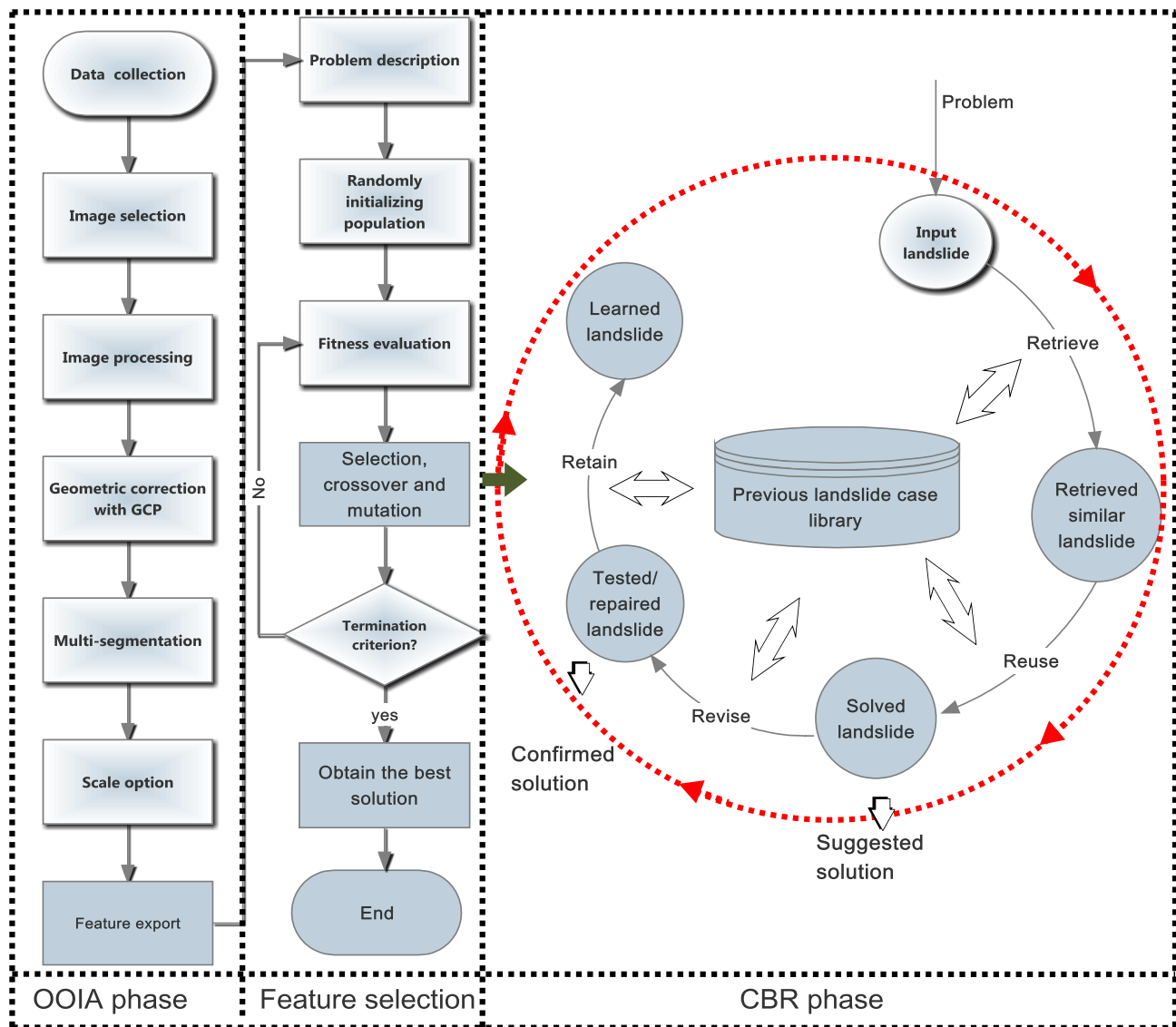
Figure 3 shows a flowchart of the integrated methodology used for the intelligent landslide detection in this study. The flowchart comprises three main phases: (1) multi-segmentation of images after data collection and preprocessing; (2) feature selection, using GAs based on the feature set; and (3) implementation of the CBR method to categorize the geomorphological features and validate the accuracy. Each phase is detailed in the subsequent sections.

### 3.1. Multiresolution Segmentation Technique

Multiresolution segmentation is a bottom-up segmentation algorithm based on a pairwise region-merging technique [27]. This algorithm generally groups image pixels that possess homogeneous spectral and textural characteristics. Smaller objects are combined into larger objects based on criteria determined by three parameters: scale, color, and shape (smoothness and compactness) [28]. The segmentation process continues until the smallest object growth exceeds a user-defined threshold. The heterogeneity (e.g., spectral and color) criterion can be defined as follows:

$$H = \sum_c w_c \left( n_{\text{merge}} \sigma_c^{\text{merge}} - \left( n_{\text{object1}} \cdot \sigma_c^{\text{object1}} + n_{\text{object2}} \cdot \sigma_c^{\text{object2}} \right) \right) \quad (1)$$

where  $H$  is an arbitrary heterogeneity criterion,  $w_c$  represents the weight attributed to each band,  $\sigma_c$  corresponds to the standard deviation of the spectral values within band  $c$ , and  $n$  denotes the number of pixels in a segment.

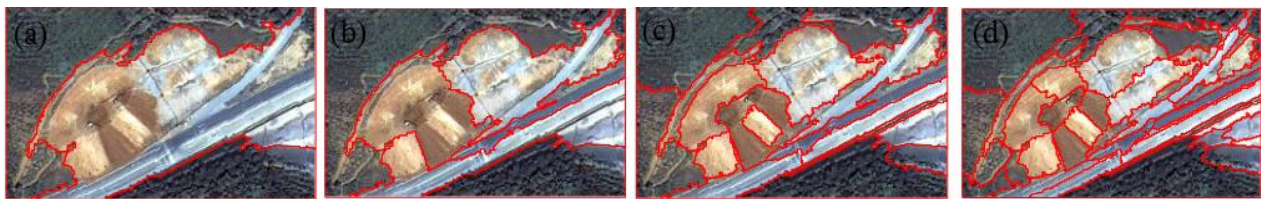


**Figure 3.** The integrated structure for automatic landslide detection, comprising three processes: (1) multi-segmentation by OOIA; (2) feature selection by GA; and (3) detection by CBR and validation using field work data.

To remove the image distortions in the different data sets, all the data were geometrically rectified based on ground control points (GCPs) by using the Autosync Workstation module of the Erdas 9.1 software. GCPs were acquired for eight locations with the help of the Trimble positioning system during the field survey. Additionally, the image geometry was evaluated by computing the root mean square error (RMSE). To ensure a high model quality over the image, the improper tie pointes were pruned after careful visual confirmation. The computed RMSE was less than one pixel (0.5 m) for the study area, which was considered satisfactory taking into account that the location is an undulating terrain. This step was crucial before characterizing the attributes of features of interest [20]. Figure 4 shows an example of image segmentation results from QuickBird images. The segmentation outcomes are determined by the spatial resolution of images and object features, and the segmentation scales are determined by the size of landslides. The segmentation scale as defined in OOIA is the maximum color difference within each chosen image layer inside square image objects [27]. Although determining the optimal



segmentation scale is difficult, the Estimation of Scale Parameter (ESP) tool, which builds on the idea of local variance of object heterogeneity, facilitates a suitable method for multi-segmentation that avoids objectivity and repetition [29]. In this study, after performing visual interpretation through trial and error, four levels (150, 100, 50, and 30) of segmentation scales were selected. The chosen scales were also tested by running ESP tools, and the results indicated that the images were segmented appropriately by using the bottom-up region-merging strategy. As shown in Figure 4, a larger segmentation scale detects larger but fewer objects, whereas a smaller segmentation scale detects smaller objects but in greater numbers. This does not necessarily mean that smaller segmentation detects more landslides. A visual comparison indicated that the objects were over-segmented at a scale of 30. Thus, the subsequent detections were based on the three scales of 50, 100, and 150. A total of 445 feature polygons were generated (Figure 5); however, after the post-processing stage (omitting outliers), 366 features were prepared for subsequent analysis. The attributes of the aforementioned objects, such as spectral, textural, and spatial information, were then exported for GA optimization.



**Figure 4.** Details of QuickBird image segmentation including four scales: (a) scale = 150; (b) scale = 100; (c) scale = 50; (d) scale = 30.

### 3.2. GAs-Based Optimization of Feature Selection

GAs are a class of stochastic search and optimization techniques based on natural selection and evolutionary principles [30]. This algorithm has been proven to be robust and effective in searching large spaces for a wide range of applications [31,32]. To minimize data redundancy, optimizing the features that are closely related to landslide occurrence was crucial because most of the segmented object features were not relevant to this study. To reduce the data dimensionality, in the second phase of the proposed approach, we applied the GA method to solve the optimization problems because of their robustness [33]. To determine the properties associated with the landslides, 366 feature attributes from the previous step (including the spectrum, shape, texture, hierarchy, and neighborhood of each object) were selected.

In using the GA for feature optimization, the feature attributes were coded as chromosomes, a type of binary string. Figure 6 presents an example of the procedure. For instance, the value of a code is set to 0 or 1. In this case, a bit value of 1 means that the corresponding instance is selected, and a value of 0 means that the corresponding instance is not selected. The populations were initially randomized before the search process was resumed, and then searched to determine the encoded chromosomes to maximize the optimal fitness function, which was computed for each of the randomly originated chromosomes. Because designing the optimal fitness function plays a major role in improving the search space efficiently and effectively, an improper fitness function can easily be trapped in a local optimum and can decrease in search effectiveness [31]. It facilitates assigning the optimal fitness value for each chromosome. The fitness function  $f(x)$  can be expressed as follows:



$$f(x) = \frac{\sum_{i=1}^n \sum_{j=1(\omega_i=\omega_j)}^n \delta(x^i, x^j)}{\sum_{i=1}^n \sum_{j=1(\omega_i \neq \omega_j)}^n \delta(x^i, x^j)} \quad (2)$$

$$\delta(x^i, x^j) = \sqrt{\sum_{k=1}^n \omega_k (x_k^i - x_k^j)^2} \quad (3)$$

where  $x^i$  is an  $n$ -dimensional feature vector of image object  $i$ ,  $x^i = (x_1^i, x_2^i, \dots, x_n^i)$  and  $\delta(x^i, x^j)$  are the Euclidean distance between vectors  $x^i$  and  $x^j$ , which is  $k$ -th feature value of the  $i$ ,  $\omega_k$  is the weight of the  $k$ -th feature, and  $n$  is the number of objects in feature optimization. The GA can compute the optimal fitness value for each individual, and under this condition, only the optimal individuals can survive. Hence, an optimized generation process can reproduce generations through mutation or crossover. Eventually, to passage a discrimination related to the fitness, the optimal individuals were decoded for use and corresponded to feature selection as inputs for landside detection and classification in the CBR process. The GA optimization process was conducted using the Gene Hunter software package.

### 3.3. CBR for Landslide Detection

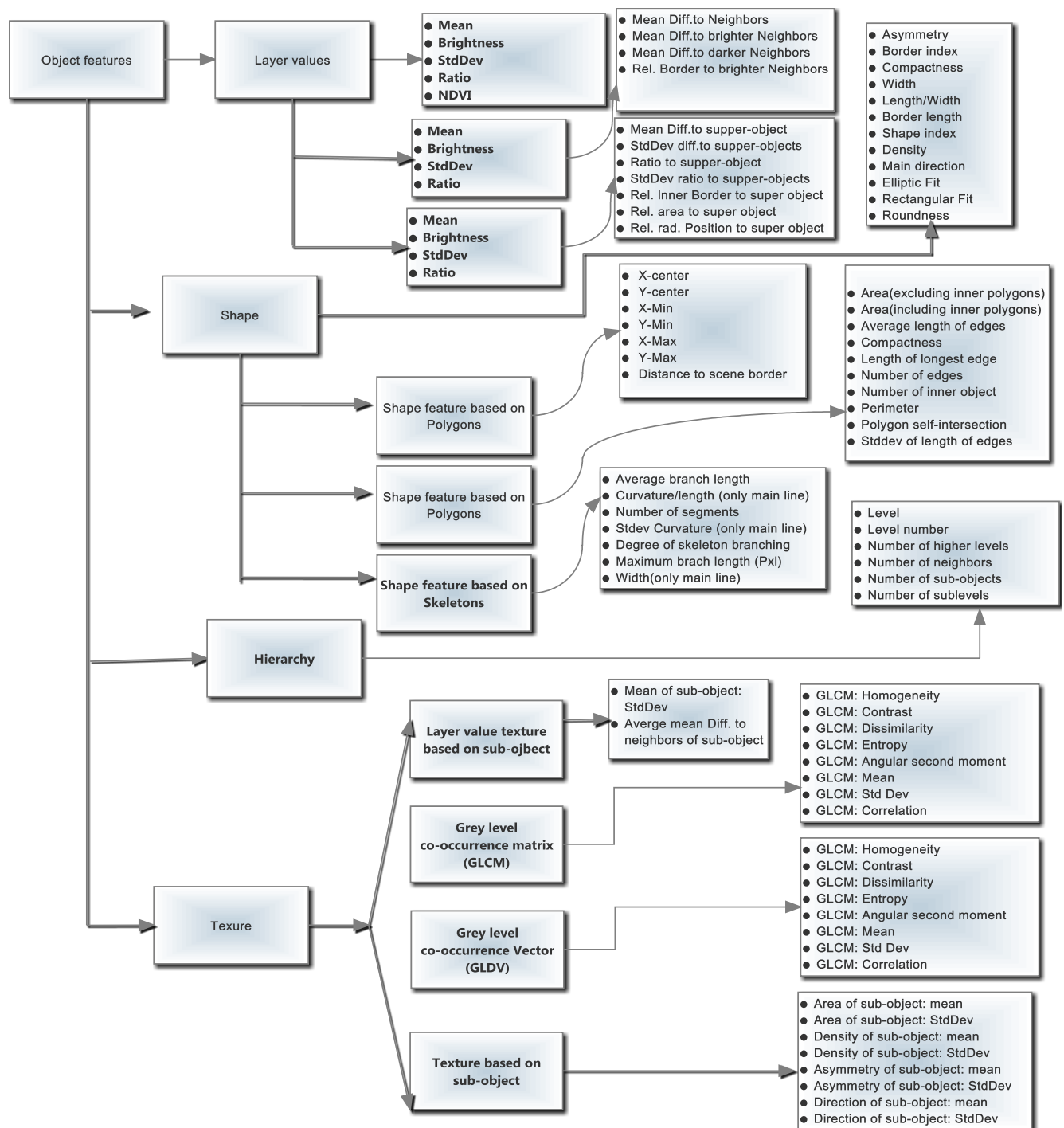
CBR is a problem-solving method that imitates the reasoning of human intelligence and involves applying past experiences to determine a solution. Hence, CBR has often been used in fields such as marketing, engineering, and economics to solve complex problems and is an effective technique because of its flexibility in representing a particular case [23,34,35]. Moreover, CBR has a higher number of appealing features than that of EKS approaches, and can overcome the disadvantages of EKS approaches while retaining advantages such as artificial intelligence, simplification of tedious tasks, and high automatic competence.

In the CBR approach, a case is the basic unit that records a problem condition and contextualizes knowledge representative of experience obtained during a problem-solving event. A case library is the core of a CBR system when applied for landslide detection. The case library of this study was established using aerial photographs and GPS data collected during the field survey. CBR offered an approach to representing the retrieved, reused, revised, and retained (4R) cases. The rapid development of CBR is attributable to its competence and maintenance of a rule-based EKS. Each case must be represented by a problem description in addition to the solution and outcome.

An advantage of CBR over conventional classifications such as MLC is that it can allow the use of both numeric and nonnumeric data and does not require a normal distribution form [14]. In this study, the data of each object were provided as the attributes of a case, thereby overcoming restrictions reported in previous studies, thus enabling the smooth interpretation of numeric data. Figure 7 shows the details of applying CBR for detecting landslides by using remote sensing data. Each case comprises two parts: the delineation of the problem and the solution of the problem (classified landslide types). A case can be expressed as follows:

$$X = (X_1, X_2, \dots, X_n; T_k) \quad (4)$$

where  $X_n$  is the  $n$ th feature related to the spectral, shape, and textural attributes of the case, and  $T_k$  represents the landslide type of the case.



**Figure 5.** Object attributes exported in the OOIA analysis.

CBR is used to detect landslides by assessing the interrelated similarity correlation degree (SCD). Several techniques have been used (e.g., Manhattan distance, grey relational analysis, k-nearest neighbor) to calculate the SCD between the input case and output case in the CBR system. Because the distance scale relational data reflects only the position of the curve rather than the trend of the data sequence changes, the actual distance to the data sequence may not be similar, as found using k-nearest neighbor or Manhattan distance techniques. However, gray relational analysis (GRA) improves the measurement of distance similarity so that it can fully express the similarity of data sequences, as shown

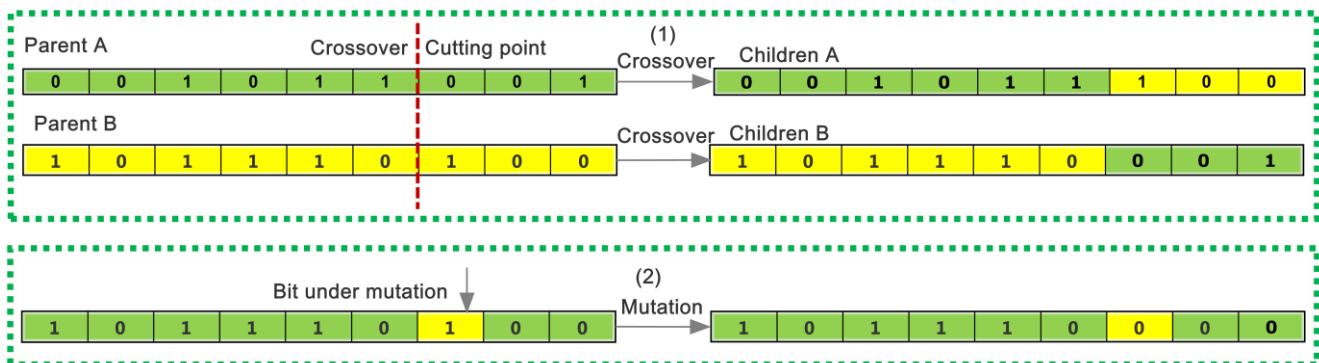
by Goldberg and other studies [29,30]. Hence, this research used the GRA method to calculate similarity because of its advantage of global comparisons. The SCD of GRA is calculated using the following equation:

$$SCD_i(n) = \frac{\min_i \min_n |X_0(n) - X_i(n)| + \zeta \max_i \max_n |X_0(n) - X_i(n)|}{|\min_i \min_n |X_0(n) - X_i(n)| + \zeta \max_i \max_n |X_0(n) - X_i(n)|} \quad (5)$$

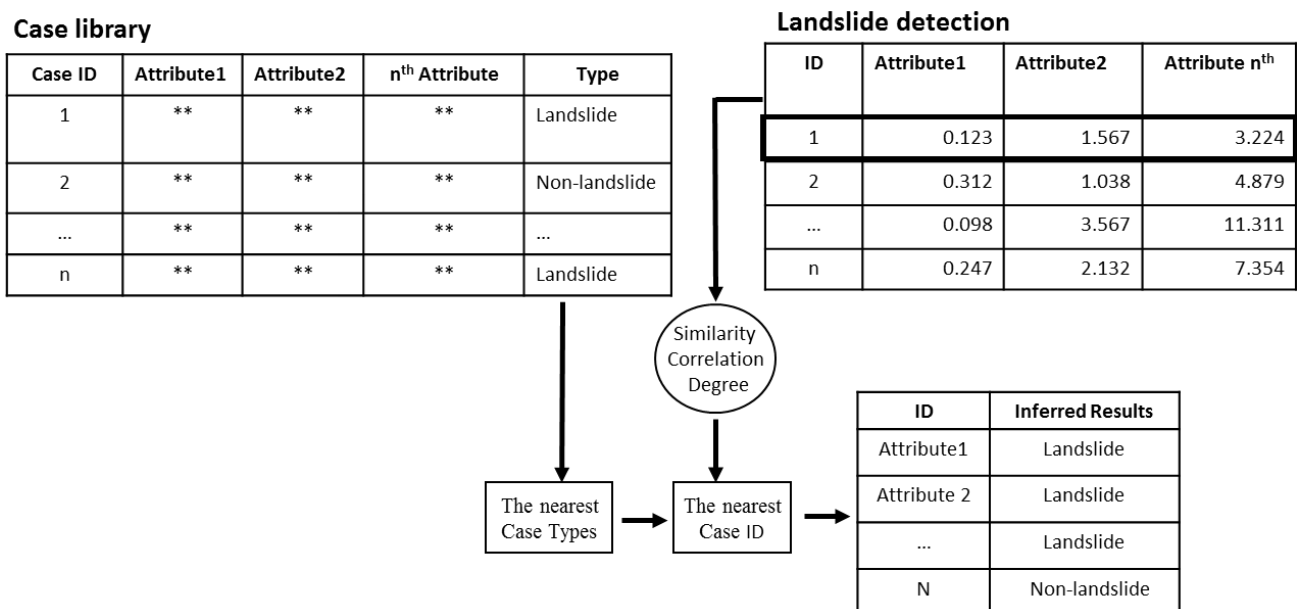
where  $SCD_i(n)$  is the related coefficient at point  $n$ , and value ranges  $[0, 1]$ ,  $X_0(n)$  and  $X_i(n)$  represent the value of the  $n$ th feature of input case  $x$  and existing case  $i$ , respectively. The term  $\zeta$  is the identification coefficient. The SCD assessment includes multiple values to prevent dispersion in a given system [36]; thus, the related grade (RG) between sequences can be defined by dividing the related coefficient by its average value:

$$RG_i = \frac{1}{N} \sum_{k=1}^N SCD_i(n) \quad (6)$$

where  $N$  is the number of features. The value of RG ranges from 0–1.



**Figure 6.** Illustration of GA: (1) an example of the single point crossover for a binary GA, randomly setting the parents and obtaining the offspring; (2) the mutation for a binary GA, in which the bits are randomly chosen and the allele's values are altered.



**Figure 7.** Illustration of landslide detection using the CBR method in this study.

The optimal features were normalized to eradicate the dominating effect of high values ranging mainly from 0–1 by using the following equation:

$$Y = \frac{X - X_{Min}}{X_{Max} - X_{Min}} \quad (7)$$

where  $Y$  is the normalized feature value ranging from 0–1, and  $X_{Max}$  and  $X_{Min}$  represent the maximum and minimum value, respectively.

Figure 8 shows examples of landslide cases in the case library. Multiscale detection must be used because each case has a unique size. Moreover, to elude repetitive detection, the segmentation results are portrayed from a larger scale to a smaller scale. Additionally, the detected landslides were extracted before performing subsequent small-scale detections.

Each segmented object was consecutively identified by CBR at different scales; the degree of similarity between each input and the existing case library was separately calculated so that the highest similarity degree could be determined to identify the landslide type. The CBR inference was programmed on the Matlab 2010b platform, thus increasing the likelihood of fast automatic detection [37].



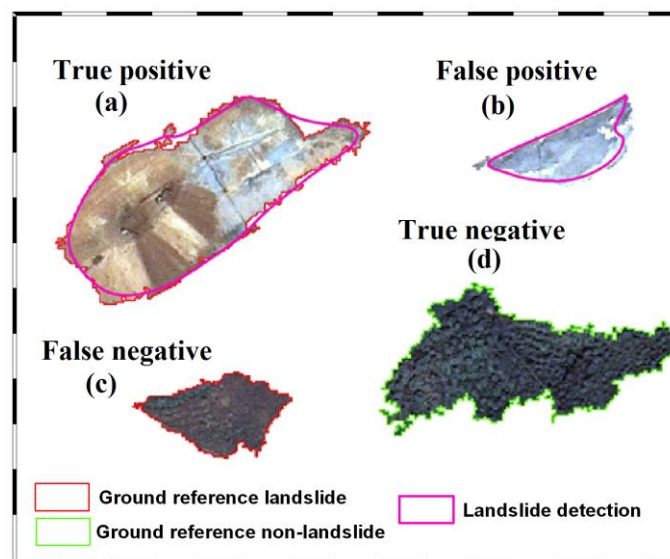
**Figure 8.** Examples showing the preparation of reference landslides in the case library, in the form of Quickbird images: (a–d) are young landslides (bright area with no vegetation cover); (e,f) are old landslides (less bright areas covered with concrete structures and vegetation).

### 3.4. Accuracy Estimation and Validation

Various methods and indices have been established for evaluating the accuracy of remote sensing products. Accuracy assessments typically rely on a confusion matrix and the definition of a sampling



unit defined in the response design [38,39]. In previous studies, indices such as the Kappa index, and overall, user, and producer accuracies were extensively used to estimate the results of map's quality [39,40]. However, recent studies have indicated a decline in the use of the Kappa index because of its flawed methodology, which involves the practical application of remote sensing [41]. Radoux and Bogaert [38] recommended the use of sample polygons to assess thematic accuracy when spatial objects are identified on a map as polygons. Mondini *et al.* [42] used receiver operating characteristic (ROC) plots (plotting the true positive value against the false positive value) for validating the results of map classification. These studies indicated that a higher ROC value matches the optimal fit results. Yang *et al.* [40] successfully applied overall accuracy to assess the accuracy of identification in the 2008 Wenchuan earthquake area in China. However, thus far, no universal optimal fit method exists for evaluating map accuracies [38].



**Figure 9.** A schmatic diagram showing the computation of true positives, false positives, false negative, and true negatives for verifying the accuracy of the model. (a) true positive: actual landslides that were correctly classified as landslides; (b) false positive: nonlandslides that were incorrectly classified as landslides; (c) false negative: landslides that were incorrectly classified as nonlandslides; and (d) true neagtive: nonlandslides that were correctly classified as nonlandslides.

In this study, the ROC index was employed for validating the model. The ROC is a popular index and has been extensively used in the field of engineering and signal detection. Swets (1988) indicated that the ROC is a highly useful indicator for evaluating the quality of deterministic and probabilistic detection and forecast systems [43]. Although it is widely accepted in binary calculations, it is seldom used in the field of remote sensing. Typically, several classes exist in calculations; however, in ROC plots, all classes are grouped into landslide and non-landslide classes. The area of the ROC was calculated using Matlab software. The true positive rate (TPR), false positive rate (FPR), false negative rate (FNR) and true negative rate (TNR) were computed by comparison with the ground reference data. Figure 9 shows the schematic diagram of the TPR, FPR, FNR, and TNR employed in this study. The area under the ROC curve (AUC) was then used to characterize the quality of a forecast system by describing the system's

ability to correctly predict the occurrence or nonoccurrence of a predefined event. The equation is expressed as follows [44].

$$Y, \text{Sensitivity} = TNR / (TNR + FPR) \quad (8)$$

$$\text{Specifity} = \frac{\sum \text{True negative}}{\sum \text{Condition negative}} = TNR / (TPR + FNR) \quad (9)$$

$$X = 1 - \text{Specifity} = 1 - \frac{\sum \text{True negative}}{\sum \text{Condition negative}} = 1 - TNR / (TPR + FNR) \quad (10)$$

$$AUC = \int_0^1 ROC(X) dX \quad (11)$$

where the values of the AUC vary from 0–1 and a higher AUC value represents a superior classifier.

## 4. Application and Resultant Analysis

### 4.1. CBR-Based Multi-Scale Landslide Detection

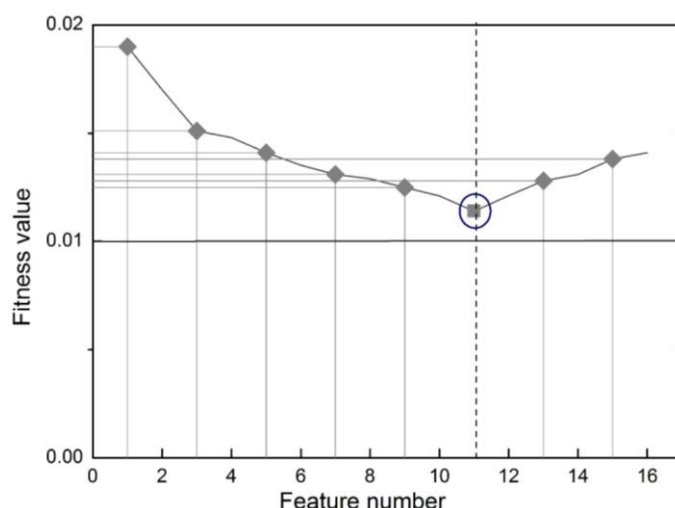
Based on previous studies and objects that were recognizable and interpretable from the satellite images [45], three classes of objects were categorized in this study: old landslides, young landslides, and non-landslides. Old landslides appear as large crescentic amphitheatres and are mostly covered by bush and grassland. Old landslides can be considered as relict or mature features that have been dormant for some time. These regions are often characterized by concrete or grass seeding protective measures for slope stabilization and therefore easily recognizable in the images. Examples of old landslides are shown in Figure 2c,e,f. Young landslides, however, appear as bright “scars” in the images and exhibit a clear unvegetated back scarp, and are therefore clearly visible in satellite images (see Figure 2a,b). Non-landslides include all classes of objects, excluding the aforementioned landslide classes.

In the GA optimization, the initial values for population, crossover, and mutation rate were set at 200, 0.6, and 0.05, respectively. Figure 10 illustrates the relationship between fitness value and feature number, indicating that the fitness number decreases for the first 11 features. After the 11th feature, the fitness value increases with the feature number (Figure 10). Based on these settings, the first 11 features were selected by the GA (Table 2). The selected features included spectral features such as the layer mean and ratio, textural features such as GLCM and GLDV, morphometric features such as elevation, and shape features such as the length-to-width ratio. CBR was then used on the basis of the features selected by the GA. Feature selection was simultaneously optimized by the GA by applying the multi-scale parameters; the feature selection was not related to the scale sizes.

The patterns of the shape of the curve for the optimized features and their values for each landslide type (including old and young) were highly similar and consistent; the curves for the non-landslide objects exhibited no obvious pattern and differed considerably from those of the objects classified as landslides (Figure 11). This implies that the SCD of the GA can be used to compare the experimental cases with the cases in the library.

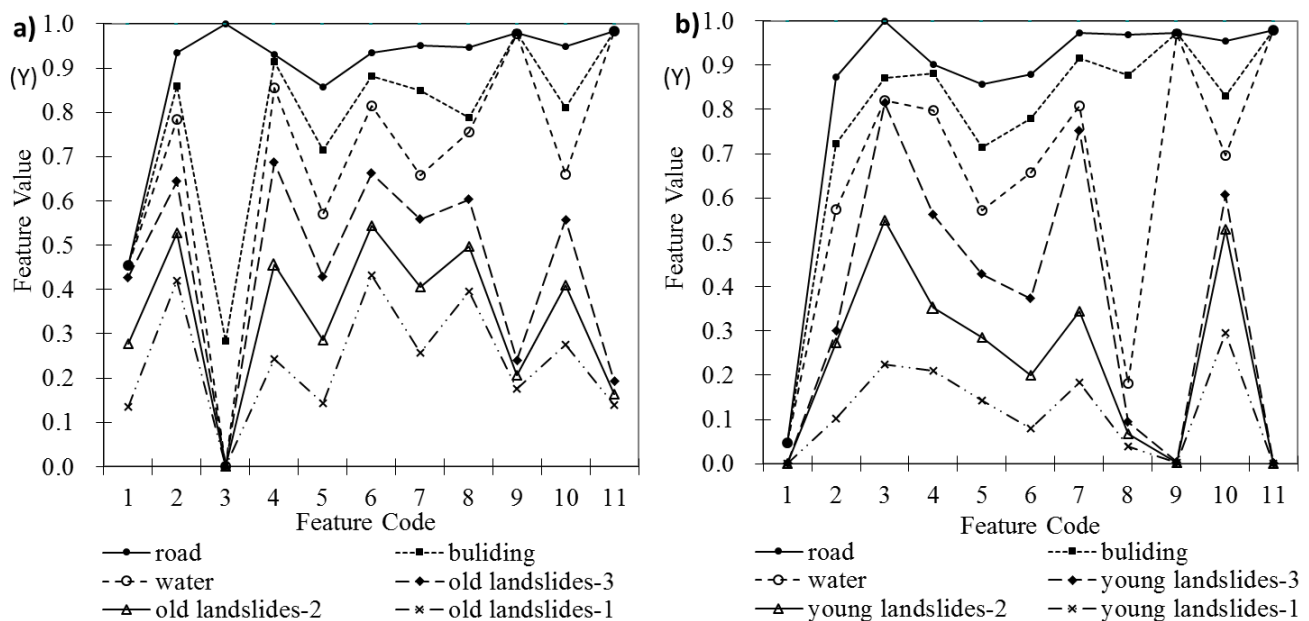
**Table 2.** Feature selection optimized by the GA.

Code	Feature Factors	Depiction
1	Mean diff to neighbors	For each neighboring object, the layer mean difference is computed and weighted with regard to the length of the border between the objects.
2	Ratio to scene	Ratio to scene of Layer L is the Layer L mean value of an image object divided by the Layer L mean value of the whole scene.
3	Length/Width	In the feature “Length/width (only main line),” the length of an object is divided by its width.
4	GLCM Std Dev(all dir)	The grey level co-occurrence matrix the layer values of all n pixels forming an image object. Feature value range: [0; depending on bit depth of data].
5	GLCM Homogeneity	If the image is locally homogenous, the value is high if GLCM is concentrated along the diagonal.
6	GLCM Dissimilarity	Texture measurement of the amount of local variation in the image objects by the grey level co-occurrence matrix (GLCM). It increases linearly and is high if the object has a high contrast.
7	GLDV Entropy (all dir)	The grey level difference vector the values are high if all elements have similar values.
8	GLCM Ang2nd moment	High if some elements are large and the remaining elements are small.
9	NDVI	Vegetation index, $NDVI = (NIR - R) / (NIR + R)$ , value range: $[-1, 1]$ .
10	Elevation	Elevation affects the distribution of vegetation and landslides typically occur at comparatively high elevation.
11	Slope	Slope = Raise/Run, $[0, 90^\circ]$ , affects the stability of slope failure.

**Figure 10.** Relationship between feature number and fitness value in the GA process.

After comparing the results, the segmentation scales of 150, 100, and 50 were determined to be suitable for identifying and delineating the landslides for the study area (Figure 12). The 800 cases in the library were then prepared from the three scales of segmentation (150, 100 and 50). These cases were separated equally but randomly into two parts: one for training cases, and the other for accuracy assessment. Each scale was treated as independent. Although, the landslides with larger sizes were detected by the large segmentation scale (150), and relatively small landslides were detected by the small segmentation scale (50), it is difficult to obtain a conclusion regarding the optimal segmentation fit

because misclassification in the form of over-segmentation in the smaller scale and omission of smaller landslides in the larger scale is always problematic. We therefore recommend a site-specific segmentation scale for an optimal fit result. The total prone area by detected landslides was calculated to be approximately 1.7 km<sup>2</sup>. The identified landslides varied in size between 59 m<sup>2</sup> and 32,700 m<sup>2</sup>. The results also indicated that implementing a multi-scale detection strategy and adopting essential measures to avoid repetition as mentioned can result in effective landslide detection. The hybrid approach can facilitate detecting landslides in large landslide-prone areas and simultaneously reduce the visual-interpretation bias.

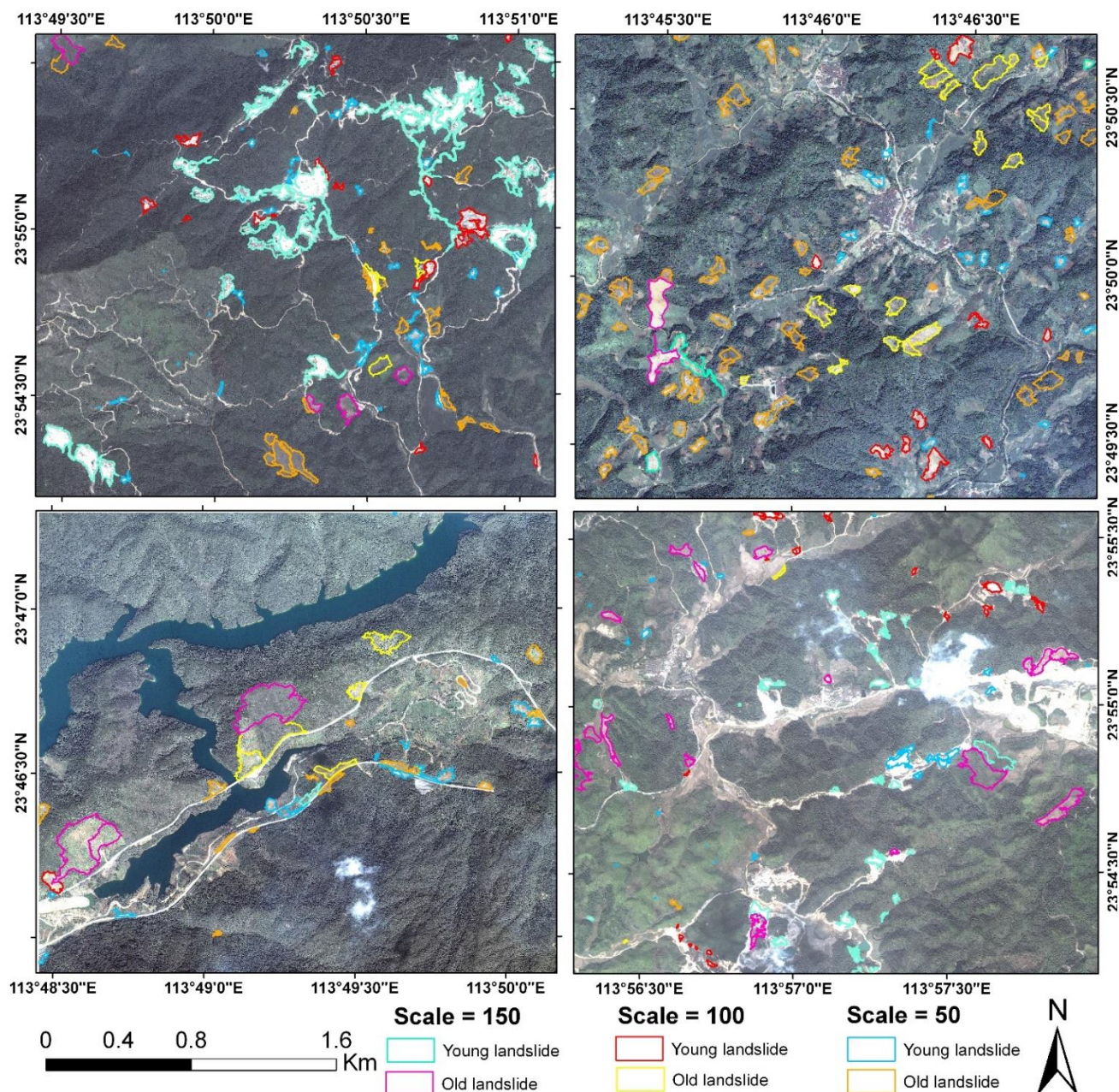


**Figure 11.** Plots of old (a) and young (b) landslide curves. The curves have similar periodic trend changes.

#### 4.2. Accuracy Assessment

Accuracy assessment of the results was verified using field investigation, GPS control points, and a QuickBird image acquired in 2007 (Figure 13). The accuracy of the proposed methodology was evaluated with the second set of samples separated from the 800 cases (see Section 4.1). ROC curves for the two types of landslide mentioned in Section 4.1 are displayed in Figure 14. AUC values for the young landslides (0.9) outperformed the old landslides (0.82). Table 3 shows the accuracy assessment results of the landslide detection, indicating that the overall accuracy was 0.87. The young-landslide detection of user and producer accuracy was 0.86 and 0.89, respectively, whereas the accuracies for old landslide classes were 0.82 and 0.85, respectively. The young landslides had obvious characteristics, compared with neighboring objects, such as less vegetation cover. These results indicated that high-resolution remote sensing data can be effectively used to detect landslides, particularly in the urbanized region in the PRD, which has pioneered China's economic development and urbanization process.



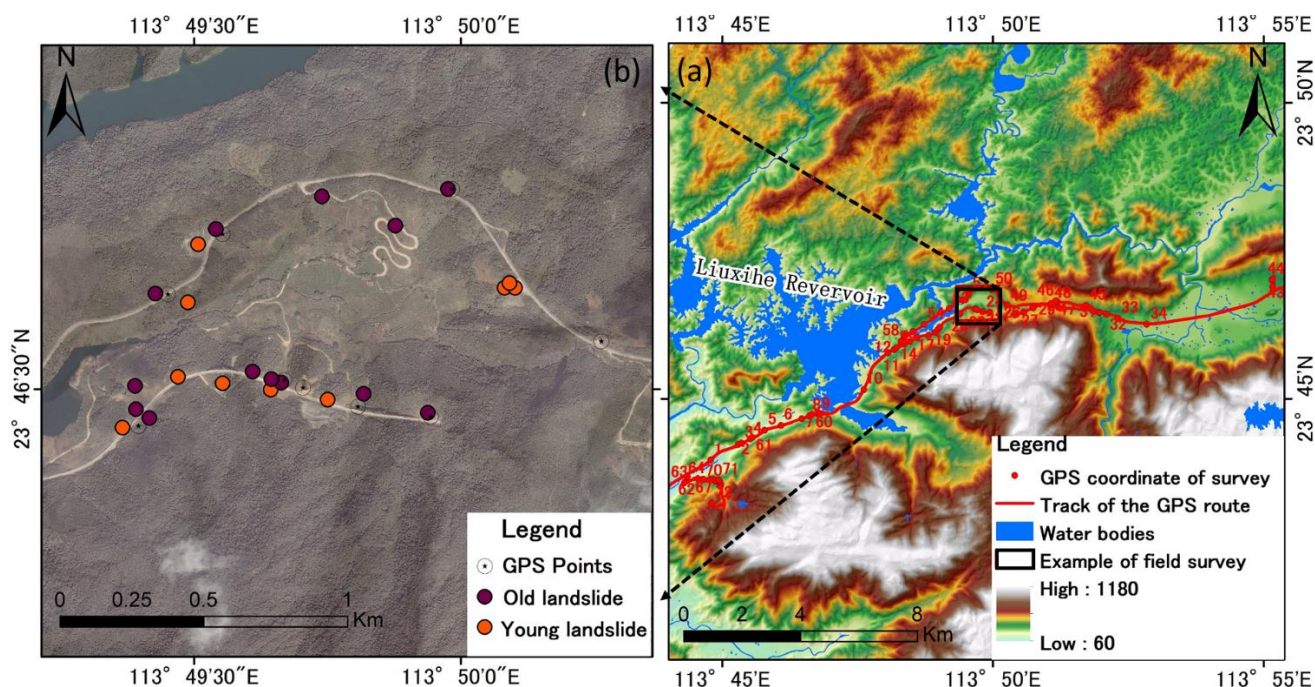


**Figure 12.** Results of landslide detection at different scales (150, 100, and 50).

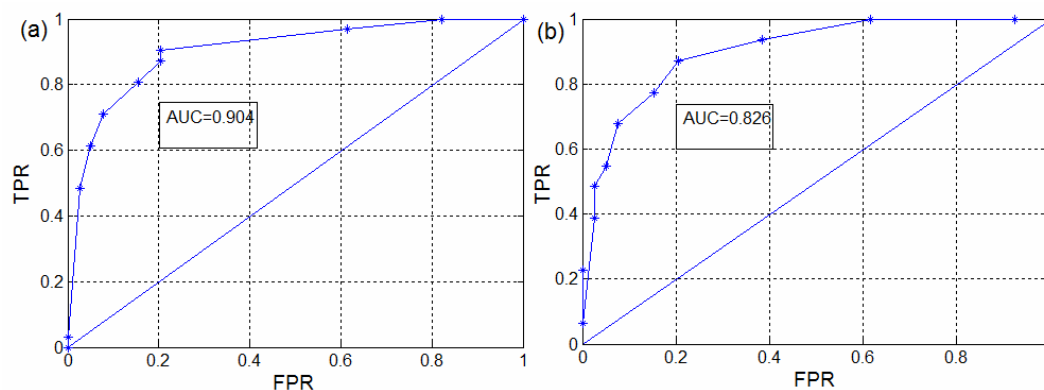
To elucidate the efficiency of the proposed method, the results were compared with the OOIA classification approach. The nearest neighbor membership function value at standard deviation ( $\sigma$ ) was set at 0.2, and feature space optimization was used to refine the classification. The same 11 features were optimized by the GA. Subsequently, the same cases were used as training data for the landslide detection. Figure 15a,b shows that the AUCs of the OOIA method for old and young landslides were 0.7 and 0.74, respectively. Table 4 shows the results of the OOIA method, indicating an overall accuracy of 0.75. Additionally, other comparisons were evaluated by applying a standard traditional per-pixel classification, viz., and the supervised maximum likelihood classification (MLC) for classifying the same images with the same training and testing samples, using Erdas software. This method was derived from the Bayes theorem, which expresses a posteriori distribution based on spectral data. The AUC of the MLC method for old and young landslides were 0.64 and 0.67, respectively (Figure 15). Table 5 shows that the



maximum likelihood method of overall accuracy was 0.68. Both results suggested that the accuracy of the OOIA and maximum likelihood classification method are considerably lower than that of the hybrid CBR model proposed in this study. According to the same accuracy estimation approach, the results of stand-alone OOIA and MLC were highly unsatisfactory with a considerably lower classification accuracy than that obtained using the OOIA-GA-CBR method. The poor classification of the supervised method may be due to the obvious spatial variations of environmental settings (e.g., roughness and soil types). However, the discrete cases in the CBR method must be suitable for representing these complexities and must facilitate obtaining more favorable classification performance.



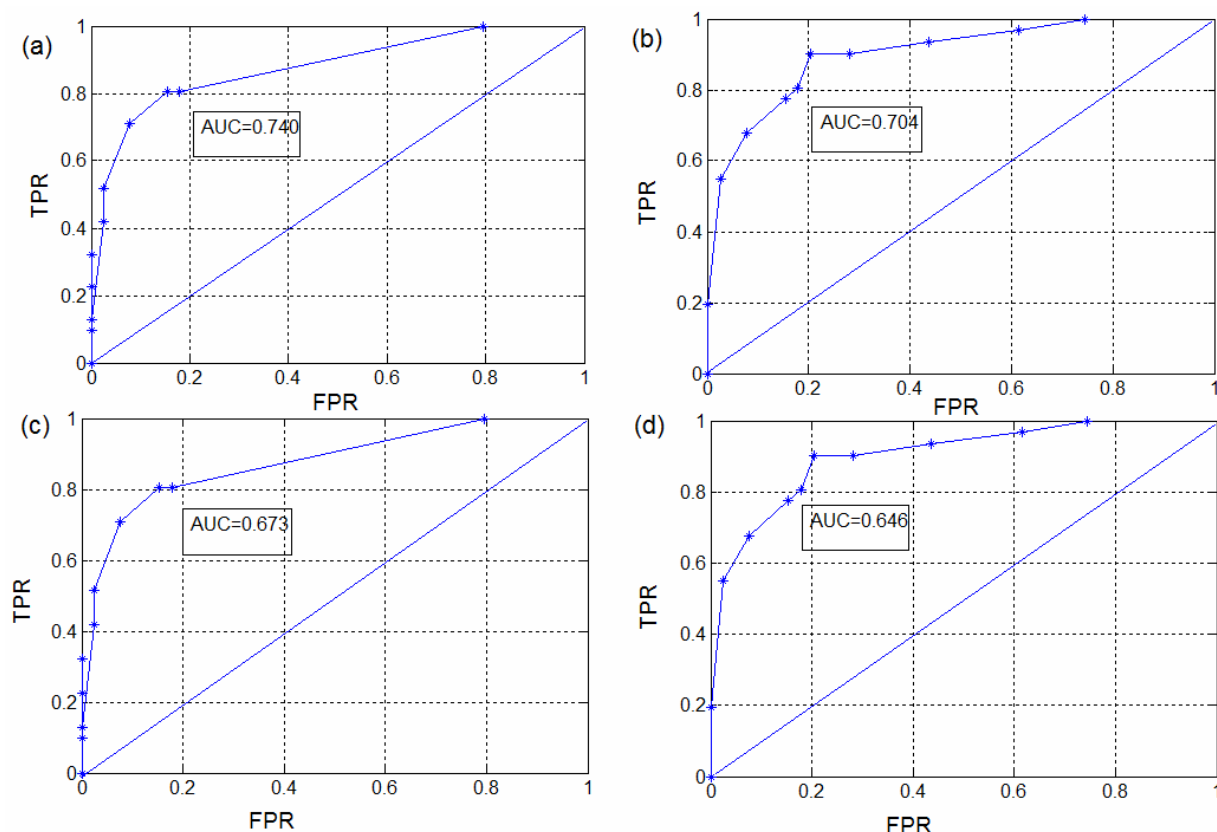
**Figure 13.** Data for validation: (a) track of GPS route for field survey; (b) examples of landslide detection points overlying the QuickBird image, using CBR.



**Figure 14.** Prediction rate of ROC curve for each landslide class using CBR: (a) young landslides; (b) old landslides.

**Table 3.** Comparison of accuracy in the CBR method.

Accuracy	User's Accuracy	Producer's Accuracy	Omission Error	Commission Error	Overall Accuracy
Old landslide	0.82	0.85	0.15	0.18	0.87
Young landslide	0.86	0.89	0.11	0.14	

**Figure 15.** Prediction rate of the ROC curve for each landslide class: (a) young landslides, using OOIA; (b) old landslides, using OOIA; (c) young landslides, using MLC; and (d) old landslides, using MLC.**Table 4.** Comparison of accuracy in the OOIA method.

Accuracy	User's Accuracy	Producer's Accuracy	Omission Error	Commission Error	Overall Accuracy
Old landslide	0.82	0.85	0.15	0.18	0.75
Young landslide	0.86	0.89	0.11	0.14	

**Table 5.** Comparison of accuracy in the supervised maximum likelihood method.

Accuracy	User's Accuracy	Producer's Accuracy	Omission Error	Commission Error	Overall Accuracy
Old landslide	0.65	0.62	0.38	0.35	0.68
Young landslide	0.67	0.63	0.37	0.33	

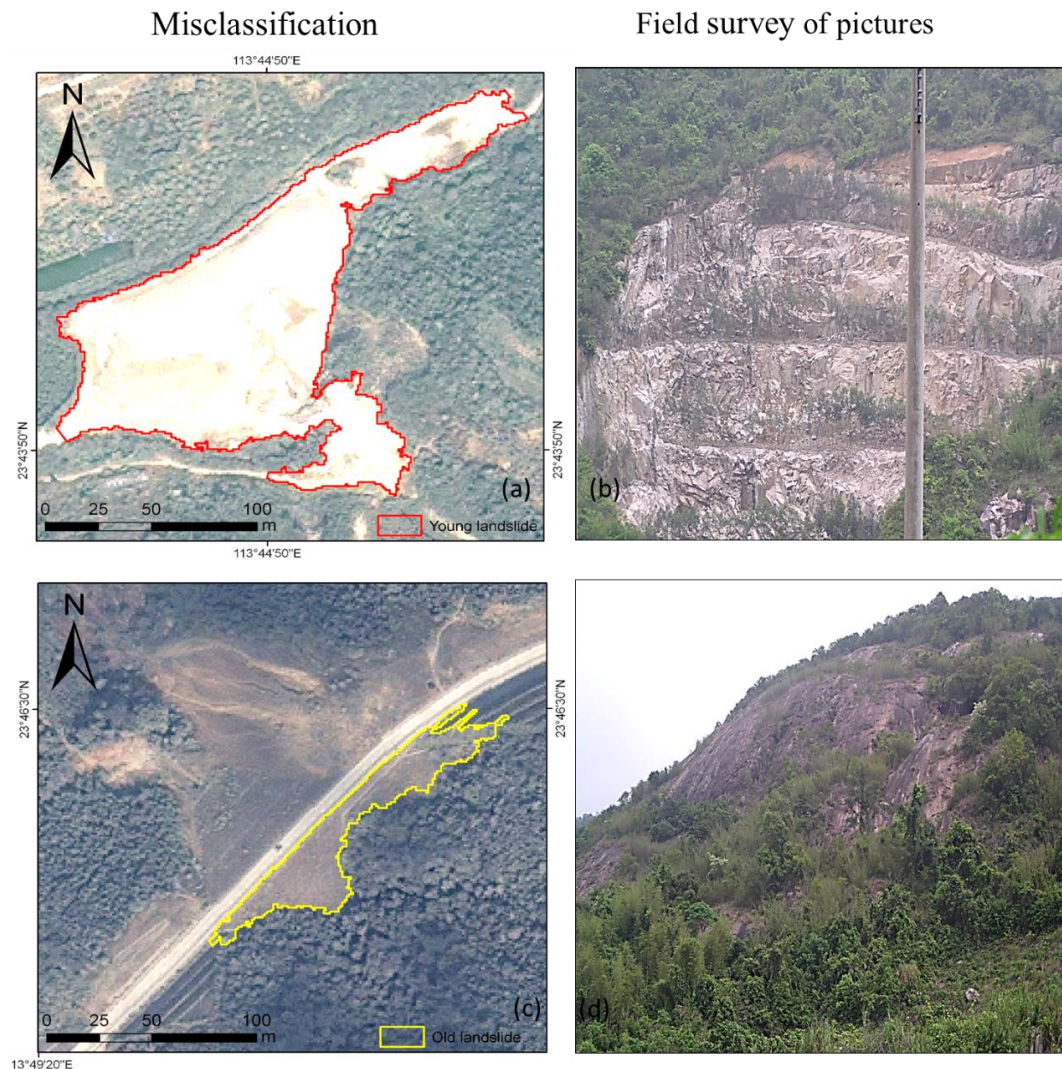
## 5. Discussion

In mountainous terrains, field investigation for landslide inventory mapping is challenging and tedious [20]. Although direct visual interpretation of remote sensing data is a reliable method, this task is laborious and time-consuming. New perspectives on detecting landslides by applying remote sensing

techniques have recently been facilitated by the availability of high-resolution images and improvements in computer technology. In this context, automated detection techniques are suitable for establishing an efficient spatial inventory landslide database that is useful for qualitative/quantitative hazard assessment. Various methods for automating landslide mapping have been proposed; however, no method has successfully demonstrated ideal results. Nichol and Wong [46] attempted using SPOT and IKONOS images to map landslides through pixel-based methods. Martha *et al.* [20] recently applied OOIA to map landslides semiautomatically through contextual analysis. They empirically used expert knowledge to select features. Chang *et al.* [18] identified shallow landslides by using high-resolution remote sensing observation data; however, they did not use a DEM and NDVI. For OOIA, applying expert knowledge to characterize landslides plays a key role in semiautomatic landslide detection. In OOIA, selecting features related to a landslide occurrence is highly subjective because of the high dependence on expert knowledge. Hence, an appropriate feature selection of landslide types is required for detection. In this study, the information of objects created through the multi-segmentation of fine resolution QuickBird (0.61 m) multispectral data and 5-m resolution elevation data were extracted, and major landslides were characterized as old and young landslides.

Multi-segmentation was performed to extract the features. Based on the vast feature information, the GA was used to optimize the selection of features related to landslides based on the data redundancy and correlation coefficient. GAs are a powerful technique for optimization parameters and can also avoid the subjectivity of quantified expert knowledge. The GA-based optimization selected 11 features from 366 features. The optimized features included spectral features such as layer mean, morphological characteristics such as slope and elevation, and textural information such as GLCM. The GLCM and orthogonal relationship between the flow and main directions of objects can be used to classify agricultural areas and roads, respectively. This process shortens the computation time, reduces the dependence on subjective expert-knowledge, and improves the accuracy of landslide identification, particularly compared with the complex rule-set-based classification of images. For example, Blaschke *et al.* [47] built a complicated rule set to detect and delineate landslides by using OOIA, which relies on expert experience and is time-consuming. A classification accuracy of 0.87 in the number of landslides was achieved in the experimental study area (Table 3). The young landslides were detected with a higher accuracy than the old landslides. Some vegetated areas were misclassified as old landslides because of the dense vegetation cover. In addition, the main misclassification of young landslides involved confusing them with bare areas (e.g., rocky outcrops, bare land, and roads) by the classifier because of the sparse vegetation; their tones were similar with a bright appearance similar to landslide-prone areas in the satellite image, as shown in Figure 16. To reduce false positive classification in the future, we must integrate topographical GLCM that can substantially reduce the misclassification of objects [47]. Furthermore, the established case library can be reused for time-independent landslide detection. The case library is crucial for CBR. Only typical cases must be revised and updated for a time suitable detection. This process can reduce the laborious aspects of visual interpretation of data and increase productivity.





**Figure 16.** Examples of misclassification: (a) a bare quarry was misclassified as a young landslide, and (b) the related field photo; (c) a rock outcrop with sparse forest was misclassified as an old landslide, and (d) the related field photo. The tones of misclassification were similar, with a bright appearance of the landslide area in the satellite image.

## 6. Conclusions

Landslide detection can facilitate establishing a landslide inventory that can potentially provide a clearer understanding of landslide patterns through spatial and temporal dimensions. Classifying high-resolution remote sensing data provides valuable landslide information for disaster management and urban planning. However, conventional visual interpretation is tedious and time-consuming. This paper proposes a new approach for automating landslide detection by integrating an object oriented approach and a GA. The results indicate that high-resolution images can enable quick identification of landslides in a fast-growing area in the PRD, South China. This study demonstrated the advantage of the proposed integration approach by using a GA to optimize feature selection and combining OOIA and CBR for detecting landslides.

The proposed model incorporates each method's advantages and avoids problems such as knowledge-based selection bottlenecks in creating EKSs. In addition, the experimental results indicated

that the hybrid model demonstrated a higher accuracy than traditional methods (supervised classification). Traditional methods generally hypothesize that a certain terrain object must maintain a stable spectral signature in the entire study area. The variations, however, can be complex under real world conditions. Moreover, the established case library in the hybrid model can be reusable for time-independent landslide detection. To reuse the case library, the typical landslides and non-landslides must be revised and updated.

This paper proposes an approach for detecting and characterizing landslide features to construct a landslide inventory for the PRD. The established inventory is the basis for forecasting the spatial and temporal distribution of future landslides. Predictive susceptibility mapping with satisfactory consistency can be transformed into hazard disasters by applying the numerous independent landslides. This valuable knowledge is suitable for disaster management and identification of landslide-prone areas, rebuilding after the occurrence of landslide disasters, and preventing future unnecessary economic losses in urban development.

## Acknowledgments

This study was supported by the project of Guangzhou geological-hazards survey (Project No. 1212010511106-05) and Science & Technology Plan Foundation of Guangdong Province of China (No. 2012A020200018, 2013B020501006). Dou thanks very much Junping Qian for her in-depth instructions. We are truly grateful to Xuan Song, Hagar Hecht, and Jian Yang, whose comments were extremely helpful in improving this paper.

## Author Contributions

Jie Dou and Shuisen Chen collected the data set. Jie Dou, Kuan-Tsung Chang, and Yunus Ali designed the experiment, performed the analysis and wrote the manuscript. Jin-King Liu, Huan Xia and Zhongfan Zhu gave constructive comments for the improvement and revision of the manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Turner, A.K.; Schuster, R. *Landslides: Investigation and Mitigation*; Transportation Research Board: Washington, DC, USA, 1996.
2. Alexander, D.E. A brief survey of GIS in mass-movement studies, with reflections on theory and methods. *Geomorphology* **2008**, *94*, 261–267.
3. Garc ía-Rodr íguez, M.J.; Malpica, J.A.; Benito, B.; D áz, M. Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. *Geomorphology* **2008**, *95*, 172–191.
4. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.-T. Landslide inventory maps: New tools for an old problem. *Earth-Sci. Rev.* **2012**, *112*, 42–66.

5. Wang, L.-J.; Sawada, K.; Moriguchi, S. Landslide-susceptibility analysis using light detection and ranging-derived digital elevation models and logistic regression models: A case study in Mizunami City, Japan. *J. Appl. Remote Sens.* **2013**, *7*, 073561.
6. Assilzadeh, H.; Levy, J.K.; Wang, X. Landslide catastrophes and disaster risk reduction: A GIS framework for landslide prevention and management. *Remote Sens.* **2010**, *2*, 2259–2273.
7. Dou, J.; Oguchi, T.; Hayakawa, Y.S.Y.; Uchiyama, S.; Saito, H.; Paudel, U. GIS-based landslide susceptibility mapping using a certainty factor model and its validation in the Chuetsu Area, Central Japan. In *Landslide Science for a Safer Geoenvironment*; Springer International Publishing: Berlin, Germany, 2014; pp. 419–424.
8. Wu, Q.; Zhang, H.; Chen, F.; Dou, J. A web-based spatial decision support system for spatial planning and governance in the Guangdong Province. In Proceedings of the 2008 Geoinformatics and Joint Conference on GIS and Built Environment, Guangzhou, China, 28–29 June 2008; Volume 7144, p. 12.
9. Zhou, S.; Fang, L. Support vector machine modeling of earthquake-induced landslides susceptibility in central part of Sichuan province, China. *Geoenviron. Disasters* **2015**, *2*, 1–12.
10. Lu, P.; Stumpf, A.; Kerle, N.; Casagli, N. Object-oriented change detection for landslide rapid mapping. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 701–705.
11. McKean, J.; Roering, J. Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology* **2004**, *57*, 331–351.
12. Dou, J.; Qian, J.; Zhang, H.; Chen, S.; Zheng, X.; Zhu, J.; Xie, Z.; Zou, Y. Landslides detection: A case study in Conghua city of Pearl River delta. In Proceedings of the Second International Conference on Earth Observation for Global Changes, Chengdu, China, 25–29 May 2009; p. 11.
13. Myint, S.W.; Gober, P.; Brazel, A.; Grossman-Clarke, S.; Weng, Q. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sens. Environ.* **2011**, *115*, 1145–1161.
14. Li, X.; Yeh, A.G. Multitemporal SAR images for monitoring cultivation systems using case-based reasoning. *Remote Sens. Environ.* **2004**, *90*, 524–534.
15. Dou, J.; Zheng, X.; Qian, J.; Liu, R.; Wu, Q. Intelligence based automatic detection and classification of ground collapses using object-based image analysis method: A case study in Paitan of Pearl River delta. *Proc. SPIE.* **2008**, doi:10.1117/12.813168.
16. Dou, J.; Zheng, X.; Qian, J.J.; Liu, R.; Wu, Q.; Chen, S.; Zhen, X. Object-based and case-based reasoning method for ground collapses detection. *J. Image Graph.* **2010**, *15*, 900–910.
17. Anders, N.S.; Seijmonsbergen, A.C.; Bouten, W. Segmentation optimization and stratified object-based analysis for semi-automated geomorphological mapping. *Remote Sens. Environ.* **2011**, *115*, 2976–2985.
18. Chang, K.-T.; Liu, J.; Wang, C. An object-oriented analysis for characterizing the rainfall-induced shallow landslide. *J. Mar. Sci. Technol.* **2012**, *20*, 647–656.
19. Qian, J.; Liu, R.; Chen, S.; Dou, J. Man-made ground collapse detection using high resolution Aerial image and object-based classification: Example of Pearl River Delta. In Proceedings of the Second International ISCRAM Workshop on Information Systems for Crisis Response and Management, Harbin, China, 26–27 August 2007; pp. 1–6.

20. Martha, T.R.; Kerle, N.; Jetten, V.; van Westen, C.J.; Kumar, K.V. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology* **2010**, *116*, 24–36.
21. Yang, M. A genetic algorithm (GA) based automated classifier for remote sensing imagery. *Can. J. Remote Sens.* **2007**, *33*, 203–213.
22. Vancoillie, F.; Verbeke, L.; Dewulf, R. Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sens. Environ.* **2007**, *110*, 476–487.
23. Minor, M.; Bergmann, R.; Görg, S. Case-based adaptation of workflows. *Inf. Syst.* **2014**, *40*, 142–152.
24. Ahn, H.; Kim, K.; Han, I. Hybrid genetic algorithms and case-based reasoning systems for customer classification. *Expert Syst.* **2006**, *23*, 127–144.
25. Du, Y.; Wu, D.; Liang, F.; Li, C. GIScience & Remote Sensing integration of case-based reasoning and object-based image classification to classify SPOT images: A case study of aquaculture land use mapping in coastal areas of Guangdong province, China. *GISci. Remote Sens.* **2013**, *50*, 574–589.
26. Varnes, D.J. Slope movement types and processes. *Natl. Acad. Sci. Transp. Res. Board Spec. Rep.* **1978**, *176*, 12–33.
27. Baatz, M.; Benz, U.; Dehghani, S.; Heynen, M. *ECognition User Guide\_4.0*; Definiens Imagine GmbH: Munich, Germany, 2004.
28. Benz, U.C.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 239–258.
29. Drăguț, L.; Tiede, D.; Levick, S.R. ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 859–871.
30. Goldberg, D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Boston, MA, USA, 1989; p. 432.
31. Tang, A.M.; Quek, C.; Ng, G.S. GA-TSKfnn: Parameters tuning of fuzzy neural network using genetic algorithms. *Expert Syst. Appl.* **2005**, *29*, 769–781.
32. Iovine, G.; D'Ambrosio, D.; Di Gregorio, S. Applying genetic algorithms for calibrating a hexagonal cellular automata model for the simulation of debris flows characterised by strong inertial effects. *Geomorphology* **2005**, *66*, 287–303.
33. Kudo, M.; Sklansky, J. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognit.* **2000**, *33*, 25–41.
34. Aamodt, A.; Plaza, E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Commun.* **1994**, *7*, 39–59.
35. Jonassen, D.H.; Hernandez-Serrano, J. Case-based reasoning and instructional design: Using stories to support problem solving. *Educ. Technol. Res. Dev.* **2002**, *50*, 65–77.
36. Deng, J.-L. Control problems of grey systems. *Syst. Control Lett.* **1982**, *1*, 288–294.
37. Matlab and Statistics Toolbox, Computer software. The MathWorks Inc.: Natick, MA, USA, 2010.
38. Radoux, J.; Bogaert, P. Accounting for the area of polygon sampling units for the prediction of primary accuracy assessment indices. *Remote Sens. Environ.* **2014**, *142*, 9–19.
39. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.



40. Yang, W.; Wang, M.; Shi, P. Using MODIS NDVI time series to identify geographic patterns of landslides in vegetated regions. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 707–710.
41. Pontius, R.G.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429.
42. Mondini, A.C.C.; Guzzetti, F.; Reichenbach, P.; Rossi, M.; Cardinali, M.; Ardizzone, F. Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sens. Environ.* **2011**, *115*, 1743–1757.
43. Swets, J.A. Measuring the accuracy of diagnostic systems. *Science* **1988**, *240*, 1285–1293.
44. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874.
45. Van den Eeckhaut, M.; Marre, A.; Poesen, J. Comparison of two landslide susceptibility assessments in the Champagne-Ardenne region (France). *Geomorphology* **2010**, *115*, 141–155.
46. Nichol, J.; Wong, M.S. Satellite remote sensing for detailed landslide inventories using change detection and image fusion. *Int. J. Remote Sens.* **2005**, *26*, 1913–1926.
47. Blaschke, T.; Feizizadeh, B.; Daniel, H. Object-based image analysis and digital terrain analysis for locating landslides in the Urmia Lake Basin, Iran. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1–12.

© 2015 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 license (<http://creativecommons.org/licenses/by/4.0/>).