

# A Supervised Object-Based Detection of Landslides and Man-Made Slopes Using Airborne Laser Scanning Data

2

Biswajeet Pradhan and Ali Alsaleh

## 2.1 Introduction

In recent years, airborne-derived products from light detection and ranging (LiDAR) measurements, such as high-resolution digital elevation models (DEMs), slope, curvature, shaded relief, and maps of landslides obtained from beneath dense vegetation, are becoming increasingly important for producing a detailed landslide inventory map (Eeckhaut et al. 2007). LiDAR applications include the construction of DEMs and shaded relief maps, detection of historical landslides under forested area, creation of topographic contours, tracking of multitemporal digital terrain model (DTM) of landslides, hydrological modeling, landform and/or soil classification, and understanding fine-scale landslide patterns (McKean and Roering 2004). The rapidness of LiDAR technology in landslide mapping of terrains through quantitative or visual analysis provides several advantages.

Landslide inventory maps provide baseline information of landslide types, location, distribution, and boundaries in landslide-prone areas. In addition, landslide inventory provides information on displacement and slope measurements that affect a failure (Galli et al. 2008). Moreover, landslide inventories are significantly useful for various purposes, such as recording of landslide magnitude, implementing the initial stage for landslide susceptibility, and hazard and risk assessments.

Object-based image analysis (OBIA) is a well-known technique resulting from the recent advances in computer vision and machine intelligence, with the main purpose of

automatically extracting both man-made and natural objects from remote sensing images (Akçay and Aksoy 2008). OBIA, in which the information content of an object is used to classify a landscape, is a step toward replicating human interpretation process (Navulur 2006). In addition, OBIA can detect landslides accurately and meaningfully by integrating contextual information to image analysis (Martha 2011), which reduces the time and cost for producing a decent landslide inventory map, especially in large areas. Several techniques have been proposed for landslide mapping, such as field observation and aerial photointerpretation; however, these techniques have some limitations, such as lack of proper resolution for aerial photographs required for the mapping of small-scale landslides caused by morphologic feature obscuration by thick vegetation cover, time-consuming, and difficulties in field mapping (Gorum et al. 2011). Remote sensing data and methods have been proven efficient in landslide mapping because of their wide area coverage, relatively cheap cost, and remarkably high-resolution data for landslide mapping, in which even minor landslides can be mapped easily. Landslide inventory maps resulting from the application of OBIA techniques can be easily converted to GIS data, which is considered as an initial stage for a more advance analysis, such as susceptibility, and hazard and risk analysis.

Landslides can be triggered by various factors. These factors can be man-made (such as mineral mining, road cutting, and urbanization) or natural (e.g., extreme rainfall events and earthquakes (Zêzere et al. 1999)). However, in tropical areas, a rapid and accurate method for landslide mapping is required because of the rapid growth of vegetation that covers the land surface characteristics in those areas. Furthermore, several cut slopes are generally created to mitigate the risk of land failure for areas that have high probability to fail because of the frequent occurrence of landslides. These requirements create new challenges for landslide identification and mapping in these areas. Thus, new methods should be developed for automatic landslide detection to produce high-quality landslide inventory map.

---

B. Pradhan (✉) · A. Alsaleh  
Department of Civil Engineering, University Putra Malaysia,  
Serdang, Malaysia  
e-mail: biswajeet24@gmail.com

## 2.2 Study Area and Data

### 2.2.1 Location of Study Area

Taman Ringlet is the first town along the Cameron Highlands stretched from Tapah, Cameron Highlands, Peninsular Malaysia. This town is a small hill that is well known for tea and strawberry farming. The town is located at an altitude of 1140 m above the sea level. Geographically, Taman Ringlet is located at latitude  $04^{\circ} 24' 45''\text{N}$  and longitude  $101^{\circ} 23' 30''\text{E}$ .

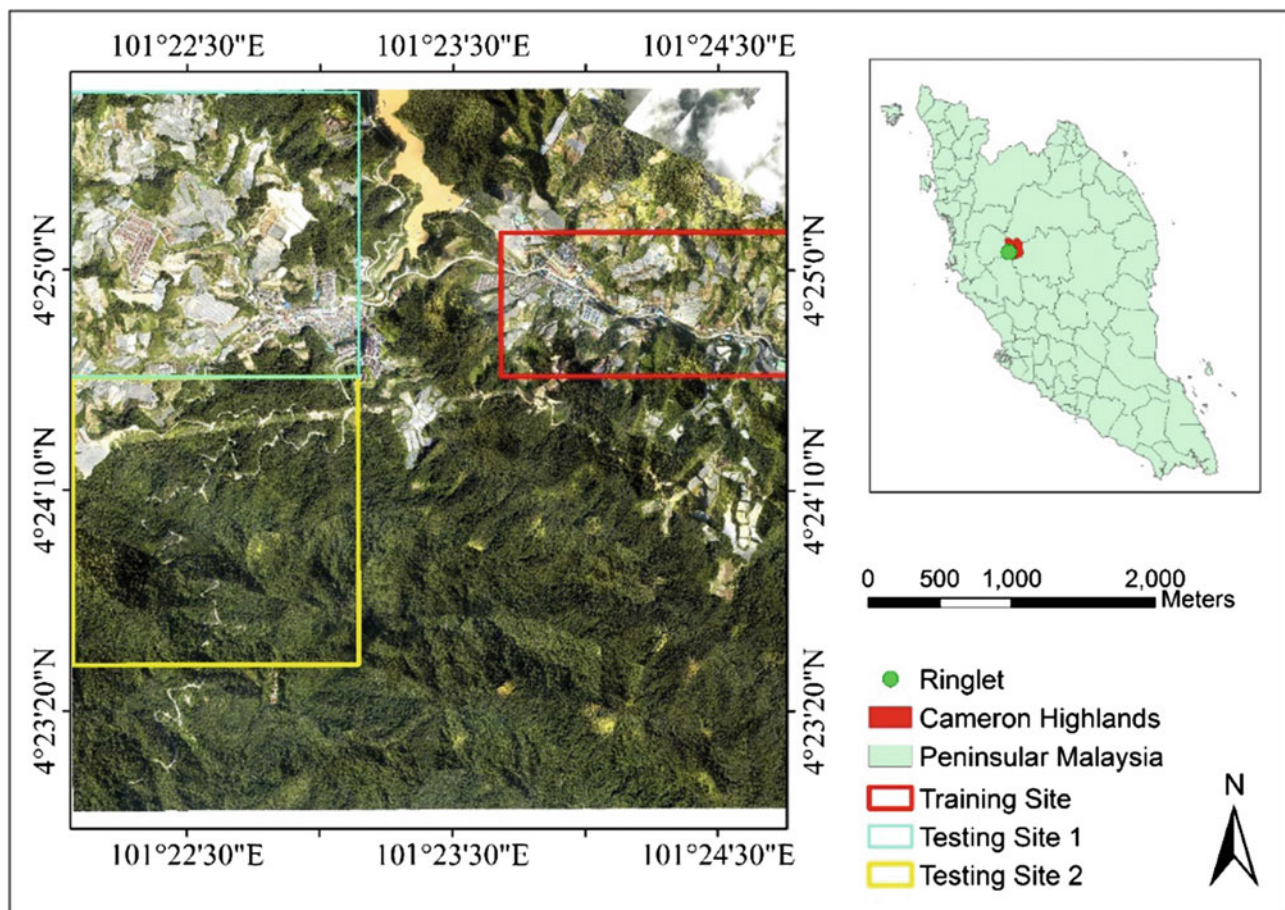
Three different subsets were selected for analysis, as shown in Fig. 2.1. One of the subsets was used to develop the methodology of landslide detection, whereas the other two were used to test the proposed method in other areas. The process of selecting subset locations was carefully implemented, with each subset having the same land cover classes but with different coverage percentage. Figure 2.1 shows that the training site and Testing Site 1 have more urban coverage compared to subset Testing Site 2, where thick vegetation covers almost the entire area.

### 2.2.2 LiDAR Data

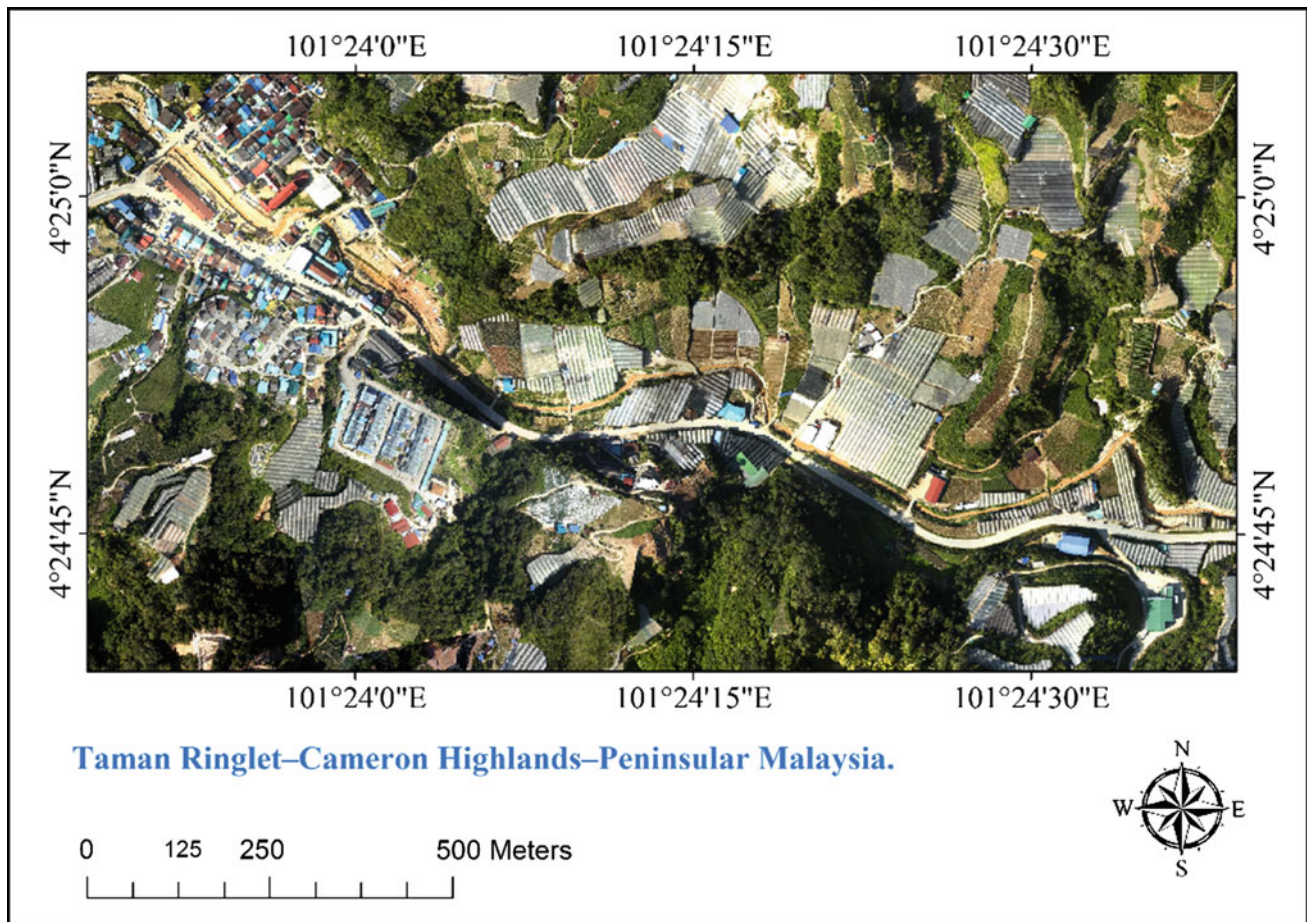
Study and data collection was implemented over Ringlet and nearby surrounding area, which covered a total area of  $25 \text{ km}^2$ . The LiDAR data were recorded for the entire  $25 \text{ km}^2$  with a flight height of 1510 m. Data were obtained on January 15, 2015. Data capturing performed well with eight points per square m and gave a pulse rate frequency of 25,000 Hz. Furthermore, the captured data were within the root-mean-square of 0.15 and 0.3 in the vertical and horizontal axes, respectively; thus, the accuracy of the captured data was reasonable. Along with LiDAR point clouds, orthophotos were also collected by the same system, as shown in Fig. 2.2.

### 2.2.3 Geological Characteristics of the Study Area

Cameron Highlands District is located in the eastern part of the main range, which is composed of granites (Bignell and



**Fig. 2.1** Geographic location of the study area



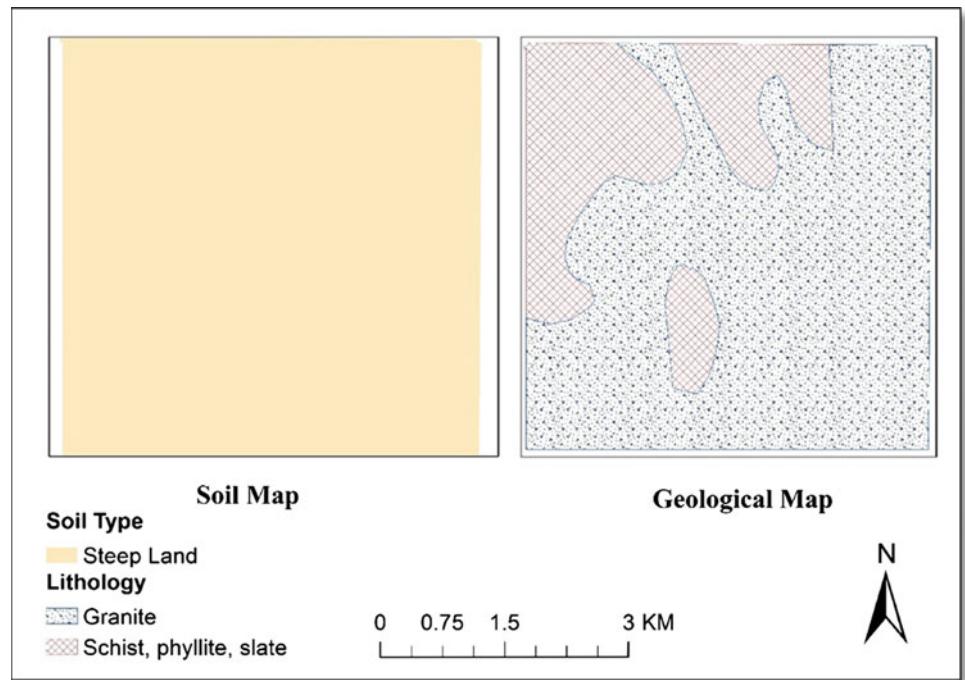
**Fig. 2.2** Aerial photograph of the training site

Snelling (1977). However, scattered outliers (roof pendants) of metasediments can be seen clearly in Fig. 2.3. Bignel and Snelling (1977) classified the granites of Cameron Highlands District as megacrystic biotite granite. Cobbing et al. (1992) mentioned that the associated microgranite and some of the granites may be mineralized and may contain muscovite. Furthermore, Chow et al. (2003) stated that metasediments consist of phyllite, schist, limestone, and slate. They also stated that minor intercalations of volcanic rocks and sandstones were found. Figure 2.3 shows the geological map of the study area and its surrounding areas. Post-Triassic–Mesozoic granites comprise most of the granite rocks, whereas a few are patches of metamorphic rocks that are mostly composed of Silurian–Ordovician schist, phyllite, limestone, and sandstone. As for the soil type, steep land soil covers the entire scene.

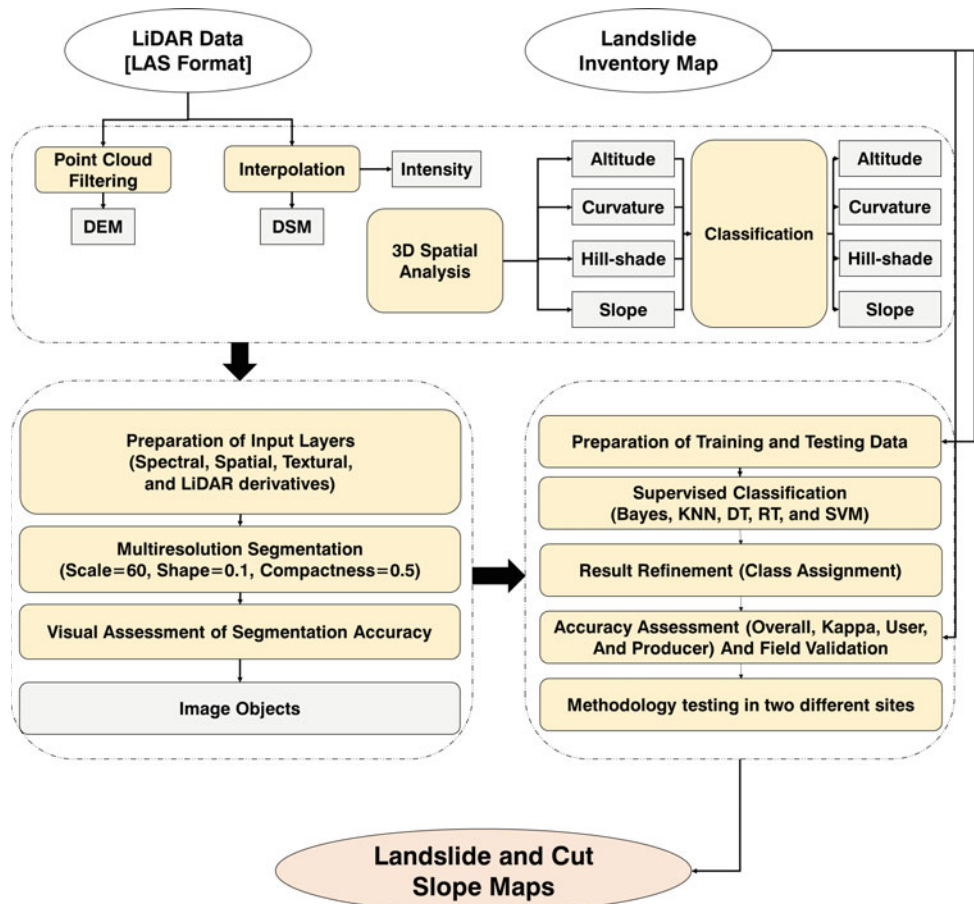
### 2.3 Methodology

Figure 2.4 shows the overall flowchart of the methodology implemented in this study. High-resolution LiDAR data with 1-m spatial resolution were used as a main data along with aerial photographs covering the Ringlet and its surrounding regions. For the ancillary data, a landslide inventory map showing the location of historical landslides was used. The overall methodology comprises three main phases: The first phase is the pre-processing and preparation of data; the second phase is image segmentation and object creation; and the final phase is image classification, and the detection of landslides and man-made slopes. The third step also includes result validation using a landslide inventory map created from field investigations based on site visits.

**Fig. 2.3** Geological characteristic maps of the entire study area



**Fig. 2.4** Overall methodology flowchart



### 2.3.1 Data Pre-processing and Preparation of Landslide Factors

Calibrated raw LiDAR data are typically processed into flight lines of 3D points and saved in LASer (LAS) format. The first step of LiDAR data pre-processing is a visual review of the flight lines to inspect the breaks or gaps between or within flight lines (Chen et al. 2004). Afterward, LiDAR data were validated with several control points. These data were obtained from the field to ensure accurate collection. Typically, these known features are in open areas free of vegetation, such as airport runways, roads, and targeted survey points. Some systematic errors can be corrected by additional bore sighting and calibration; however, if the digital data are inaccurate or corrupted, the mission may have to be repeated (Chen et al. 2004). Basically, the two former steps should be done before the acquisition team leaves the field. If errors are discovered at this stage, the system can be redeployed for another mission. After the LiDAR data were reviewed and a few reference points were validated, noise points were then filtered out; these noises could be of an extremely high or low value with unrealistic elevation values or with unexpected values in the project area (Fang and Huang 2004). After noise removal, layer extraction was done by attributing all the remaining points into layers using the multiple return system (Hodgson et al. 2005).

Subsequently, landslide conditioning factors (i.e., digital surface model (DSM), DTM, curvature, slope, hill shade, and altitude) were prepared. This process starts with the conversion of the LAS format to raster using ArcMap software by applying the Triangular Irregular Network (TIN)-based interpolation to create the DSM layer (Fig. 2.5a) with a spatial resolution of 1 m. Next, DEM layer (Fig. 2.5b) was generated by filtering out the point cloud into ground and non-ground points using the former interpolation technique. The 3D spatial analysis tool in ArcMap software was used to extract slope, curvature, hill shade, and altitude layers (Fig. 2.5c–f, respectively) from the DSM and DEM layers. Evidently, slope is the principal factor affecting landslide occurrences (Pradhan and Lee 2010). Slope is considered as an important factor for land stability because of its direct impact on landslide phenomenology (Martha et al. 2011). That is, a steeper slope means higher risk of landslide caused by gravity-induced high shear (Long 2008). The hill-shade map shows a good image and movement of the terrain, which supports landslide mapping (Olaya 2009). Curvature layer defines the convex/concave character of the surface. Curvature values are calculated as positive, negative, and zero values, which refers to concavity, convexity, and flattening ground surface, respectively (Pradhan and Lee 2010). Plan curvature is considered important, because it reliably

indicates convergence and divergence of slope surfaces in depletion (concave forms of the landslide crowns, tension cracks and depressions, and zones of local water accumulation) and accumulation zones (convex forms of the landslide foot and toe; (Ohlmacher 2007).

### 2.3.2 Image Segmentation

Before classifying a feature of interest, such as landslides, delineating image objects that separately or aggregately discriminate a specific feature (i.e., trees, buildings, and parcels of land) is important. This process is called image segmentation, which divides an image into objects or regions based on the homogeneity of pixel values (Martha 2011). The precision and quality of segmentation have a direct impact on the accuracy of the generated classification map (Laliberte et al. 2004).

This research analysis was conducted in eCognition software. Several types of algorithms can be found for the purpose of image segmentation, with multiresolution, quad tree, and chessboard being the most efficient ones (Definiens 2007). These algorithms provide an effective application for segmentation and perform good accuracy results.

Multiresolution algorithm, which belongs to the region-based algorithm category, was utilized in this study for segmentation (Möller et al. 2007). This algorithm performs various steps, which is initiated with one pixel and continues until all the criteria specified by the user are covered (Benz et al. 2004). Multiresolution segmentation algorithm uses three parameters: scale, shape, and compactness. Selecting the value of these three parameters should be carefully implemented to achieve meaningful classification results (Gibril et al. 2016). In this study, a trial-and-error approach was used to select the parameters, and evaluation was based on visual interpretation.

### 2.3.3 Classification

Image segmentation was also examined visually. The software calculated different parameters for each object, including the mean of slope, curvature, DEM, brightness, and density, as well as geometrical parameters, such as shape index, texture, length/width, area, and compactness. Each of these parameters was later used to classify an image object into several classes using supervised object-based classification approach.

#### 2.3.3.1 Classifier

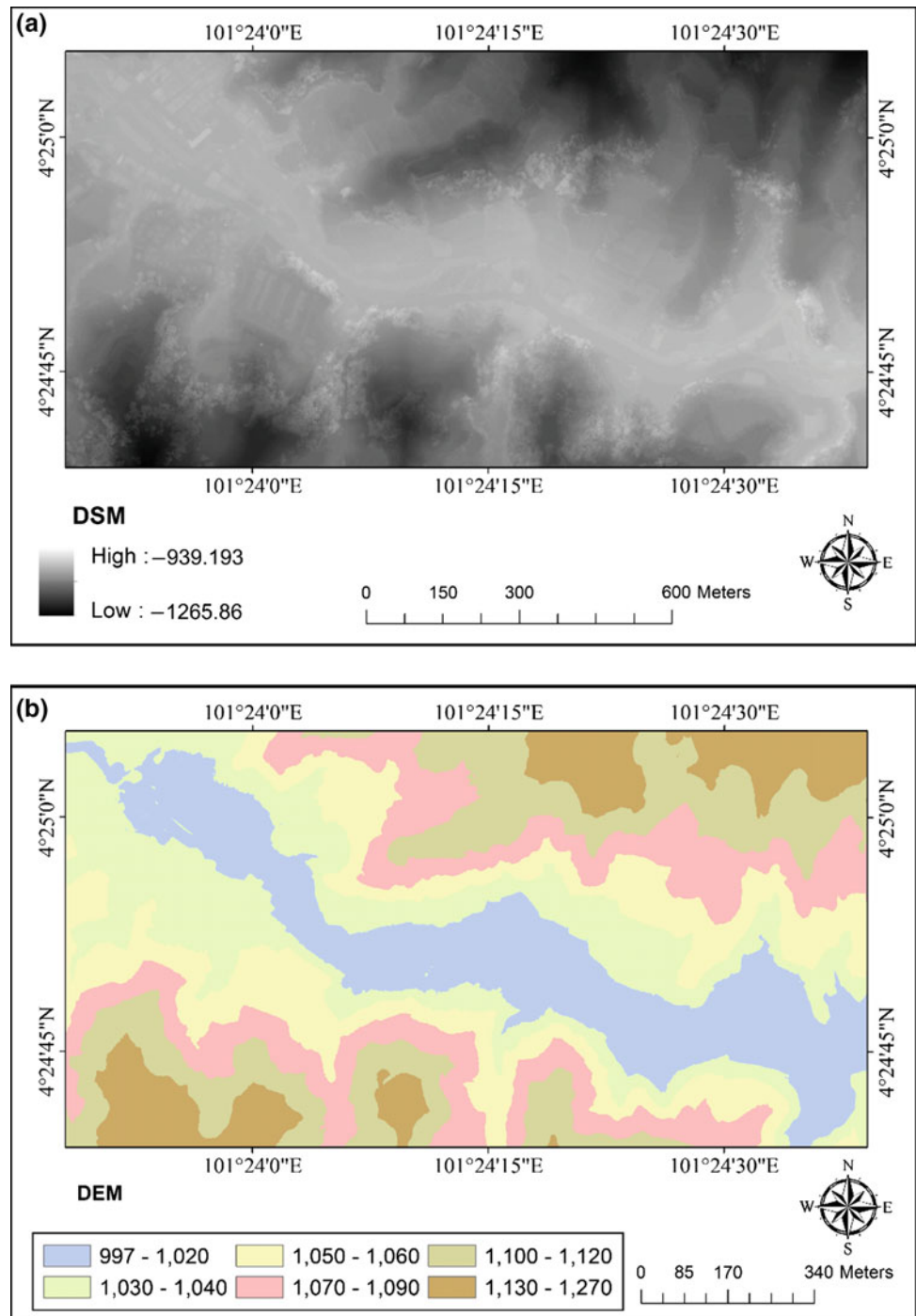
Classifier algorithm allows analysts to apply machine-learning functions in a two-step process. First, a

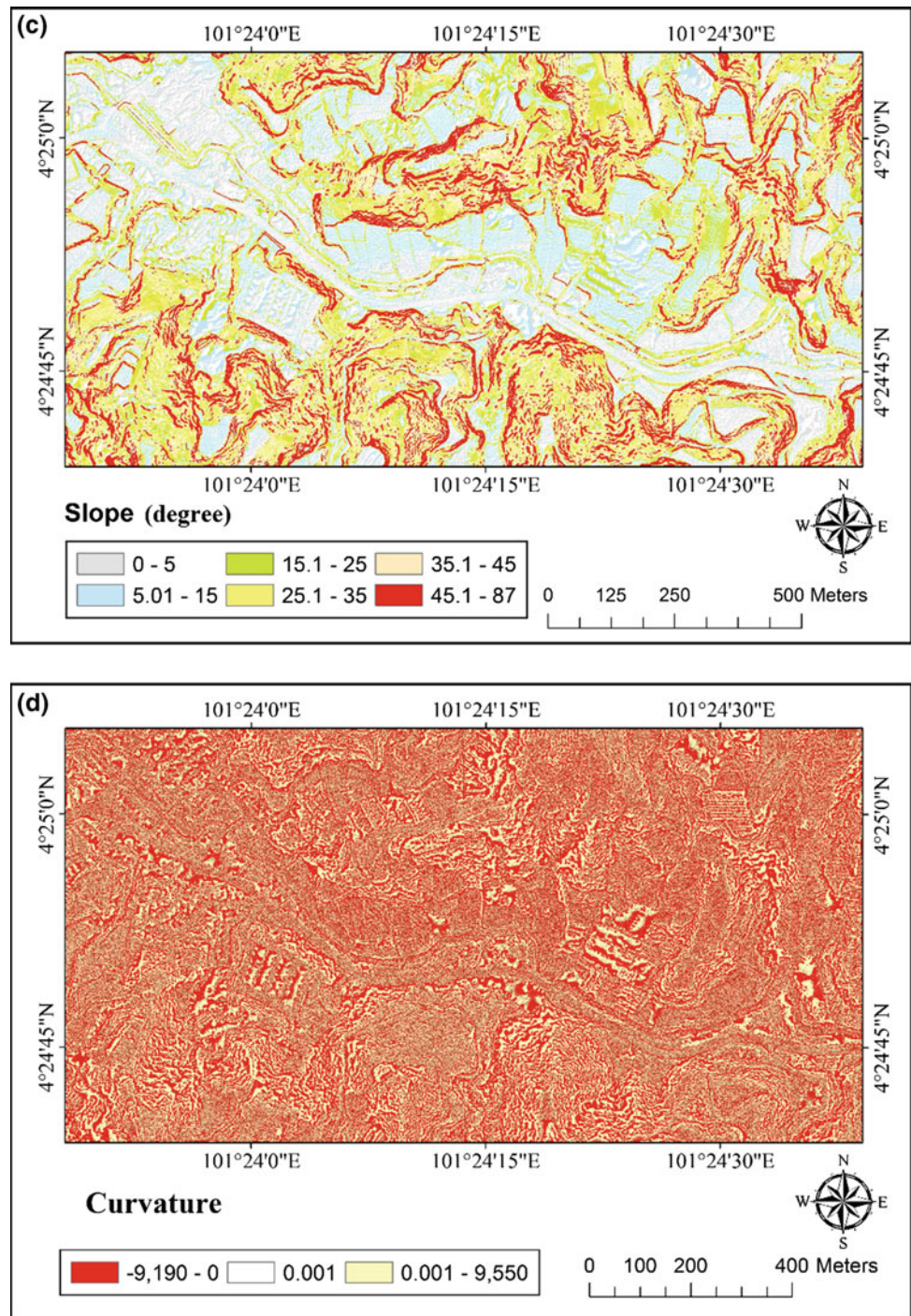
classifier is trained using the classified domain objects as training samples. The trained classifier was stored as a string variable in the configuration settings. Second, the trained classifier was applied to the domain, classifying the image objects according to the trained parameters. Classification accuracy ensures the proper selection of the sampling method (Chen et al. 2014). In this study, training samples were selected randomly and distributed fairly over the entire study area, in which 60% of the samples were

used for training purpose and 40% were used for testing the result.

Generally, five different algorithms [i.e., Bayes, *k*-nearest neighbor (*k*-NN), decision tree (DT), random forest (RF), and support vector machine (SVM)] can be applied to the classifier algorithm. Each of the aforementioned algorithms is best suited for a specific purpose. In this study, these algorithms were tested to identify the optimum algorithm for landslide and cut slope detection.

**Fig. 2.5** Landslide conditioning factors: **a** DSM; **b** DEM; **c** slope; **d** curvature; **e** altitude; and **f** hill shade



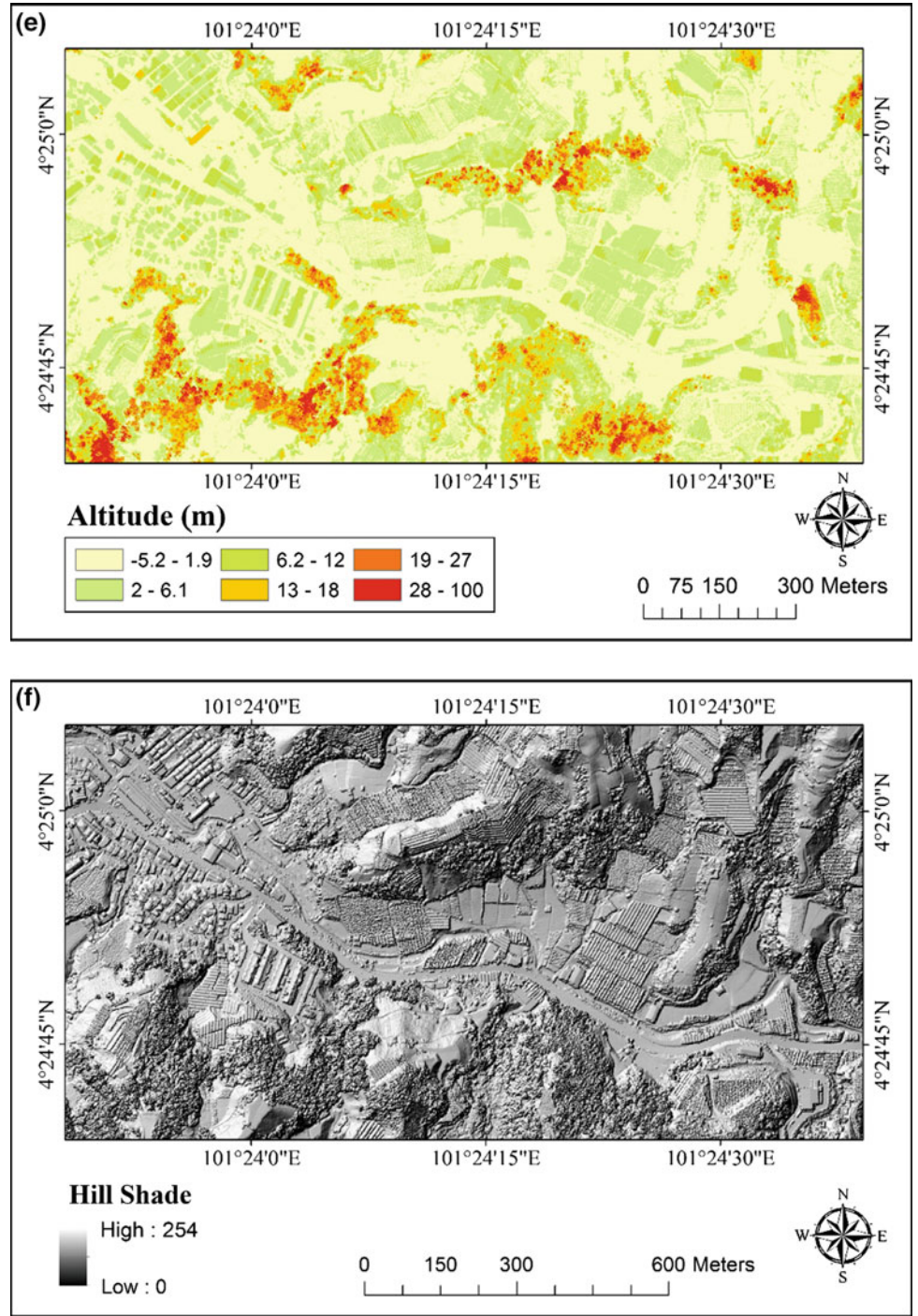
**Fig. 2.5** (continued)

### 2.3.3.2 Bayes

Conditional independence assumption is a machine-learning classification system derived from Bayes' theorem, which strongly supports the assumption of the independence between features (Soria et al. 2011). One advantage of this classifier is the simplicity of its construction, which does not need any complicated estimation schemes of iterative

parameters (Wu et al. 2008). In addition, Naive Bayes (NB) classifier is unaffected by noise or irrelevant attributes. Numerous successful experiments and studies of this classifier have been conducted in the literature (Xie et al. 2005).

Given an observation consisting of  $k$ -attributes  $x_i$ ,  $i = 1, 2, \dots, K$  ( $x_i$  is a landslide conditioning factor), and  $y_j$ ,  $j = \text{landslide}$ , man-made slopes are the output class. NB

**Fig. 2.5** (continued)

estimates the probability  $P(y_j/x_i)$  for all possible output class. The class can be predicted depending on the largest posterior probability using Eq. (2.1).

$$y_{NB} = \underset{y_i \in [\text{landslide, non-landslide}]}{\arg \max P(y_j)} \prod_{i=1}^n P(x_i/y_j) \quad (2.1)$$

The prior probability  $P(y_j)$  can be estimated using the proportion of the observations with output class  $y_j$  in the

training dataset. Conditional probability is calculated using Eq. (2.2).

$$p\left(\frac{x_i}{y_j}\right) = \frac{1}{\sqrt{2\pi}\delta} e^{-(x_i-\mu)^2/2\delta^2}, \quad (2.2)$$

where  $\mu$  is mean, and  $\delta$  is standard deviation of  $x_i$ .

The Bayes classifier has a simple design and assumptions and was applied successfully in many practical situations.

The basic assumption of this classifier (conditional independence) is rarely true in real-world applications (Zhang 2004). Caruana and Niculescu-Mizil (2006) applied a comprehensive comparison with other classification methods, which showed a better performance compared with other approaches (e.g., boosted tree). Friedman et al. (1997) stated that Bayes classifier requires only a small number of training data to evaluate the necessary classification parameters, which is considered an advantage.

### 2.3.3.3 $k$ -NN

$k$ -NN is one of the simplest algorithms (Mitchell 1997);  $k$ -NN classifies pixel instance  $x$  containing  $x_i$  coordinates (including an  $n$ -dimensional input space  $x = (x_1, x_2, \dots, x_n)$ )  $x \in R^n$ , where dimensions represent the values of the conditioning factors related to that particular pixel) by class values  $c_j$  of the  $k$ -closest neighboring pixels  $x_r$  surrounding  $x$  ( $c_j$  is previously assigned in the training set by a practitioner as  $fc(x_r)$ ). The nearest neighbors are defined in terms of Euclidean distance  $d(x, x_r)$ . Thus, the classifier initially calculates the distances to  $k$ -neighbors for each  $x$  instance in the training set. Subsequently, a simple voting system assigns  $c_j$  class value (landslide class) to that particular pixel by class, which predominates the neighboring instances (Eq. 2.3), or alternatively assigns its mean value to the pixel if the data are ordinal numerical [Eq. (2.4); Fig. 2.6].

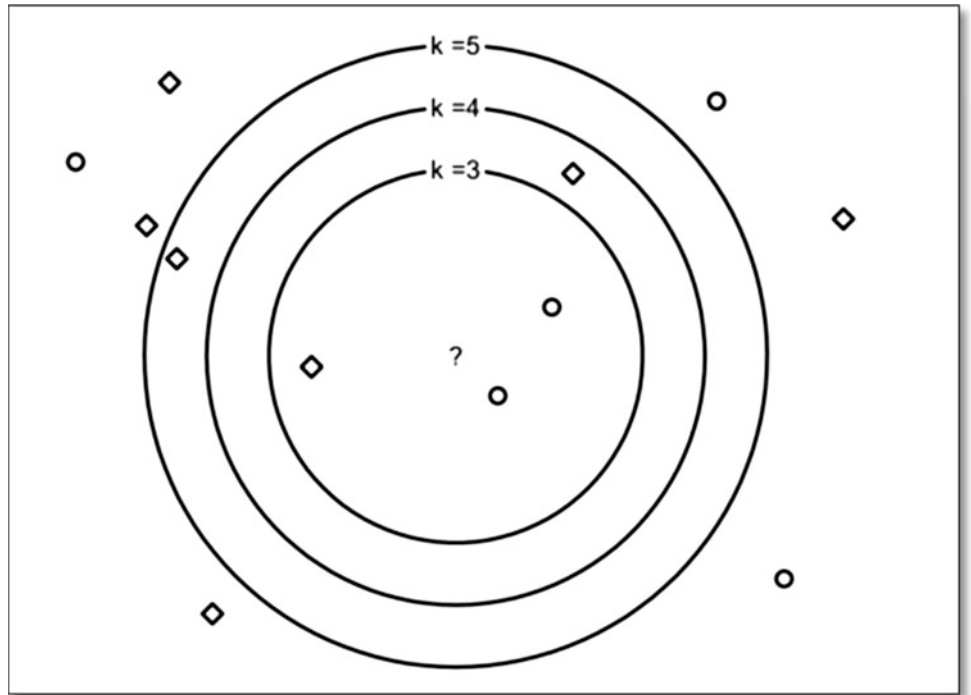
$$f'_c \leftarrow \arg \max \sum_{i=1}^k f(d(x, x_r), f_c(x_r)), \quad \forall (x \wedge x_r) \in \text{nominal data type} \quad (2.3)$$

$$f'_c \leftarrow \arg \max \frac{\sum_{i=1}^k f_c(x_r)}{k}; \quad \forall (x \wedge x_r) \in \text{ordinal data type} \quad (2.4)$$

Typically,  $k$ -NN does need conventional training/testing procedures;  $f'_c$  is simply calculated based on the remaining (testing) part of the dataset, which is similar with the training mode (Varmuza and Filzmoser 2016). The number of neighbors is necessarily an odd number ( $k = 1, 3, 5, 7, \dots$ ) to avoid even votes. Closer neighbors tend to have a greater impact; thus, it is further desirable to ponder the proximity of each neighbor, thereby upgrading to weighted  $k$ -NN (Mitchell 1997). Thus, the algorithm becomes global (Sheppard's method) but requires sorting and weighting of distances per each pixel element (and each conditioning factor is assigned to it) in the training set, resulting in a hardware-demanding and time-consuming procedure.

Therefore,  $k$ -NN classifier can be biased if all, relevant and irrelevant, conditioning factors are fed together to the algorithm because it builds a weighted or regular  $k$ -NN relation per each conditioning factor, thus misleading the classification. In other words,  $k$ -NN is extremely sensitive to

**Fig. 2.6**  $k$ -NN classification principle. Unclassified instance (?) is classified by the majority of neighbors into landslide (circle) or non-landslide (square) instance. Note that for  $k = 3$ , the instance is classified as landslide; for  $k = 4$ , the instance remains unclassified (2:2 even votes); and for  $k = 5$ , the instance is classified as non-landslide



the relevance of the conditioning factor with landslide occurrence; thus, a strict attribute selection should be performed prior to the analysis. Alternatively, Euclidean distance axis can be stretched in the case of weighted  $k$ -NN, so different conditioning factors will have different weights according to their relevance. Nonetheless, this process does not solve the computational demands of this algorithm, especially when mixed data types are present, which results in a double procedure because of varied distance calculations.

Nevertheless, the distances are the classification criteria; thus,  $k$ -NN algorithm is straightforward and does not require a true black-box model. Furthermore, the algorithm can originate from a remarkably sparse data that are randomly sampled throughout the training set, which are sometimes convenient but are of little relevance to the concept of landslide assessment and prediction of the spatial landslide distribution.  $k$ -NN classifier is also convenient for experimenting because it only needs one parameter, which is the number of  $k$ -neighbors  $k$ , to be optimized.

#### 2.3.3.4 SVM

Support vector machine (SVM) is a supervised learning method that analyzes data and recognizes patterns. In other words, given a labeled training data (supervised learning), the algorithm outputs an optimal hyperplane that categorizes new samples (Vapnik and Vapnik 1998). SVM classifies the original entry space into a more detailed feature space using training samples. Thereafter, the ideal hyperplane within this feature space is assigned by doubling the class boundary margins (Abe 2005). The nearest training samples to the ideal hyperplane are called support vectors. After determining the decision surface, it will be used to classify new data. Consider a training dataset of instance-labeled pairs  $(x_i, y_i)$  with  $x_i \in R^n$ ,  $y_i \in (1, -1)$ , and  $i = 1 \dots m$ . In this study of landslide and man-made cut slope detection,  $x$  is a vector of entry space, which includes slope, curvature, hill shade, soil type, distance to road, and altitude.

The two classes  $(1, -1)$  stand for the pixels of landslide and man-made slope, respectively. Finding the ideal hyperplane separation that discriminates the two classes from the set of training data is the aim of SVM classification. In case of linear data separation, a separating hyperplane can be defined as follows:

$$y_i(w \cdot x_i + b) \geq 1 - \delta_i, \quad (2.5)$$

where  $w$  is a coefficient vector that determines the orientation of the hyperplane in the feature space,  $b$  is the offset of

the hyperplane from the origin, and  $\delta_i$  is the positive slack variables (Cortes and Vapnik 1995).

Determining an optimal hyperplane leads to solving the following optimization problem using Lagrangian multipliers (Samui 2008):

$$\text{minimize} \quad \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j (x_i x_j) \quad (2.6)$$

$$\text{subjected to} \quad \sum_{i=1}^n a_i y_i = 0, \quad 0 \leq a_j \leq C, \quad (2.7)$$

where  $a_i$  are Lagrange multipliers,  $C$  is the penalty, and the slack variables  $\delta_i$  allow the violation of penalized constraint.

The decision function, which is used to classify new data, can then be written as

$$g(x) = \text{sign} \left( \sum_{i=1}^n y_i a_i x_i + b \right). \quad (2.8)$$

In some cases, where determining the separating hyperplane is impossible through the linear kernel function, data entry can be transferred to a high-dimensional feature space using a few nonlinear kernel equations. The classification decision equation is then written as

$$g(x) = \text{sign} \left( \sum_{i=1}^n y_i a_i k(x_i x_j) + b \right), \quad (2.9)$$

where  $k(x_i x_j)$  is the kernel function.

#### 2.3.3.5 DT

Decision tree (DT) is a nonparametric supervised learning method that is usually used for data mining. In this method, a series of decisions are made to segment the data into homogeneous subgroups. DT model is more likely to look like a tree with several branches. In some cases, DT can be remarkably complex with the involvement of a large number of splits and nodes. DT aims to build a model that can estimate the value of a target variable depending on several input variables regarded as training samples. The tree model can be learned by breaking the main set into subsets depending on an attribute value test. Thereafter, this operation is repeated for each derived subset in a repetitive manner called recursive partitioning (Last et al. 2002). Once the subset of all nodes has the same value as the target variable, or when the breaking operation does not add any more value to the predictions, the recursion step is

considered complete. The main objective of using a tree-building algorithm is to determine a set of if-then logical or split conditions.

### Important DT Parameters

The minimum number of samples needed per node that are defined by the parameter is called Min sample count. Finding the optimum-sized tree can be challenging, because the prediction of a tree with a few splits may be inaccurate. Conversely, a tree with a multitude splits will add unnecessary complications to the analysis operation. Cross-validation can be performed to address this issue by setting cross-validation folds using eCognition parameters. In this process, computing for the classification tree is done by learning the samples and then evaluating the prediction accuracy by testing these samples. Cross-validation gives a poor indication in cases where the test sample cost is more than the learning sample cost and a good indication in instances with a different-sized tree.

#### 2.3.3.6 RF

Random forest (RF) is a machine-learning algorithm used for the purpose of classification and regression, as proposed by (Breiman 2001). This supervised method was successfully applied in several areas and domains. Remote sensing field is one of the major domains and has been applied in landslide detection (Chen et al. 2014), urban trees (Puissant et al. 2014), agricultural soil mapping (Grimm et al. 2008), and biomass estimation (Mutanga et al. 2012). RF is a multiple DT classifier based on classification and regression tree [CART; (Breiman et al. 1984)]. This method implements a bootstrap sampling for each DT, which enables the estimate calculation of errors to be based on the remaining instances, which is known as “out-of-bag” (OOB). RF applies a different process to determine the best split threshold, in comparison with CART. RF is considered as a random subset of the original set of the feature, whereas CART considers all variables at each node. Users can estimate the variables per the number of node by using the square root of the total variable number. Two mechanisms, sampling and the use of random variables for each node, generate significantly different uncorrelated trees. Furthermore, having a relatively large number of trees is necessary to obtain the full variability of the training data, which gives good classification performance with high accuracy. The final step is assigning a feature into a class by considering the votes of all the trees in the forest. The class will then be assigned based on majority voting. The RF package (Liaw and Wiener 2002) for the open-source statistical language R (R Development Core Team 2013) was used for all experiments in this study.

Random forest (RF) has several advantages. First, RF is a nonparametric method; thus, the values of variables are not required to follow a particular statistical distribution. Second, it is insensitive to overfitting and noise. Furthermore, RF is relatively fast compared with other techniques, such as the boosted method (Breiman 2001). The calculation time for training RF is defined by Eq. (2.10).

$$cT\sqrt{MN} \log N, \quad (2.10)$$

where  $c$  is a constant dependent of data complexity (i.e., small or large dataset),  $T$  is the number of tree,  $M$  is the number of variables, and  $N$  is the number of instances (Breiman 2003).

When RF and SVM, whose complexities vary between  $N^2$  (when  $c$  is small) and  $N^3$  (when  $c$  is large), are compared (Bottou and Lin 2007), RF will give a better adaption for larger datasets. Also, RF requires less tuning (Rodriguez-Galiano et al. 2012) and can implement the actual measures of variables, which can be estimated by alternating the value of variables on the OOB sample and calculating the difference in OOB errors before and after the alternation process. Those measures are used to analyze and interpret the classification (Rodriguez-Galiano et al. 2012) and define the type of sensor (Guo et al. 2011). Otherwise, defining the scale of segmentation (Duro et al. 2012) is more suitable for identifying a particular geographic object.

#### 2.3.3.7 Landslide and Cut Slope Detection

The supervised landslide detection and cut slope detection were done in two successive steps. The first step was training the classifier with an adequate number of samples. The samples were selected randomly based on landslide inventory data. In this study, 60% of the samples were selected for training the classifier for the classes: landslide, cut slope, and non-landslide. These samples were examined based on the aerial photographs, slope, and hill-shade layer of the study area to ensure that each sample was selected accurately. The classifiers were then trained using these samples. In the second step, in each classifier method, several user-defined parameters should be carefully selected. In this study, the user-defined parameters were selected based on a trial-and-error approach. Table 2.1 shows the classifiers with their user-defined parameters that were selected for supervised landslide detection.

#### 2.3.3.8 Validation

The efficiency and quality of the presented methodology for each study and research must be properly examined and tested, which can be achieved by a proper validation technique. In this study, the validation was done in three steps: The first step is to examine the classification results visually;

**Table 2.1** Selected value of each parameter for each classification algorithm used

Classifier	Parameter	Selected value
Bayes	NA	NA
<i>k</i> -NN	<i>K</i>	1
SVM	Kernel type	Linear
	<i>C</i>	1
DT	Depth	0
	Min sample count	0
	Use of surrogates	Yes
	Max categories	16
	Cross-validation folds	3
	Use of 1 standard error (SE) rule	No
	Truncate pruned tree	Yes
RF	Depth	0
	Min sample count	0
	Use of surrogates	Yes
	Max categories	16
	Active variables	0
	Max tree number	50
	Forest accuracy	0.01
	Termination criteria type	Both

the second step is to transfer the methodology to a different subset of the study area and examine the ability of the method to detect landslides and cut slopes; and the final step is field validation, which is a site visit to the field, and is necessary to confirm the location and boundary of few landslides detected by the methodology presented in this study.

### Visual Interpretation

In the first validation step, where the results of each classifier are examined visually, some classifiers (e.g., Bayes) produced results with high level of uncertainty and misclassification, thereby making visual interpretation useful for the rejection of the result of such classifier. In addition, a few classifiers have numerous user-defined parameters (e.g., DT and RF), which need to be fine-tuned; visual interpretation is considerably useful for this purpose.

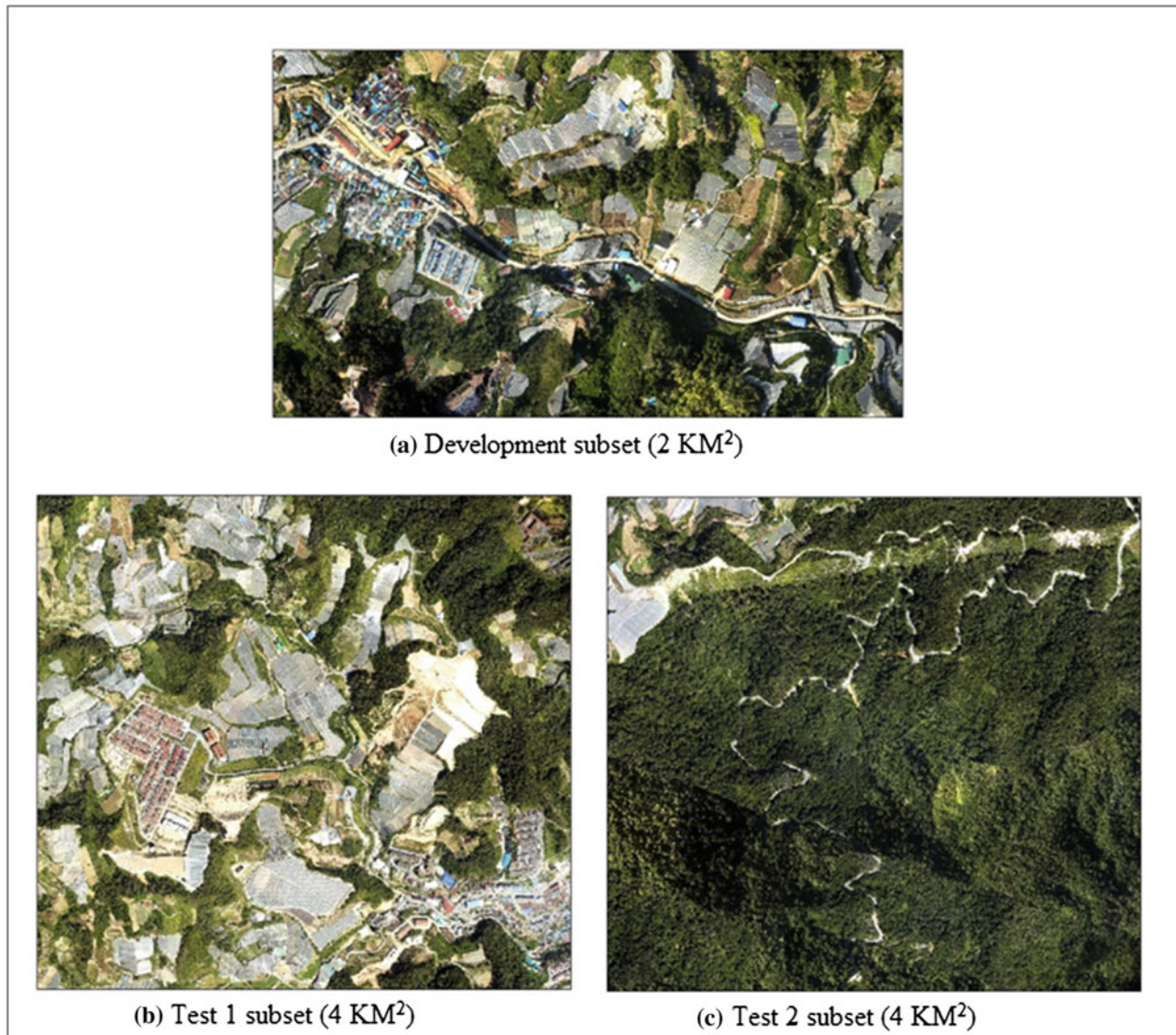
### Transferring to Testing Subset

The efficiency and quality of the presented methodology should be properly examined and tested. In this research, the study area was divided into three different subsets, and the same methodology was replicated on the testing sites to examine its validity and accuracy. Dividing the study area into three different subsets was implemented carefully. The

first subset (Fig. 2.7a) has the smallest land area (2 km<sup>2</sup>) and was used to develop the method. The first subset was easy to process and interpret because of its small size. Moreover, the training subset has various types of land-use classes (e.g., landslides, cut slopes, vegetation, and urbanization) that are fairly distributed over the entire study area, thereby making it a challenging task during the development of the methodology. The other two subsets, which have a larger land area of 4 km<sup>2</sup>, were used to test the proposed method. The first testing subset (Fig. 2.7b) is considerably similar to the training subset but larger in size. Conversely, the second subset (Fig. 2.7c) has different distribution of land-use classes; the vegetation covers almost 85% of the entire area.

### Field Validation

Multiple field visits were conducted using Global Positioning System (GPS) devices to examine the location and the boundary of detected landslides. Documenting these landslides in the field was challenging, because most of the landslides are in private farms, and other landslides are within thick-vegetated forests. Only landslides parallel to the road or in open areas were well documented through multiple field inspections. Most of the landslides were covered by vegetation and became invisible because of the rapid growth of vegetation in tropical areas, thereby posing a new



**Fig. 2.7** Subsets of the study area

challenge for documentation. Figure 2.8 shows some examples of landslides from the study area of Ringlet.

### Accuracy Assessment

Accuracy assessment is based on the comparison of supervised object-based classification result with an actual landslide inventory map. Actual landslide inventory data are generally derived from ground truth data, ground reference data, or other dependable datasets. Performing accuracy assessment of the features detected through remote sensing images is highly pertinent (Lillesand et al. 2004).

Ground truth information (or reference data source) consists of gathered observation about features and phenomena that are captured by data. In terms of validation, ground truth information is used for the accuracy assessment of landslide inventory map. Land cover maps from remotely sensed data have minor practical value without accuracy assessment. Accuracy assessment aims to evaluate the pattern classification landslide location map. In addition, this assessment reports the importance of classification schemes, so other researchers can easily interpret and apply them. Accuracy assessment involves two steps:



**Fig. 2.8** Photographs taken during field validation

1. Collection of ground truth data. Ground truth data are independent from the training data that have been used in the process of image classification. Based on the methodology, ground truth data are collected in specific locations to be found in remotely sensed data. Sources of

ground truth data consist of high-spatial resolution remotely sensed data, such as aerial photography, a high-resolution satellite imagery (including Ikonos, Quick-bird, and Worldview2), or a field survey measurement using GPS.

2. Comparison. Ground truth data are compared with data that are determined from different classes to classify the imagery.

A confusion matrix is a cross-tabulation of the classified and actual class labels for the study area (Foody 2004). This matrix is a square array of dimension  $r \times r$ , where  $r$  is the number of categories. Confusion matrix represents the correlation between two samples of measurement from the classified region. The overall, user, and producer accuracies, and the kappa coefficient can be measured using confusion matrix. The overall accuracy is attained by dividing the aggregate of the main diagonal entries of the confusion matrix by the entire number of samples. The kappa coefficient ( $K$ ) was measured using Eq. (2.11).

$$K = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad (2.11)$$

## 2.4 Results

### 2.4.1 Landslide Detection Results

Several supervised classification methods were applied for landslide detection, including Bayes, DT, RF,  $k$ -NN, and SVM. The results of classification using  $k$ -NN, DT, and Bayes algorithms showed poor accuracy results, because most of the landslides were not detected correctly. Furthermore, landslides were misclassified as man-made cut slopes and bare lands in some cases. SVM and RF algorithms performed better compared with the previous three; many landslides were correctly detected, positioned, and delineated. Two testing sites were used to evaluate the consistency of the used classifiers for landslide detection. This section presents the results of landslide detection in the two sites. The first testing site contained several landslides, while few were detected in the second testing site.

### 2.4.2 Results of Landslide Detection in the Training Site

#### 2.4.2.1 RF

In this study, RF classifier was also used for landslide detection. Results of RF landslide detection are shown in Fig. 2.9. Initial observations for the map indicate that this method performed better than  $k$ -NN, DT, and, Bayes

algorithms. Evidently, most of the landslide inventories were detected accurately. Few cut slopes were misclassified as landslides, as shown in the northeastern part of the study area. RF detected 30 out of the 40 landslide inventories found in the study area. However, some landslides were undetected despite being visible in the middle part of the study area. The challenge with RF algorithm is that it requires the fine-tuning of several parameters. The current study optimized the parameters by trial-and-error approach. However, best results were not achieved. Using more robust algorithms for fine-tuning the parameters of RF could improve the landslide detection results.

#### 2.4.2.2 SVM

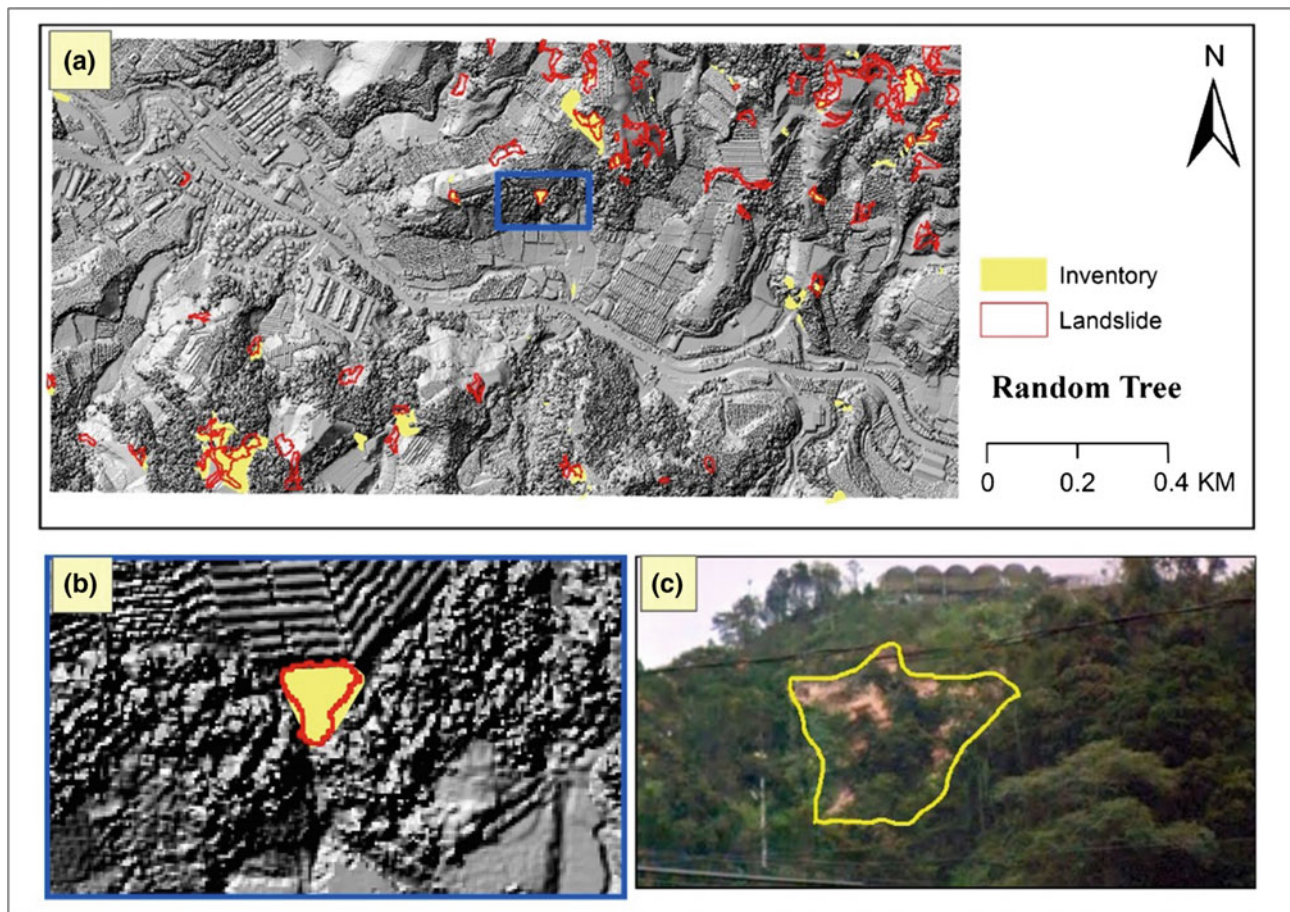
Support vector machine (SVM) has been widely used for landslide susceptibility mapping, and its advantages are well established in several landslide studies. In this study, SVM was used for landslide detection. Figure 2.10 shows the results of SVM landslide detection. Results indicate that SVM is the best among the other four methods; most of the landslides were detected, few cut slopes were misclassified as landslides, and few landslides were undetected. SVM works on the concept of optimization and error reduction; therefore, it performs well for landslide detection. Having accurate landslide inventories is difficult; thus, a methodology that can detect and reduce errors is significantly important. SVM utilizes this concept; thus, it detected landslides accurately, leaving only few undetected.

Figure 2.10c shows a landslide photograph taken during the field visit. The capturing angle does not show the entire boundary of the landslide; thus, proper documentation of the landslide was challenging.

### 2.4.3 Results of Landslide Detection in Testing Site 1

#### 2.4.3.1 RF

The result of RF landslide detection for Testing Site 1 is shown in Fig. 2.11. The first examination of the map shows that several landslides were accurately detected, and few cut slopes were misclassified. This shows the main difference between the result of RF and those of other methods presented previously. The RF algorithm tends to separate landslides from cut slopes better than Bayes, DT, and  $k$ -NN techniques. Although RF requires several user-defined parameters for fine-tuning, its results are better than the other techniques, using several combinations of the parameters.



**Fig. 2.9** Detected landslides using RF classification algorithm

#### 2.4.3.2 SVM

A landslide inventory map for Testing Site 1 was produced by SVM method, as shown in Fig. 2.12. This method is shown to be suitable for the training site. When the results were examined, the algorithm produced a good landslide inventory map; the landslides were accurately detected, and few cut slopes were misclassified.

### 2.4.4 Results of Landslide Detection in Testing Site 2

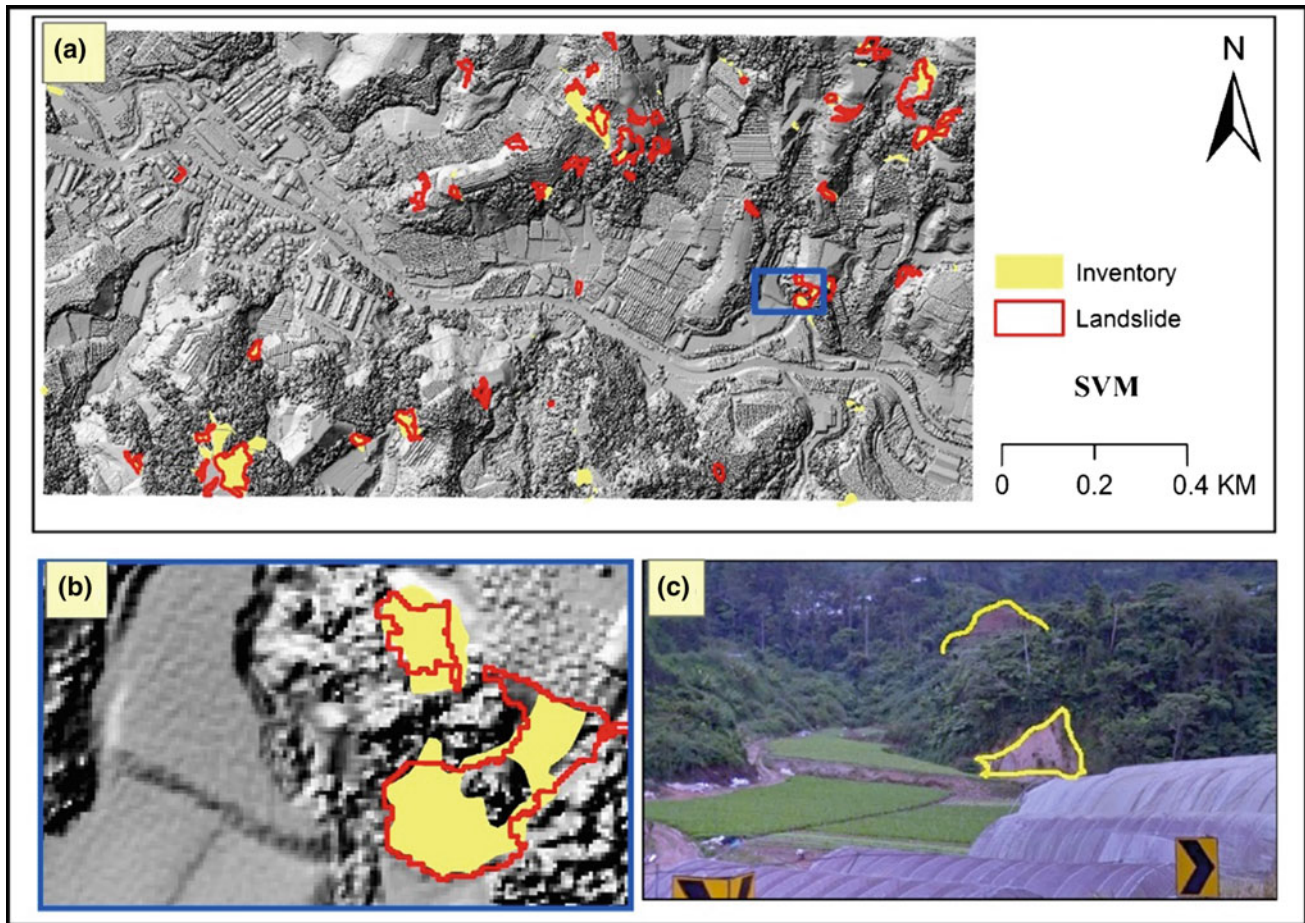
#### 2.4.4.1 RF

Figure 2.13 shows the results of landslides detected using RF for Testing Site 2. Results are far from excellent, as many

landslides were undetected. However, this method remains better than Bayes, DT, and  $k$ -NN, because few cut slopes were misclassified. This result shows that RF is a good classifier for landslide detection in the presence of man-made slopes.

#### 2.4.4.2 SVM

Figure 2.14 shows the result of SVM landslide detection in the presence of man-made slopes for Testing Site 2. SVM produced an accurate landslide inventory map with few undetected landslides. In addition, results show that SVM is better than RF through visual examination. Several landslides in the upper left part of the study area were detected by SVM but not by RF. However, both the SVM and RF performed well in landslide detection when man-made slopes are present in the study area.

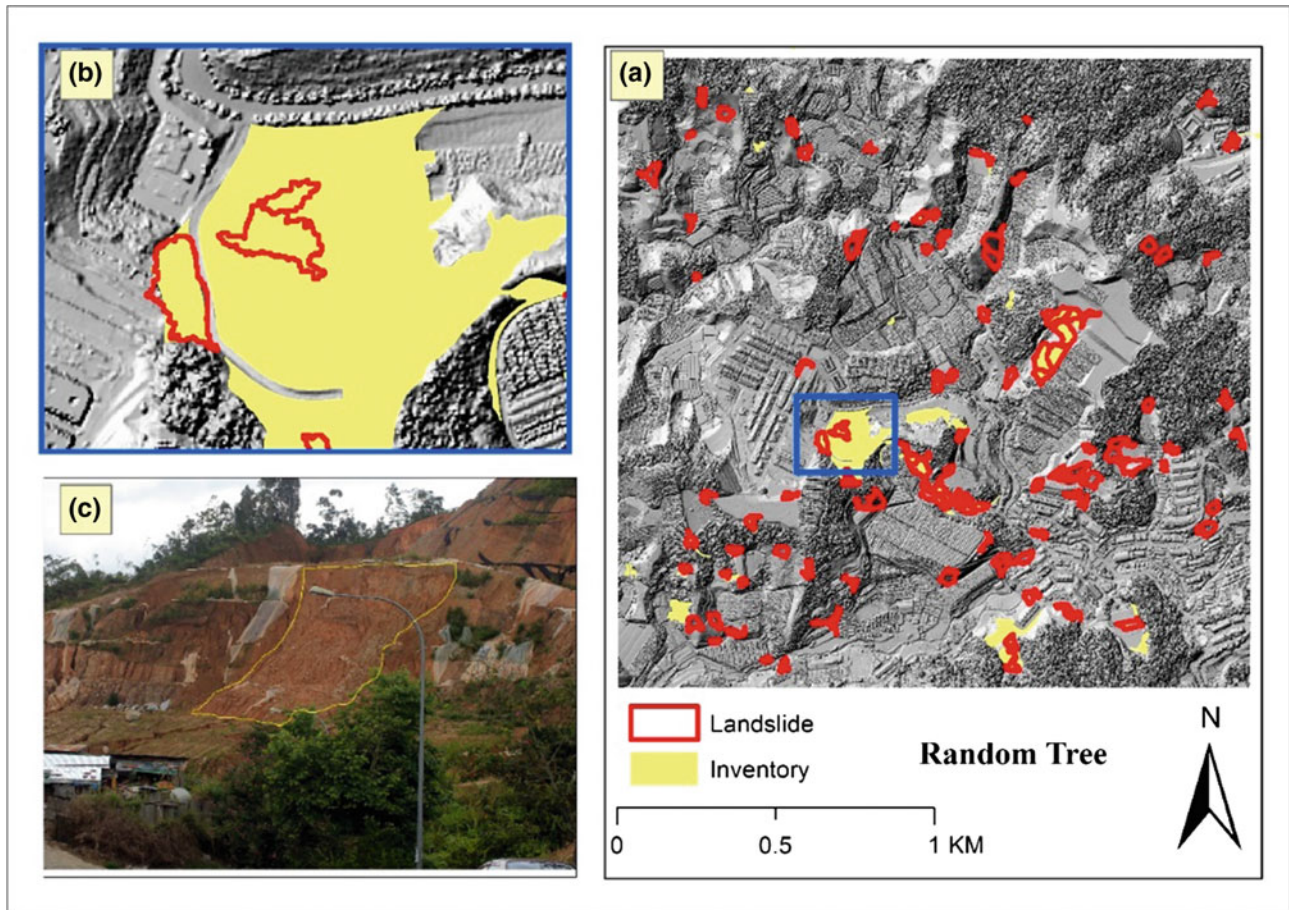


**Fig. 2.10** Detected landslides using SVM classification algorithm

#### 2.4.5 Cut Slope and Landslide Detection Results

This section presents the results obtained from landslide and cut slope mapping of the training site and two testing sites. Figure 2.15 shows the detected landslides and cut slopes by SVM and RF models. These two algorithms were considered as the most effective among others because of their high accuracy. The map shows the detected landslides in dark blue, the cut slopes in light blue, and the other features in pink. Both landslides and cut slopes are randomly distributed in the study. However, the study area exhibited a clustered pattern of landslides in the upper right corner, characterized by having a high slope, concave curvature, and is mostly barren.

Figures 2.16 and 2.17 show the landslides and man-made slopes of the two testing sites. The landslides are shown in dark blue, whereas man-made slopes are highlighted in light blue. The landslides and cut slopes are randomly distributed in the study area. The area has large and small landslides and cut slopes. Landslides may also vary in types in this study such as landslides and debris flows. Figure 2.17 illustrates few landslides and cut slopes in Testing Site 2. Most of the landslides and man-made slopes are located in the north part of the area; the middle and south parts are mostly forested area. Some landslides may have occurred in forest area, which could not be detected because LiDAR point clouds only have one return. Multiple LiDAR data returns are important to detect landslides in forested areas. Overall, 123



**Fig. 2.11** Detected landslides using RF classification algorithm

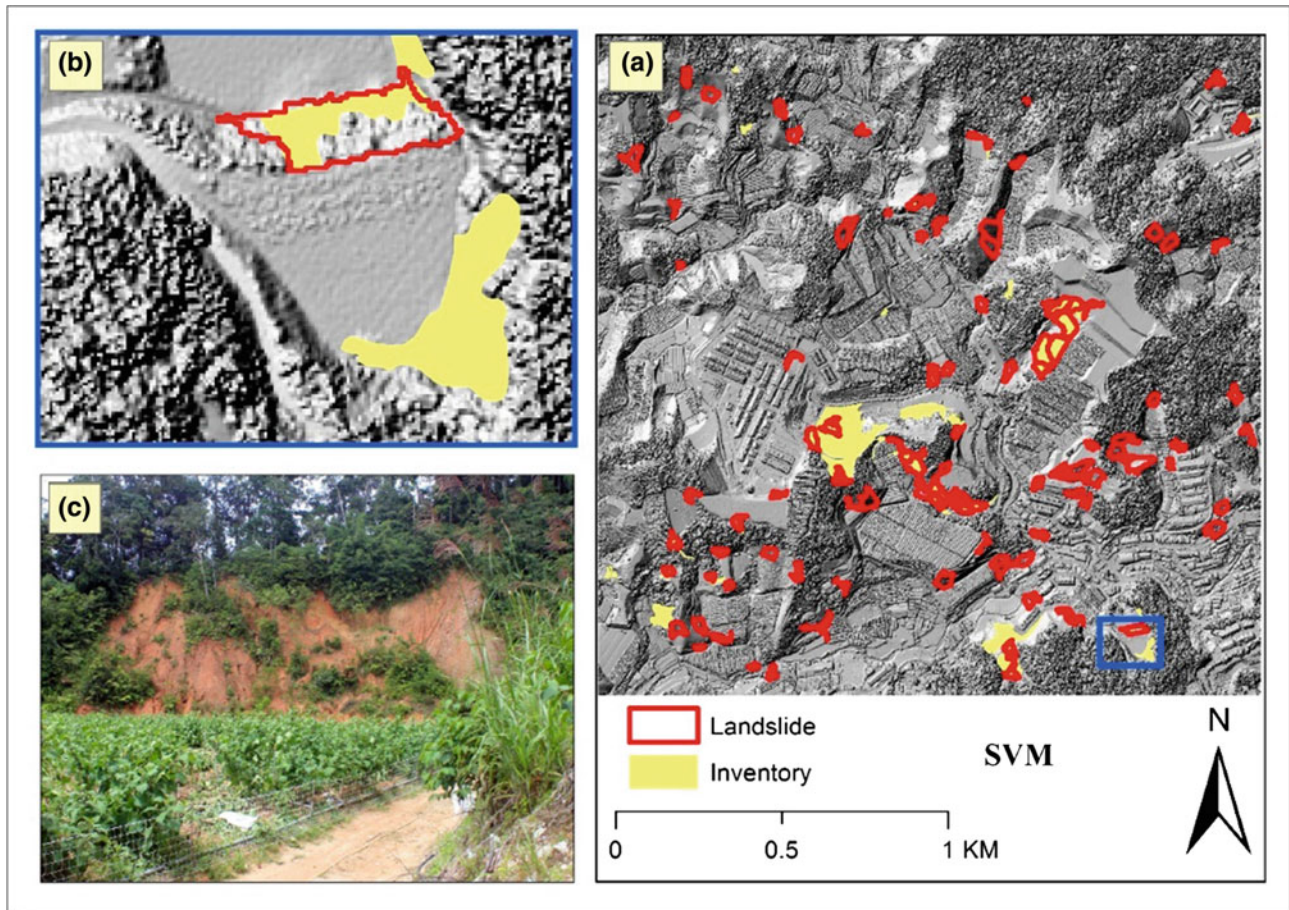
landslides and 205 cut slopes were detected in the first testing site. Similarly, 18 landslides and 51 man-made slopes were detected in the second testing site.

#### 2.4.6 Results of Image Segmentation

Once the input layers were prepared, spectral and LiDAR-based features were combined in one raster dataset for segmentation. Then, a multiresolution segmentation algorithm was utilized for segmentation. The parameters of segmentation were set as scale (60), shape (0.1), and compactness (0.5). These values were selected based on trial-and-error experiments in eCognition software. Segmentation result of the training site is shown in Fig. 2.18.

Landslide features are accurately delineated. Accurate segmentation is important for efficient landslide detection by various features. For example, in Fig. 2.18a, the segments show that the landslide scarp is accurately defined, whereas in Fig. 2.18b, the landslide scarp is only partially defined. Moreover, in Fig. 2.18c, the landslide scarp is defined inaccurately.

Some landslides are defined accurately because of fewer variations in slope, curvature, and altitude values. Landslide scarps are defined inaccurately, because the slope, curvature, and height values vary significantly within the landslide objects. Therefore, one landslide scarp may be segmented as two or more landslide objects, thereby reducing the accuracy of landslide detection as several spatial features can be useless.

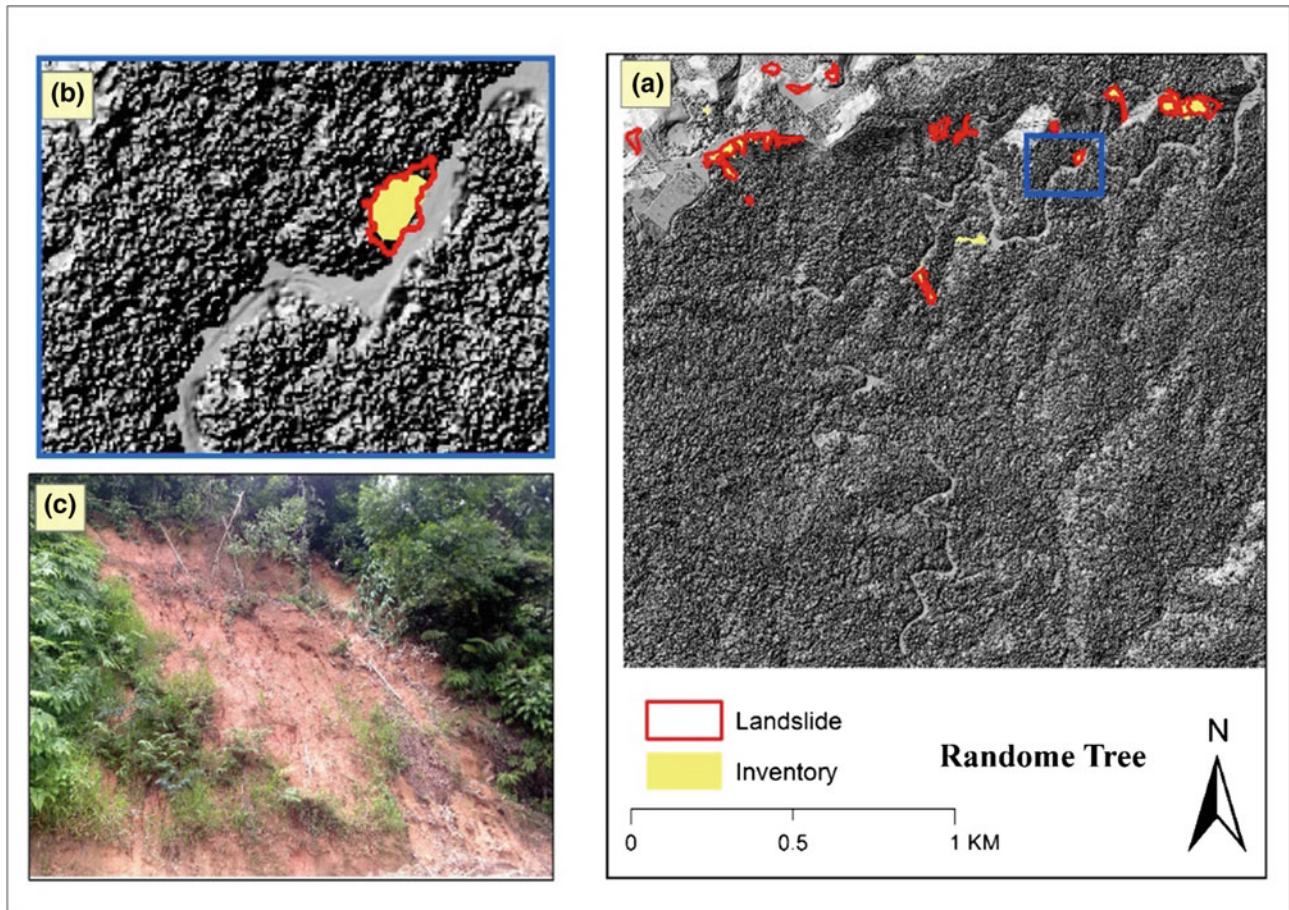


**Fig. 2.12** Detected landslides using SVM classification algorithm

#### 2.4.7 Results of Accuracy Assessment

Table 2.2 shows the overall accuracies and kappa indices of landslide detection using several classifiers and datasets. For the training site, the SVM method had the highest overall accuracy (0.90) and kappa index (0.83), in contrast to the DT algorithm that had the lowest overall accuracy and kappa index of 0.61 and 0.37, respectively. In general, the accuracy assessment shows that SVM and RF performed well in landslide detection compared to other methods. For Testing Site 1, the highest and lowest overall accuracies were 0.80

and 0.61 for SVM and Bayes methods, respectively. Moreover, the highest and lowest kappa indices were 0.74 and 0.33 for the same classifiers, respectively. Thus, RF is considered as a good classifier for landslide detection. Results confirmed that SVM and RF are the best methods for landslide detection. For Testing Site 2, RF had the highest overall accuracy of 0.91, followed by SVM with 0.90. The lowest overall accuracy was achieved by Bayes algorithm with 0.65. The kappa indices indicate that SVM is better than RF and other methods. The kappa index of SVM and RF is 0.85 and 0.80, respectively. Quantitative assessments



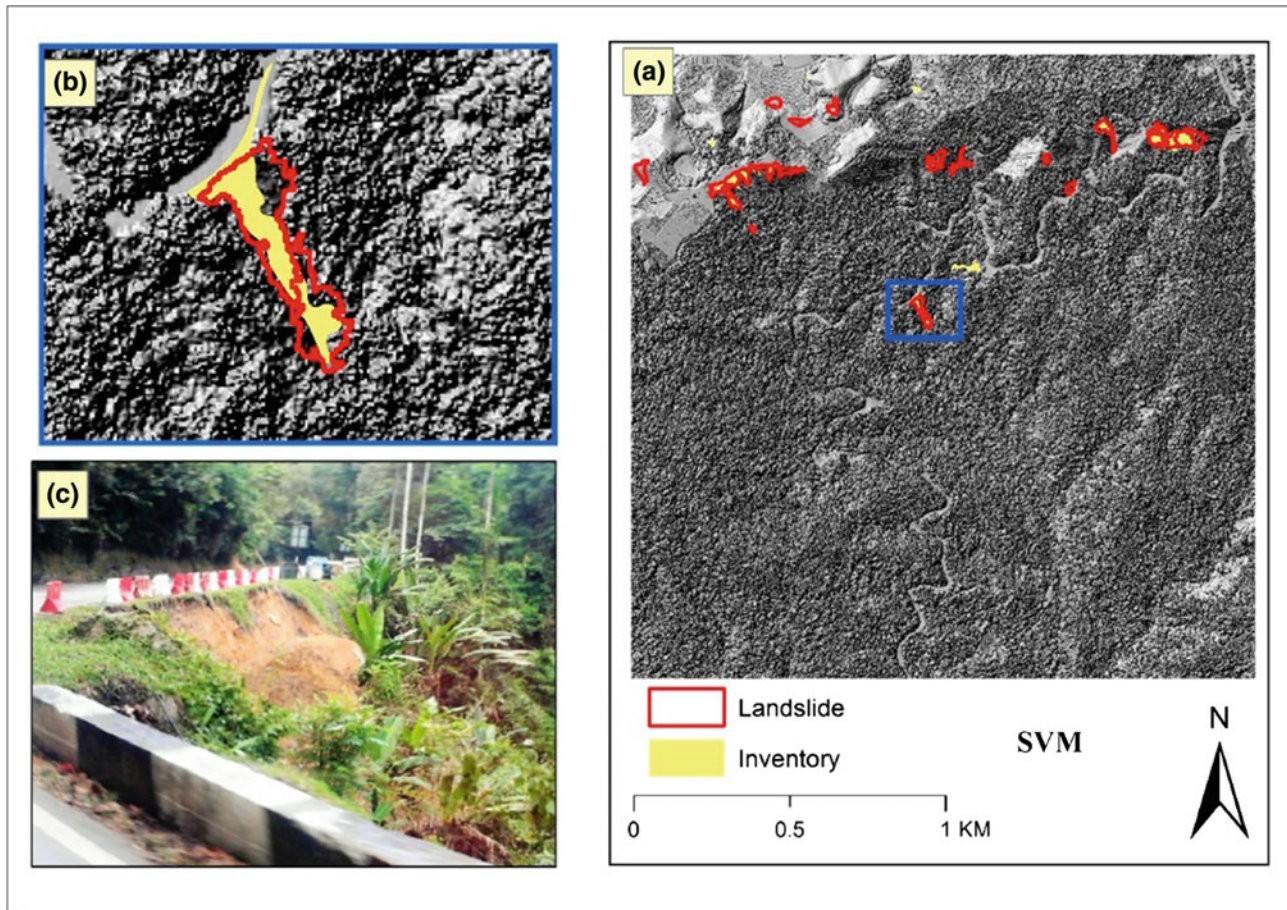
**Fig. 2.13** Detected landslides using RF classification algorithm

show that SVM and RF can be good classifiers for supervised landslide detection in LiDAR data and orthophotos.

The study aimed to detect landslides in the presence of man-made slopes and produce an accurate inventory map. Man-made or cut slopes create challenges in landslide detection because their geometry and geomorphology are relatively similar. This study attempted to separate these classes. Table 2.3 shows the user and producer accuracies obtained from various classifiers for cut slope and landslide classes. Although Bayes method achieved the highest user accuracy, evaluation of simultaneous user and producer accuracies is important. This evaluation ensures that the detected landslides are accurate and that only few landslides will be undetected. The highest user and producer accuracies for landslide class were achieved by Bayes and SVM

methods, whereas the highest user and producer accuracies for cut slope were observed for Bayes (1) and RF methods (0.90). However, SVM and RF achieved relatively high user and producer accuracies simultaneously, indicating a good classification of landslides and cut slopes. SVM performed better than RF for landslide and cut slope classification.

For Testing Site 1, the highest user and producer accuracies for landslide class were achieved by RF and Bayes algorithms, whereas the highest user and producer accuracies for cut slope were found for SVM and RF algorithms. Kappa indices showed that SVM and RF are best for landslide and cut slope separation. The highest kappa index was achieved by SVM (0.78, 0.78) and RF (0.67, 0.83) for landslide and cut slope classes, respectively. In addition, the user and producer accuracies and kappa indices for Testing Site 2



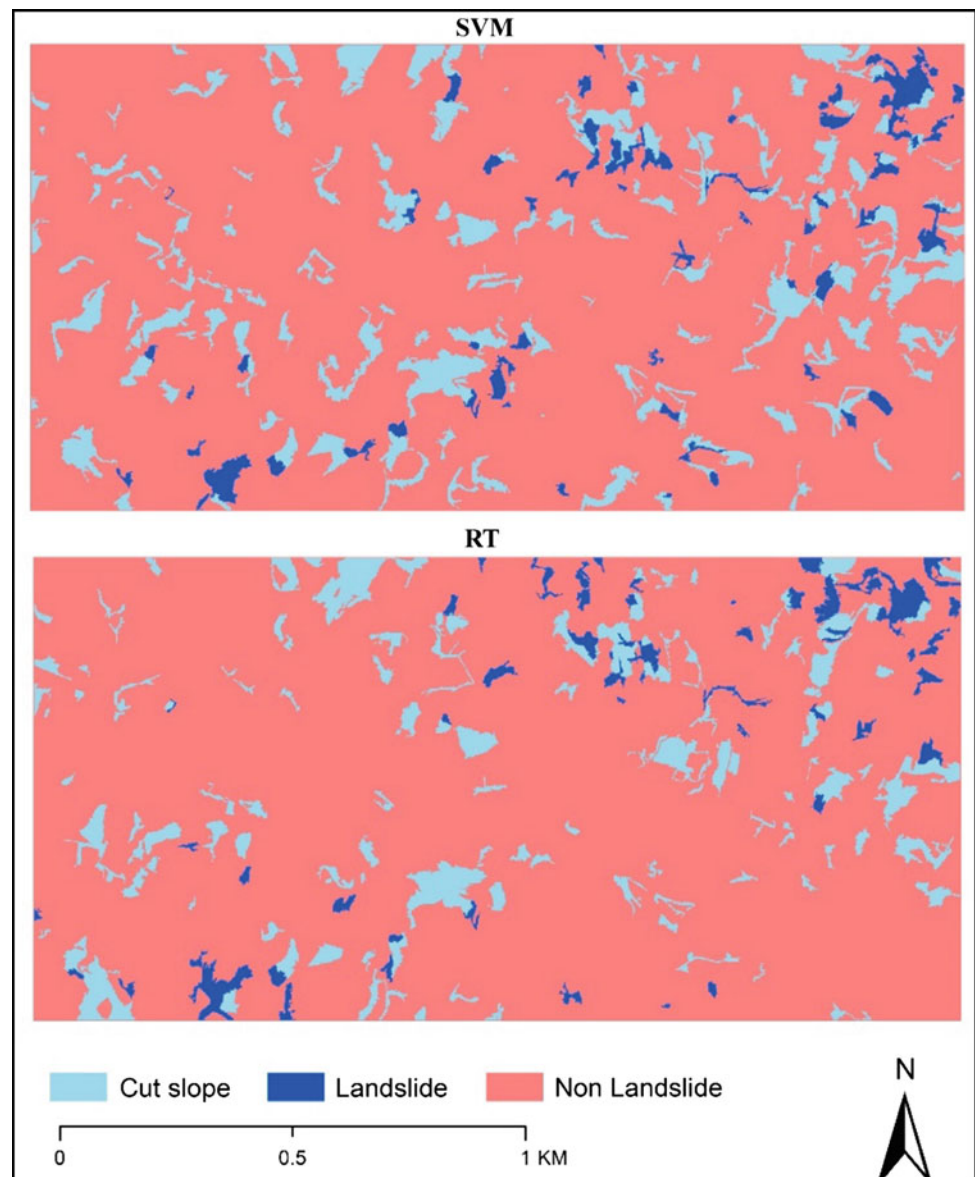
**Fig. 2.14** Detected landslides using SVM classification algorithm

confirmed that SVM and RF classifiers are the best algorithms for landslide detection. Overall assessment shows that the separation between cut slopes and landslides using LiDAR data and orthophotos through supervised classification is possible and can be improved. The current study achieved satisfactory results of landslide detection and separation between landslides and man-made slopes; however, further research is needed to detect the type of and improve the accuracy in cut slopes. The proposed supervised framework provides a rapid and efficient guideline for landslide mapping, which is valuable for landslide susceptibility mapping, and hazard and risk assessments.

## 2.5 Discussion

Several methods of determining segmentation parameters, such as supervised and Taguchi approaches (Gibril et al. 2016), were reviewed. In supervised approaches, segmentation is usually optimized based on multiple features found in the image of the study area. Conversely, in Taguchi approaches, segmentation parameters are optimized for a single feature only. Because several types of landslides are present in the study, optimizing the parameters for only one feature without considering the type of landslide created a huge challenge for the Taguchi approach. Supervised method

**Fig. 2.15** Landslide and cut slope mapping (Training Site)



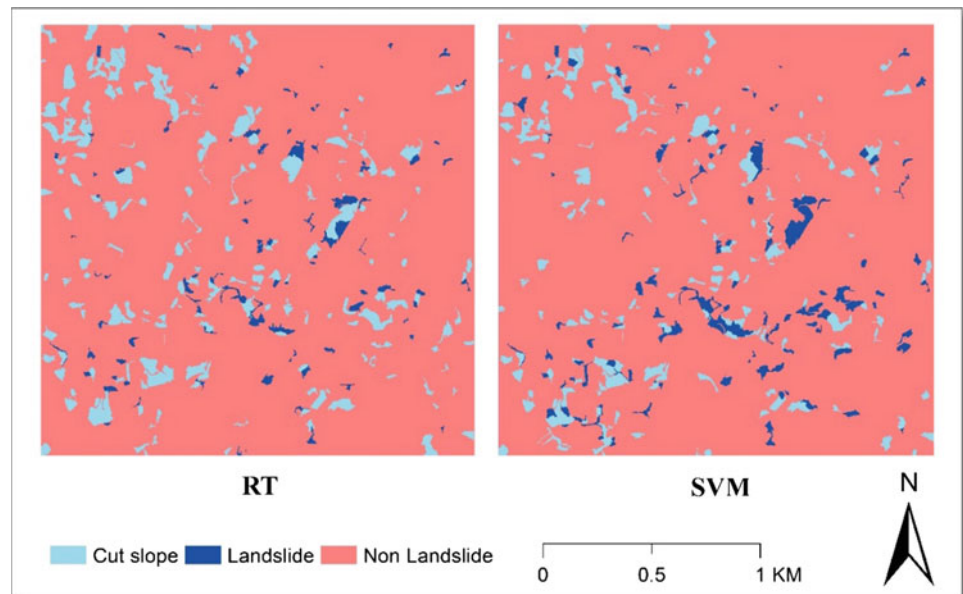
is a good option after the Taguchi method. Although supervised approaches are user independent and require less time than trial-and-error method, they also depend on the selected subsegments that aimed to be merged into a target segment.

Preparing the input layers, selecting a classifier, and the fine-tuning of user-defined parameters of the classifier are important in supervised landslide detection (W. Chen et al. 2014). The current study analyzed several input layers derived from LiDAR-based DEM and DSM for improved landslide detection. Significant layers were selected based on their importance for landslide detection using the training site and analyzed through trial-and-error approach. Overall,

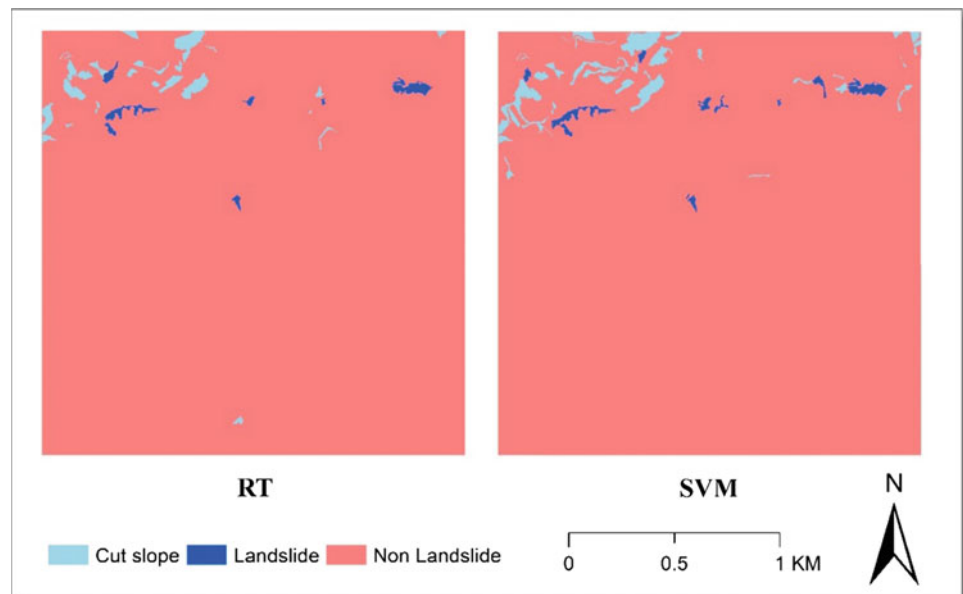
13 features, including spectral, LiDAR, spatial, and texture, were used. Furthermore, several classifiers were analyzed by measuring their accuracies for landslide detection. The user-defined parameters of the classifiers were also fine-tuned by trial-and-error method.

From LiDAR point clouds, six features were produced: DEM, DSM, height, slope, curvature, and hill shade. Training site elevation ranged from 997 to 1270 m. Conversely, the height feature showed that the height of objects in the study areas varies from 0 m (flat objects) to 100 m (hilly lands). In addition, slope of the study area includes flat and hilly lands. The slope ranged from 0 to 87°. As in

**Fig. 2.16** Landslide and cut slope mapping (Testing Site 1)



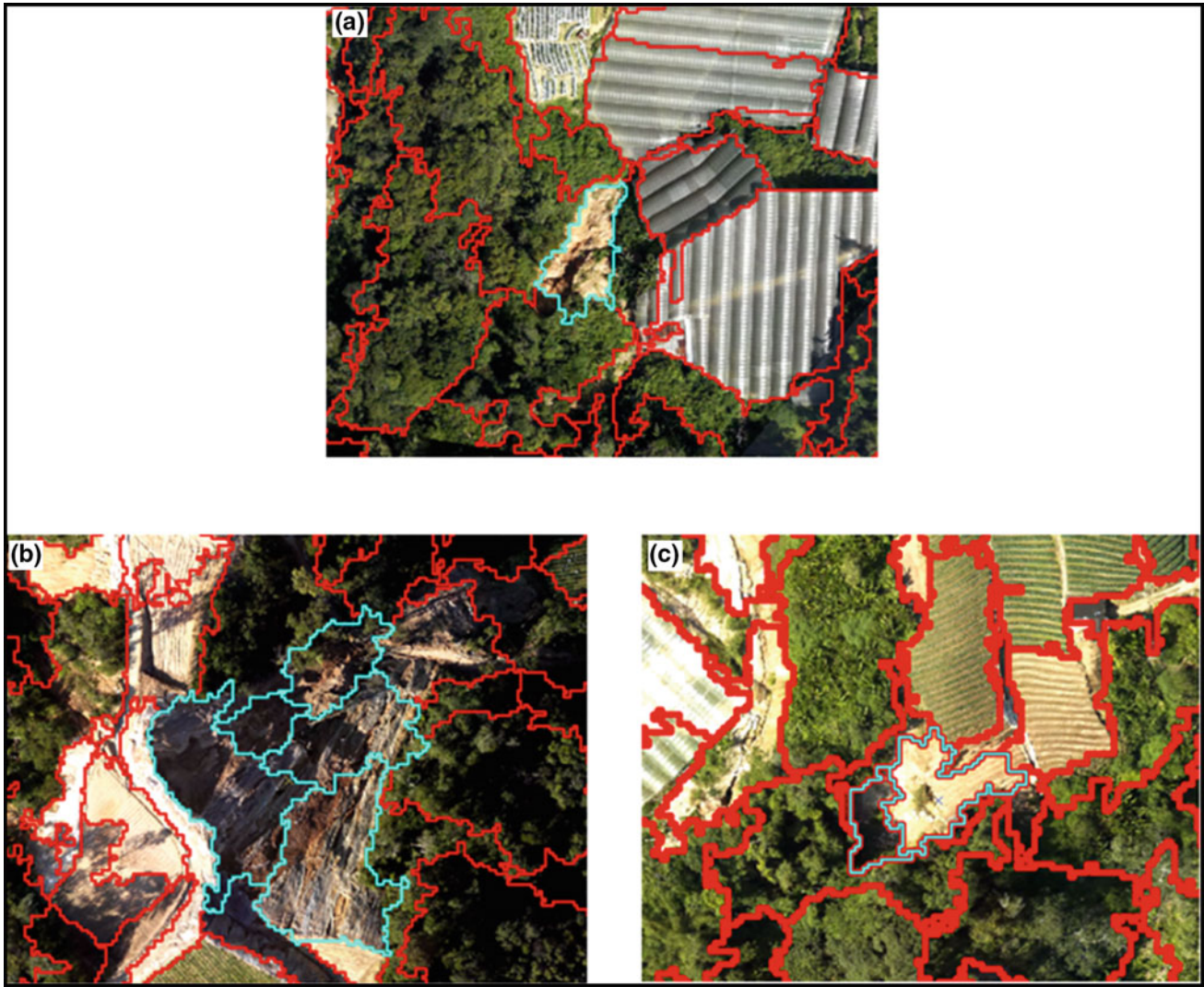
**Fig. 2.17** Landslide and cut slope mapping (Testing Site 2)



segmentation results which is shown in Fig. 2.18, the study area has flat, concave, and convex type of curvature surfaces.

Figure 2.19 shows that landslides and cut slopes are difficult to separate using only LiDAR-derived parameters because both landslides and cut slopes have relatively similar characteristics. Therefore, investigating other parameters and orthophotos is important.

Three spectral bands (R, G, and B) of the orthophotos were used for landslide detection. Orthophotos are useful information to separate landslides from other features, such as grassland, buildings, and water bodies. Analysis of typical values of spectral bands for landslide and non-landslides is presented in Fig. 2.20. The chart presents the minimum, maximum, and mean values of RGB



**Fig. 2.18** Segmentation results

**Table 2.2** Overall accuracies and kappa indices of landslide detection

Dataset	Classifier	Overall accuracy	Kappa index
Training Site	Bayes	0.75	0.46
	<i>k</i> -NN	0.65	0.4
	DT	0.61	0.37
	RF	0.82	0.7
	SVM	0.9	0.83
Testing Site 1	Bayes	0.61	0.33
	<i>k</i> -NN	0.72	0.51
	DT	0.72	0.53
	RF	0.78	0.64
	SVM	0.86	0.74
Testing Site 3	Bayes	0.65	0.37
	<i>k</i> -NN	0.71	0.43
	DT	0.78	0.56
	RF	0.91	0.8
	SVM	0.9	0.85

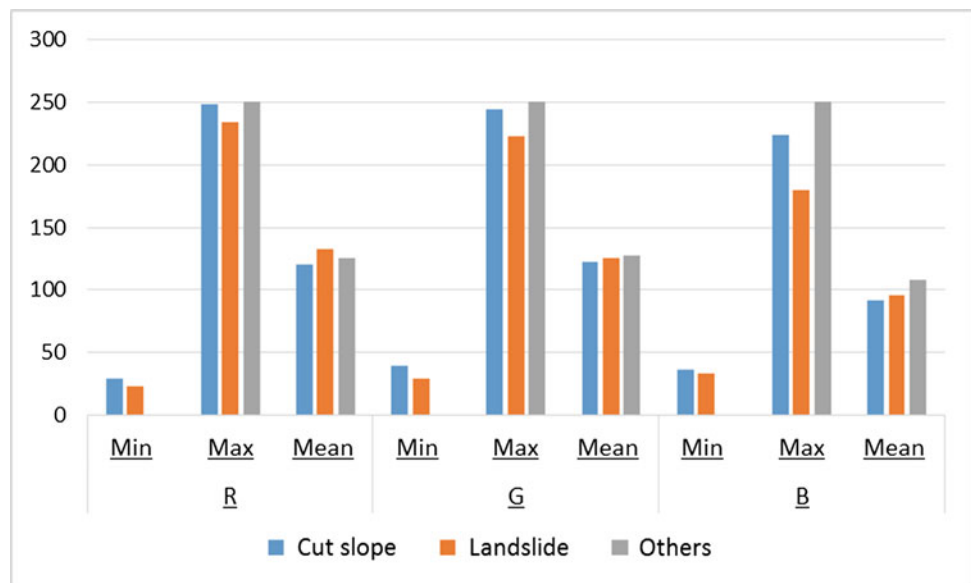
**Table 2.3** User and producer accuracies obtained from various classifiers for cut slope and landslide classes

Dataset	Classifier	Class	User	Producer	KIA
Training Site	Bayes	Landslide	1	0.5	0.44
		Cut slope	1	0.26	0.22
		Others	0.7	1	1
	<i>k</i> -NN	Landslide	0.37	0.45	0.27
		Cut slope	0.54	0.65	0.53
		Others	0.82	0.71	0.4
	DT	Landslide	0.58	0.35	0.25
		Cut slope	0.42	0.75	0.59
		Others	0.77	0.67	0.36
	RF	Landslide	0.92	0.6	0.53
		Cut slope	0.69	0.9	0.86
		Others	0.86	0.88	0.72
	SVM	Landslide	0.82	0.95	0.93
		Cut slope	0.8	0.8	0.74
		Others	0.98	0.92	0.83
Testing Site 1	Bayes	Landslide	0.85	0.2	0.15
		Cut slope	0.66	0.4	0.27
		Others	0.58	0.96	0.84
	<i>k</i> -NN	Landslide	0.77	0.45	0.37
		Cut slope	0.51	0.75	0.63
		Others	0.82	0.8	0.55
	DT	Landslide	0.6	0.45	0.34
		Cut slope	0.52	0.73	0.62
		Others	0.86	0.82	0.62
	RF	Landslide	0.94	0.58	0.52
		Cut slope	0.56	0.91	0.86
		Others	0.91	0.8	0.6
	SVM	Landslide	0.79	0.76	0.71
		Cut slope	0.69	0.71	0.65
		Others	0.93	0.93	0.82
Testing Site 2	Bayes	Landslide	1	0.33	0.26
		Cut slope	1	0.33	0.26
		Others	0.58	1	1
	<i>k</i> -NN	Landslide	0.4	0.66	0.54
		Cut slope	0.75	0.5	0.44
		Others	0.83	0.76	0.37
	DT	Landslide	0.66	0.57	0.5
		Cut slope	0.8	0.44	0.36
		Others	0.8	0.96	0.85
	RF	Landslide	1	0.71	0.67
		Cut slope	0.85	0.85	0.83
		Others	0.9	0.96	0.87
	SVM	Landslide	1	0.83	0.78
		Cut slope	1	0.83	0.78
		Others	0.83	1	1

**Fig. 2.19** Typical values of LiDAR-based features for landslide and non-landslide features. *Note* (1) slope  $\times 10$ ; (2) hill shade, intensity, and DEM  $\times 102$



**Fig. 2.20** Typical values of spectral bands for landslide and non-landslide features



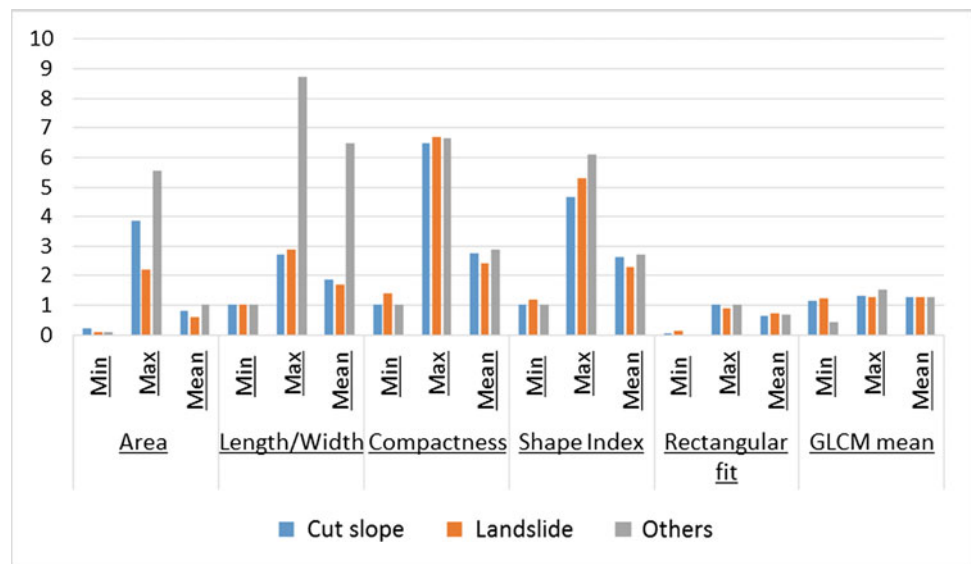
bands for landslide, cut slope, and other features in the study area. The three classes have approximately similar average, minimum, and maximum values of RGB bands. However, slight differences in these bands can also separate the three classes at some extent depending on the classifier used. When these parameters are combined with other features, they can work more significantly in landslide and cut slope detection.

Spatial and texture features are also important for landslide detection. Area, length/width, compactness, shape index, rectangular fit, and Gray-Level Co-Occurrence Matrix

(GLCM) mean are the spatial and texture features used in this study. Figure 2.21 shows the extracted values of these features for landslide and non-landslide features.

The chart of spatial parameters reveals that landslide and cut slope can be well separated using these parameters than other previously discussed parameters. The values of most of the parameters for landslide, cut slope, and other features are different. Minimum, maximum, and average values vary for the three classes. Area, length/width, and texture are the most important parameters for the separation of landslide and cut slope features in the study area.

**Fig. 2.21** Typical values of landslide and non-landslide features. *Note* (1) area  $\times 103$ ; (2) GLCM mean  $\times 102$



## 2.6 Conclusion

Landslides are one of the most destructive natural disasters that mountainous areas, such as Malaysia, suffer from and are known to threaten human lives and properties. Landslide inventory maps are valuable sources of information and are essential for various studies and investigations, such as landslide susceptibility, and hazard and risk assessments, as well as various decision-making processes and policies. Providing an efficient method in detecting and distinguishing landslides and man-made slopes is a challenging task; most methods concentrate on landslide detection only, and these methods require much time and are costly.

This study aims to provide a rapid and accurate method that can create a landslide and man-made slope inventory map semiautomatically. To achieve this goal, few steps were applied: (1) the preparation and analysis of several spatial, spectral, and texture features, and LiDAR-derived parameters; and (2) the evaluation of five well-known classifiers to determine the best algorithm for landslide and man-made slope detection.

In general, five classifiers, i.e., Bayes, DT,  $k$ -NN, RF, and SVM, were evaluated to propose a semiautomatic supervised landslide and man-made slope detection approach using airborne LiDAR data coupled with orthophotos.

The research findings provided an effective solution for supervised and semiautomatic landslide and man-made mapping in tropical areas. Analysis showed that RF and SVM are the suitable classifiers for object classification using LiDAR data. The accuracies of these models were consistent in the three subsets of data that were used for validations. Overall evaluations of the studied classifiers showed that using supervised classification at object level,

separation between cut slopes and landslides using LiDAR data and orthophotos is possible and can be improved.

## References

- Abe, S. (2005). *Support vector machines for pattern classification* (Vol. 53). Berlin: Springer.
- Akçay, H. G., & Aksoy, S. (2008). Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7), 2097–2111.
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3), 239–258.
- Biguel, F., & Snelling, G. (1977). The geochronology of the main range Batholith: Cameron highlands road and Gunong Bujang Melaka. *Overseas Geol Miner Resour*, 47, 3–35.
- Bottou, L., & Lin, C.-J. (2007). Support vector machine solvers. *Large Scale Kernel Machines*, 301–320.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Breiman, L. (2003). RF/tools: A class of two-eyed algorithms. *Paper presented at the SIAM workshop*.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Boca Raton: CRC Press.
- Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Paper presented at the proceedings of the 23rd international conference on machine learning*.
- Chen, L. C., Teo, T.-A., Shao, Y.-C., Lai, Y.-C., & Rau, J.-Y. (2004). Fusion of LiDAR data and optical imagery for building modeling. *International Archives of Photogrammetry and Remote Sensing*, 35 (B4), 732–737.
- Chen, W., Li, X., Wang, Y., Chen, G., & Liu, S. (2014). Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the Three Gorges, China. *Remote Sensing of Environment*, 152, 291–301.
- Chow, W., Zakaria, M., Ferdaus, A., & Nurzaidi, A. (2003). Geological terrain mapping. JMG unpublished report. JMG. SWP. GS, 16, 1–42.

- Cobbing, E., Pitfield, P., Darbyshire, D., & Mallick, D. (1992). The granites of the SE Asian tin belt. British Geological Survey, Overseas Memoir No. 10: HMSO, London.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Definiens, A. (2007). *Definiens developer 7 reference book* (pp. 21–24). München: Definiens AG.
- Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *International Journal of Remote Sensing*, 33(14), 4502–4526.
- Eeckhaut, M., Poesen, J., Verstraeten, G., Vanacker, V., Nyssen, J., Moeyersons, J., et al. (2007). Use of LiDAR-derived images for mapping old landslides under forest. *Earth Surface Processes and Landforms*, 32(5), 754–769.
- Fang, H.-T., & Huang, D.-S. (2004). Noise reduction in LiDAR signal based on discrete wavelet transform. *Optics Communications*, 233(1), 67–76.
- Foody, G. M. (2004). Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, 70(5), 627–633.
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2–3), 131–163.
- Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., & Reichenbach, P. (2008). Comparing landslide inventory maps. *Geomorphology*, 94(3), 268–289.
- Gibril, M. B. A., Bakar, S. A., Yao, K., Idrees, M. O., & Pradhan, B. (2016). Fusion of RADARSAT-2 and multispectral optical remote sensing data for LULC extraction in a tropical agricultural area. *Geocarto International*, 1–14.
- Gorum, T., Fan, X., van Westen, C. J., Huang, R. Q., Xu, Q., Tang, C., et al. (2011). Distribution pattern of earthquake-induced landslides triggered by the 12 May 2008 Wenchuan earthquake. *Geomorphology*, 133(3), 152–167.
- Grimm, R., Behrens, T., Märker, M., & Elsenbeer, H. (2008). Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using Random Forests analysis. *Geoderma*, 146(1), 102–113.
- Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1), 56–66.
- Hodgson, M. E., Jensen, J., Raber, G., Tullis, J., Davis, B. A., Thompson, G., et al. (2005). An evaluation of lidar-derived elevation and terrain slope in leaf-off conditions. *Photogrammetric Engineering & Remote Sensing*, 71(7), 817–823.
- Laliberte, A. S., Rango, A., Havstad, K. M., Paris, J. F., Beck, R. F., McNeely, R., et al. (2004). Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of Environment*, 93(1), 198–210.
- Last, M., Maimon, O., & Minkov, E. (2002). Improving stability of decision trees. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(02), 145–159.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18–22.
- Lillesand, T. M., Kiefer, R. W., & Chipman, J. (2004). *Remote sensing and image interpretation*. New York: Wiley.
- Long, N. T. (2008). *Landslide susceptibility mapping of the mountainous area in A Luoi district, Thua Thien Hue province, Vietnam*. Faculty of Engineering, Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Belgium.
- Martha, T. R. (2011). *Detection of landslides by object oriented image analysis*. University of Twente, Faculty of Geo-Information Science and Earth Observation. Enschede, The Netherlands: ITC Printing Department.
- Martha, T. R., Kerle, N., Van Westen, C. J., Jetten, V., & Kumar, K. V. (2011). Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12), 4928–4943.
- McKean, J., & Roering, J. (2004). Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology*, 57(3), 331–351.
- Mitchell, T. M. (1997). *Machine learning* (Vol. 45, p. 37). Burr Ridge, IL: McGraw Hill.
- Möller, M., Lymburner, L., & Volk, M. (2007). The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation*, 9(3), 311–321.
- Mutanga, O., Adam, E., & Cho, M. A. (2012). High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 18, 399–406.
- Navulur, K. (2006). *Multispectral image analysis using the object-oriented paradigm*. Boca Rotan: CRC Press.
- Ohlmacher, G. C. (2007). Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Engineering Geology*, 91(2), 117–134.
- Olaya, V. (2009). Basic land-surface parameters. *Developments in Soil Science*, 33, 141–169.
- Pradhan, B., & Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling and Software*, 25(6), 747–759.
- Puissant, A., Rougier, S., & Stumpf, A. (2014). Object-oriented mapping of urban trees using Random Forest classifiers. *International Journal of Applied Earth Observation and Geoinformation*, 26, 235–245.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93–104.
- Samui, P. (2008). Slope stability analysis: A support vector machine approach. *Environmental Geology*, 56(2), 255–267.
- Soria, D., Garibaldi, J. M., Ambrogi, F., Biganzoli, E. M., & Ellis, I. O. (2011). A ‘non-parametric’ version of the naive Bayes classifier. *Knowledge-Based Systems*, 24(6), 775–784.
- Team, R. C. (2013). *R: A language and environment for statistical computing*.
- Vapnik, V. N., & Vapnik, V. (1998). *Statistical learning theory* (Vol. 1). New York: Wiley.
- Varmuza, K., & Filzmoser, P. (2016). *Introduction to multivariate statistical analysis in chemometrics*. Boca Rotan: CRC Press.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., et al. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37.
- Xie, Z., Zhang, Q., Hsu, W., & Lee, M. L. (2005). Enhancing SNNB with local accuracy estimation and ensemble techniques. *Paper presented at the international conference on database systems for advanced applications*.
- Zêzere s, J. L., de Brum Ferreira, A., & Rodrigues, M Ls. (1999). The role of conditioning and triggering factors in the occurrence of landslides: A case study in the area north of Lisbon (Portugal). *Geomorphology*, 30(1), 133–146.
- Zhang, H. (2004). The optimality of naive Bayes. *AA, I*(2), 3.

Laser Scanning Applications in Landslide Assessment

Pradhan, B. (Ed.)

2017, XVIII, 359 p. 197 illus., 186 illus. in color.,

Hardcover

ISBN: 978-3-319-55341-2