Doctoral Dissertation
博士論文

Evaluation of landslide inventory, classification and susceptibility assessment based on case studies in Japan and China

(斜面崩壊の分布図作成、分類、発生しやすさの評価ー日本と中国を例に)

竇　傑

(Dou Jie)

The University of Tokyo

東京大学

ABSTRACT

Landslides are one of the most widespread geological hazards affecting mountainous regions all over the world. In recent years, extreme weather has brought about many weather-related disasters worldwide, causing significant casualties and economic losses. In hilly terrains such as in densely populated regions of China and Japan, there are always the risks of landslides that cause multiple fatalities and serious consequences by extreme rainfall or serious tremors. Concerted efforts of the government and the general public are crucial in enhancing the community's resilience against landslide disasters and reducing the potential loss of life and damage to properties. Hence, expeditious construction of landslide inventory maps and prediction of landslide occurrence have become an important but challenging issue.

This research mainly presents the results of testing the strengths and weaknesses of contemporary landslide inventories and susceptibility mapping techniques, to facilitate the construction of accurate regional landslide susceptibility maps for varying geo-environments, through case studies in Japan and China. Three main topics were selected based on extensive reviews of literature and considered as the most important to improve landslide susceptibility mapping. The details of three specific tasks are listed as below:

1) Landslide inventories are often prepared by manual analysis of post-event aerial photographs or satellite images. This is time-consuming and may lead to misinterpretations. This work presents an improved automated model for rapid preparation of landslide inventories. The experimental results indicated that the proposed integrated method demonstrates higher classification performance than the stand-alone object oriented image analysis (OOIA) technique for detecting landslides. The area under curve (*AUC*) of the receiver operating characteristics (ROC) 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.

2) Many previous studies successfully evaluated susceptibility of landslides in a wider area, however prediction of landslide types in to deep and shallow slides which are crucial for risk analysis has rarely been conducted. This work examines the differences in landslide depth, volume and the risk imposed between shallow and deep-seated landslide types. Shallow and deep-seated landslide prediction is useful in utilizing emergency resources by prioritizing target areas while responding to sediment related disasters. Ten factors, including elevation, slope, aspect, curvature, lithology, distance from the nearest geologic boundary, density of geologic boundaries, distance from drainage network, the compound topographic index (*CTI*) and the stream power index (*SPI*) derived from the DEM and a geological map were analyzed using support vector machine (SVM) technique. Iterated over 10 random instances the average training and testing accuracy of landslide type prediction was found to be 89.2% and 77.8%, respectively. The overall accuracy of SVM does not rapidly decrease with a decrease in training samples. The trained model was then used to prepare a map showing probable future landslides differentiated into shallow and deep-seated landslides.

3) Different studies use different numbers of causative factors for the development of susceptibility maps. The selection of the causative factors so far largely remains random and subjective. Selection of essential factors improves the prediction accuracy of landslide susceptibility mapping (LSM). This work proposes a rule-based statistical method for an objective selection of causative factors fitting to differently triggered landslides. The certainty factor (CF) model was then applied to select the best subset from the original available factors. Using all factors and the best subset factors obtained, landslide

susceptibility maps were produced using statistical index (SI) and logistic regression (LR) models. The susceptibility maps were validated and compared using landslide locations in the validation data. The prediction performance of two susceptibility maps was estimated using *AUC*. The result shows that *AUC* values for the LR model (0.817 for Niigata and 0.837 for Dongjiang) are slightly higher than those obtained from the statistical index (SI) model (0.801 for Niigata and 0.794 for Dongjiang). Our findings can help to understand the main causative factors with landslide occurrence.

These spatial-temporal aggregation and scenario models for detecting and evaluating landslide susceptibility, which could be adopted as a prototype for warning systems in the other similar landslide-prone areas. Additionally, the optimization of causative factors such as slope angle, and lithology can be used in other susceptible regions, especially for data scarcity areas. Moreover, the susceptibility maps could assist urban planners, designers, civil engineers and earth scientists to specify where a problem may exist and to determine what type of failure may occur at the hazardous regions in the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS BY ALPHABET

Acronyms	**Explanation**
ANN	Artificial neural network
AUC	Area under curve
BP	Back propagation
BPNN	Back propagation neural network
CBR	Case-based reasoning
CF	Certainty factor
CTI	Compound topographic index
DEM	Digital elevation model
Dev	Deviation
DL	Deep learning
DTC	Decision tree classification
EKS	Expert-based knowledge systems
ESP	Estimation of Scale Parameter
FNR	False negative ratio
FoS	Factor of safety
FPR	False positive ratio
FR	Frequency ratio
GA	Genetic algorithm
GCPs	Ground Control Points
GIS	Geographic information system
GLCM	Grey level co-occurrence matrix
GLDV	Grey level difference vector

Acronyms	**Explanation**
GPS	Global Positioning System
GRA	Gray relational analysis
GRNN	General regression neural networks
GSI	Geospatial Information Authority of Japan
GSJ	Geological Survey of Japan
H	Hyperplane
IHRR	Institute of Hazard, Risk and Resilience
ILR	Initial learning rate
JAXA	Japan Aerospace Exploration Agency
JMA	Japan Meteorological Agency
LF	Linear functions
Lidar	Light Detect ion and Ranging
LR	Logistic regression
LSI	Landslide susceptibility index
LSM	Landslide susceptibility mapping
Meter	m
Magnitude	M
MLC	Maximum likelihood classification
NDVI	Normalized Difference Vegetation Index
NH	National Highway
NIED	National Research Institute for Earth Science and Disaster Prevention
NIR	Near-infrared

Acronyms	**Explanation**
OLS	Ordinary least squares regression
OOIA	Object-oriented image analysis
PF	Polynomial functions
PGA	Peak ground acceleration
PNN	probabilistic neural networks
PRD	Pearl River Delta
4R	Retrieved, reused, revised, and retained
RBF	Radial basis functions
RF	Random forest
RG	Related grade
RMSE	Root Mean Squared Error
ROC	Receiver operating curve
RS	Remote sensing
SCD	Similarity correlation degree
SF	Sigmoid (SF) functions
SI	Statistical index
SPI	Stream power index
SVM	Support vector machines
TNR	True negative ratio
TOL	Tolerance
TPR	True positive ratio
VIF	Variance inflation factor
WOE	Weight of evidence

CHAPTER 1 INTRODUCTION

1.1 Literature review

Landslides are the result of natural geologic processes that have worked to shape the landscape, which are one of the most widespread geological hazards affecting the mountainous regions all over the world. Landslides are induced by earthquakes, rainfall, snow melt and human interventions, resulting in significant casualties and property damage every year around the world. The annual losses due to landslides are more than those of any other types of natural disasters such as earthquakes, floods, and sinkhole formation (Guzzetti 1999; García-Rodríguez et al. 2008). In recent years, especially the extreme weather has brought about many more weather-related disasters worldwide, including the landslide hazards which cause significant casualties and economic losses. In hilly terrains such as in densely populated regions of China and Japan, there are always the risks of landslides that cause multiple fatalities and serious consequences by extreme high-intensity rainfall, adverse geological environment, anthropogenic activities or serious tremors. As indicated by Turner and Schuster (1996), this trend will continue and be clearer under the influence of urbanization, economic development, deforestation, and increased regional precipitation in landslide-prone areas due to changing climate as shown in Figure 1-1 (IHRR 2010). To mitigate from serious disasters, landslide susceptibility, hazard, and risk must be predicted (Guzzetti 1999; De Waele et al. 2011). Concerted efforts of the Government and the general public are crucial in enhancing the community’s resilience against landslide disasters and reducing the potential loss of life and damage to property. Therefore, fast constructing inventory map, classification, and susceptibility of landslide occurrence have become an important but challenging issue in the hazard mitigation research field (Chang and Chao 2006; Dou et al. 2014; Danneels et al. 2007; Lu et al. 2011).

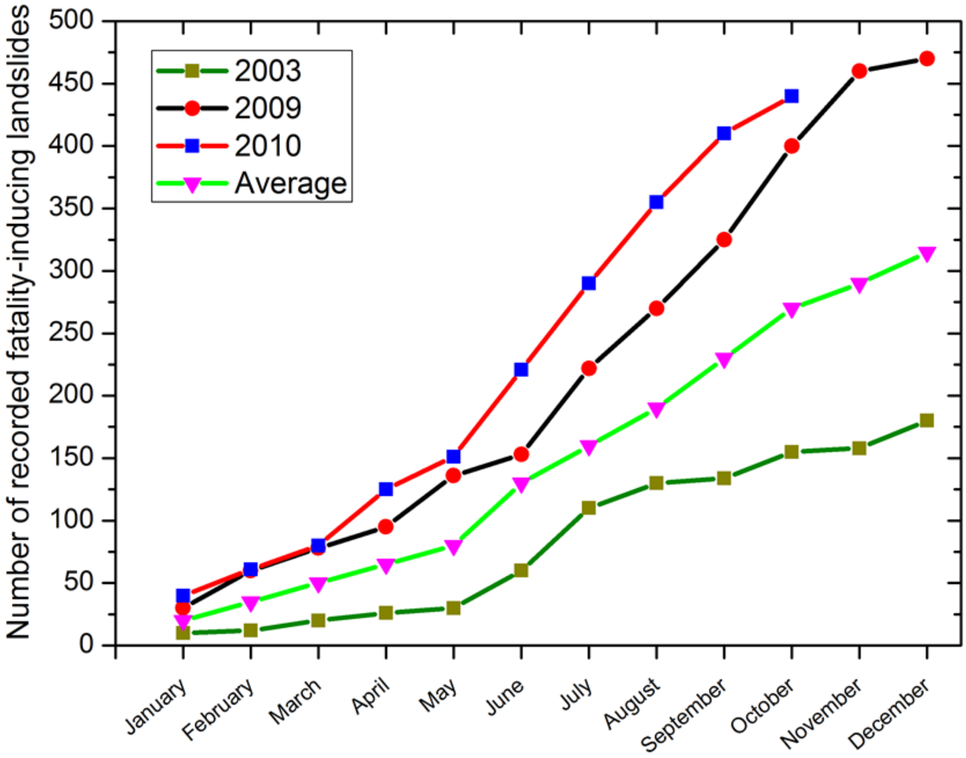

Figure 1-1 Number of fatality-inducing landslides in 2003, 2009 and 2010 compared with average revised from International Landslide Centre part of Institute of Hazard, Risk and Resilience (IHRR), 2011.

For a successful qualitative or quantitative landslide hazard evaluation, compiling a historical landslide-event inventory is particularly crucial for pre-disaster and post-disaster analyses (Guzzetti et al. 2005; Lu et al. 2011). 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 (McKean and Roering 2004a). Herein, to address the above mentioned problems, this work proposes a relatively new approach for semi-automatically 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).

Additionally, both shallow failures and erosion into bedrock play important roles in shaping landscapes in mountainous areas (Oguchi 1996). However, previous studies tended to focus on the spatial prediction of only a single type of landslides (Chang and Chao 2006; Lee and Tsai 2008; Cheng et al. 2010). Therefore, few studies differentiated the probabilities of shallow and deep-seated landslides. Differentiating the two landslide types is conducive to assessing the geomorphic hazards contributing to the hillslope erosion and sediment discharge for the protection of human settlements and infrastructures (Dramis and Sorriso-Valvo 1994; Korup 2005, 2006; Tsai and Chen 2009; Larsen et al. 2010). In this study, we propose a nonlinear algorithm of support vector machine (SVM) to classify the landslide types.

Furthermore, Landslide susceptibility maps play a vital role in assisting and managing hazards for land use planning and risk mitigation (Yalcin et al. 2011; Tofani et al. 2014; Shahabi et al. 2014; Dou et al. 2015a; Dragićević et al. 2015). These maps provide information on the likelihood of landslides occurring in an area given the local terrain conditions (Brabb 1984). This involves not only finding where the risk of landslide-related problems is spatially located, but also qualitatively and quantitatively assessing the significance of any such hazards and associated risk factors. Using GIS (geographical information systems), various methods for landslide susceptibility mapping have been proposed, which can be divided into qualitative and quantitative methods (Felicísimo et al. 2012; Peng et al. 2014). Qualitative methods is the knowledge driven approaches that denote susceptibility levels in descriptive terms using expert knowledge (Conoscenti et al. 2014). Qualitative methods are relatively subjective and were extensively used during the 1970s and 1980s (Aleotti and Chowdhury 1999; Yilmaz et al. 2011). The main limitation of qualitative methods is that accuracy depends on the knowledge of the experts who carried out the research. Quantitative methods statistically and numerically investigate the relationship between landslides and causative factors to predict landslide occurrence probabilities (Neuhäuser et al. 2011; Anbazhagan and Ramesh 2014). Compared to qualitative methods, a more realistic susceptibility map can be obtained from statistical methods based on the data-driven approaches (Yalcin et al. 2011) because they reduce subjectivity and biases in the process of weighting landslide causal factors. This study will apply different methods of landside susceptibility mapping (LSM) and compare the results. For LSM, different scientists use different causative factors. Ayalew and Yamagishi (2005) reported that neither universal criteria nor guidelines are followed to select the landslide causative factors. Thus, it is difficult to determine whether the collected factors are appropriate or not. Moreover, factor redundancy and a method to quantify factors pose a challenge (van Westen et al., 1997). Hence, we propose a method to optimize the factors for landslide susceptibility assessment.

1.2 Research objectives

The overall objectives of this dissertation are to construct the landslide inventory map, classify the landslide types, and compare the robustness of different models for landslide susceptibility assessments based on case studies in Japan and China with varying geo-environments. The specific three objectives are as follows:

1) Propose an integrated model to rapidly map the spatial landslide distribution for the future landslide assessment in the one of fast economic developing area, south China. As noted, landslide inventories are often prepared by the manual analysis of post-event aerial and satellite images, which is time-consuming and may lead to misinterpretations. Our work includes an improvement to the method for preparing landslide inventories and

classification of their types.

2) Classify landslides into shallow and deep-seated based on existing inventories. As noted, this kind of work has been needed but rarely conducted.
3) Optimize landslide causative factors, apply different landslide susceptibility models, and compare their results. Selection of positive factors improves the prediction accuracy of an LSM. This work proposes a rule-based statistical method for an objective selection of causative factors fitting to differently triggered landslides.

1.3 Structure of the doctoral thesis

This thesis is organized in eight chapters. The following chapters are as follows:

Chapter 2 describes the four study regions in Japan and China.

Chapter 3 uses data for the four study areas including 1) image data (satellite image and aerial photographs); 2) topographic data (different resolution DEMs); 3) landslide data from NIED, and 4) others such as land use data from GSI (Geospatial Information Authority of Japan), geological maps (1:50000) from GSJ (Geological Survey of Japan), rainfall data from JMA (Japan Meteorological Agency), and field photographs.

Chapter 4 presents an integrated model to automatically detect landslides for constructing landslide inventories.

Chapter 5 provides a method to classify landslides into shallow and deep-seated ones using a machining learning method applied to the existing inventories.

Chapter 6 discusses the selection and significance of landslide causative factors through the application of different mapping techniques.

Chapter 7 evaluates the approaches and results obtained in the previous chapters. An overall discussion to summarize the work is also included.

Chapter 8 concludes the thesis with listing the advantages and limitations of landslide susceptibility assessment using remote sensing and GIS in varying geo-environments.

CHAPTER 2 STUDY AREAS

In this research, four typical landslide-prone mountainous areas in China and Japan are selected as case study sites. They are: 1) Conghua district, Pearl River Delta (PRD), south China (frequent rainfall and human intervention activity); 2) Chuetsu area, Niigata Prefecture, Japan (earthquake-triggered landslides); 3) Sado Island, Niigata Prefecture, Japan (snowmelt and rainfall triggered landslides); and 4) Dongjiang Reservoir watershed, Hunan Province, China (torrential rainfall and human intervention activity). These areas have been suffered from severe landslides which caused considerable damage to properties and loss of life. Therefore, it is necessary to understand the mechanism of landslides for quickly detecting and forecasting landslides, to alert the authorities and local people in these regions.

2.1. Conghua district, China

The Conghua district 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 2-1). It is a district of Guangzhou City, which connects the Pearl River Delta (PRD) with the mountainous area in the north of Guangdong Province. The elevation of the district ranges from 4 to 1185 m with a mean of 600 m. The district is known as "Guangzhou Garden" because of its vast forest coverage of approximately 67%, a total area of 1974.5 km^2, in the proximity of the Tropic of Cancer. The average temperature is approximately 21 ℃. Nearly half of the study area consists of mountains and highlands. In the past few years, the study area has experienced fast land-use/land-cover changes in response to the economic boom; 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 been occurring frequently (Figure 2-2), resulting in considerable economic losses from the road damage and destruction of fundamental infrastructure. In particular, a road section near National Highway 105 (NH 105) is frequently damaged by landslides, causing event injury to people and severe destruction of buildings and property. Principal landslide types observed in Conghua are shown in Figure 2-2. The majority of them are debris slides, soil creep, and rock falls, according to the classification by Varnes (1978). Evidence of previous landslides can be observed along NH 105 (Figure 2-2 d–f) including concrete structures for slope stabilization. Hence, understanding the spatial distribution of landslides and mitigating hazards are urgent in this region.

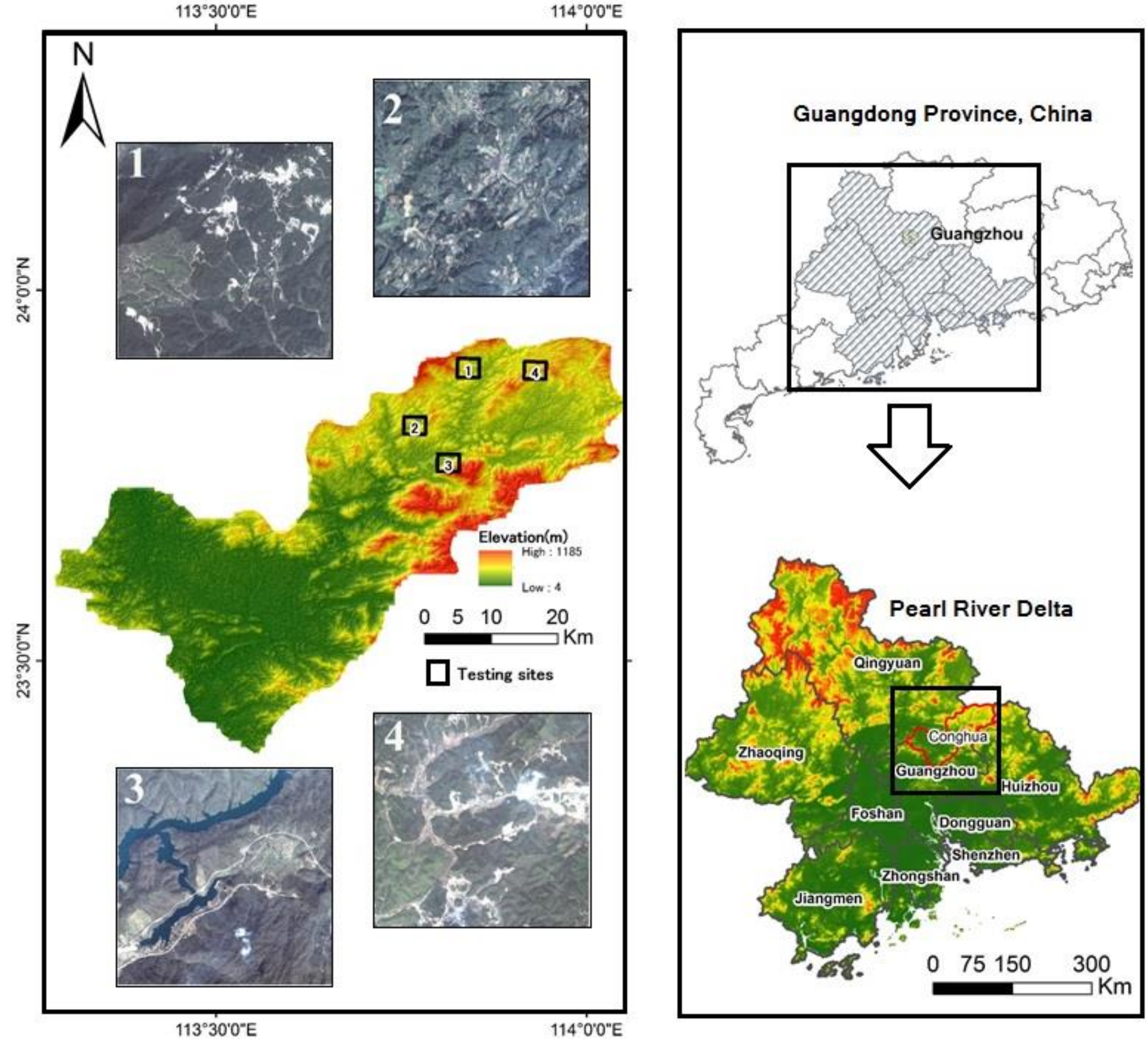

Figure 2-1 Location of the Conghua district with elevation distribution.

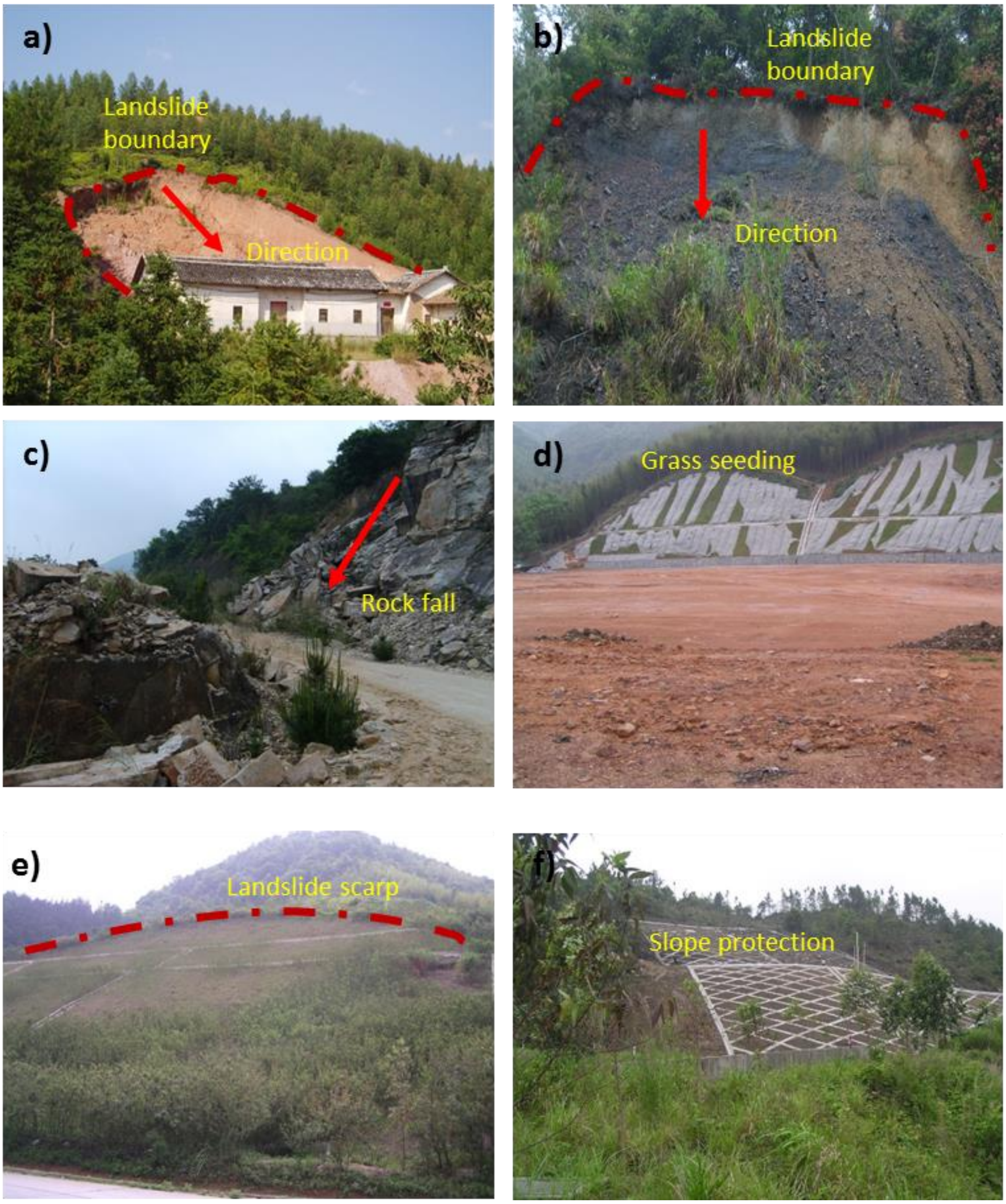

Figure 2-2 Examples of landslides in the Conghua district. (a) Debris slide threatening a house. (b) Soil creep. (c) Rock fall. Evidence of previous landslide scarps in the form of (d) terraced off and (e, f) concrete structures for slope stabilization.

2.2. Chuetsu area, Japan

The study area in the Chuetsu area is located in a mountainous region of Niigata Prefecture, Japan (Figure 2-3). The 300 km^2 study area is situated at 138°47'E – 138°58'E and 37°14'N – 37°22'N, where the elevation ranges from 22 to 734 m with a mean of 206 m. It annually receives approximately 2000 mm of precipitation, a few typhoons, and heavy snow. The area is underlain by sedimentary and metamorphic rocks from the Paleogene to Quaternary (Takeuchi and Yanagisawa 2004). Land use/cover in the area is characterized by sparse settlements, agro-industrial activities such as paddy farming, and deciduous broad-leaved beech forests.

The area experienced an earthquake of M 6.8 on October 23, 2004, with the depth of the hypocenter being 13 km (Japan Meteorological Agency, 2004; Figure 2-4). The ground shaking caused by the earthquake and numerous aftershocks (four $\geq M$ 6 and ten M 5–6) triggered numerous shallow and deep-seated landslides (Figure 2-5), in the Hagashiyama Mountains in the Chuetsu area. The Fire and Disaster Management Agency of Japan (2004) reported that 40 people died and 4,496 were injured during this event. The damage of property and infrastructure from the landslides alone had been initially estimated at U.S. $8 billion, making it one of the costliest landslide events in the history (Kieffer et al. 2006). Therefore, it is necessary to assess this event in detail. We analyzed an airborne LiDAR digital elevation model (DEM), and collected 13 possible landslide-conditioning factors: elevation, slope angle, slope aspect, total curvature, plan curvature, profile curvature, drainage density, distance from drainage networks, the compound topographic index (*CTI*), the stream power index (*SPI*), lithology, density of geological boundary, and distance from geological boundary.

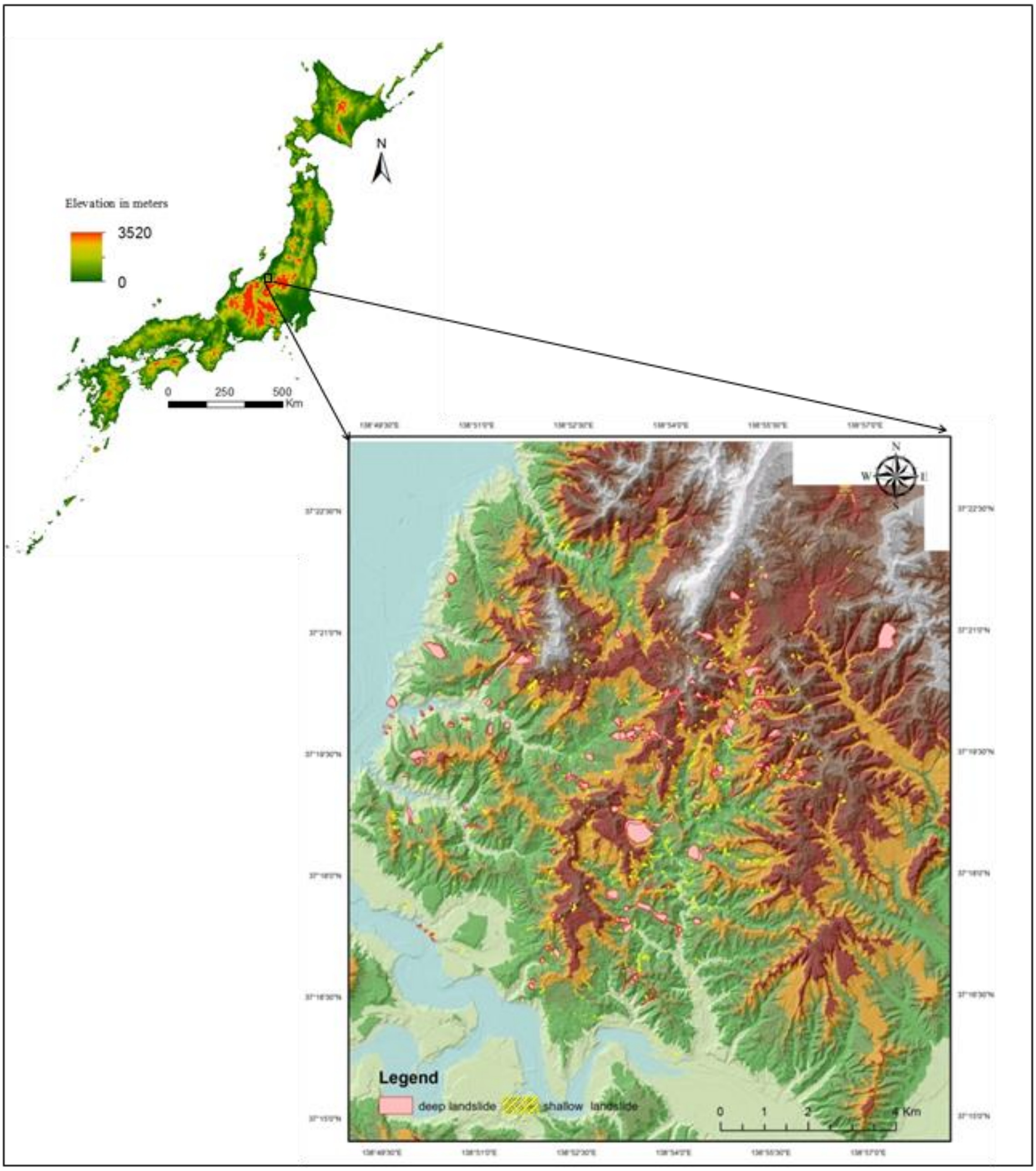

Figure 2-3 Location of the study area in Chuetsu and distribution of landslides. Landslide data are from the National Research Institute for Earth Science and Disaster Prevention (NIED), Japan.

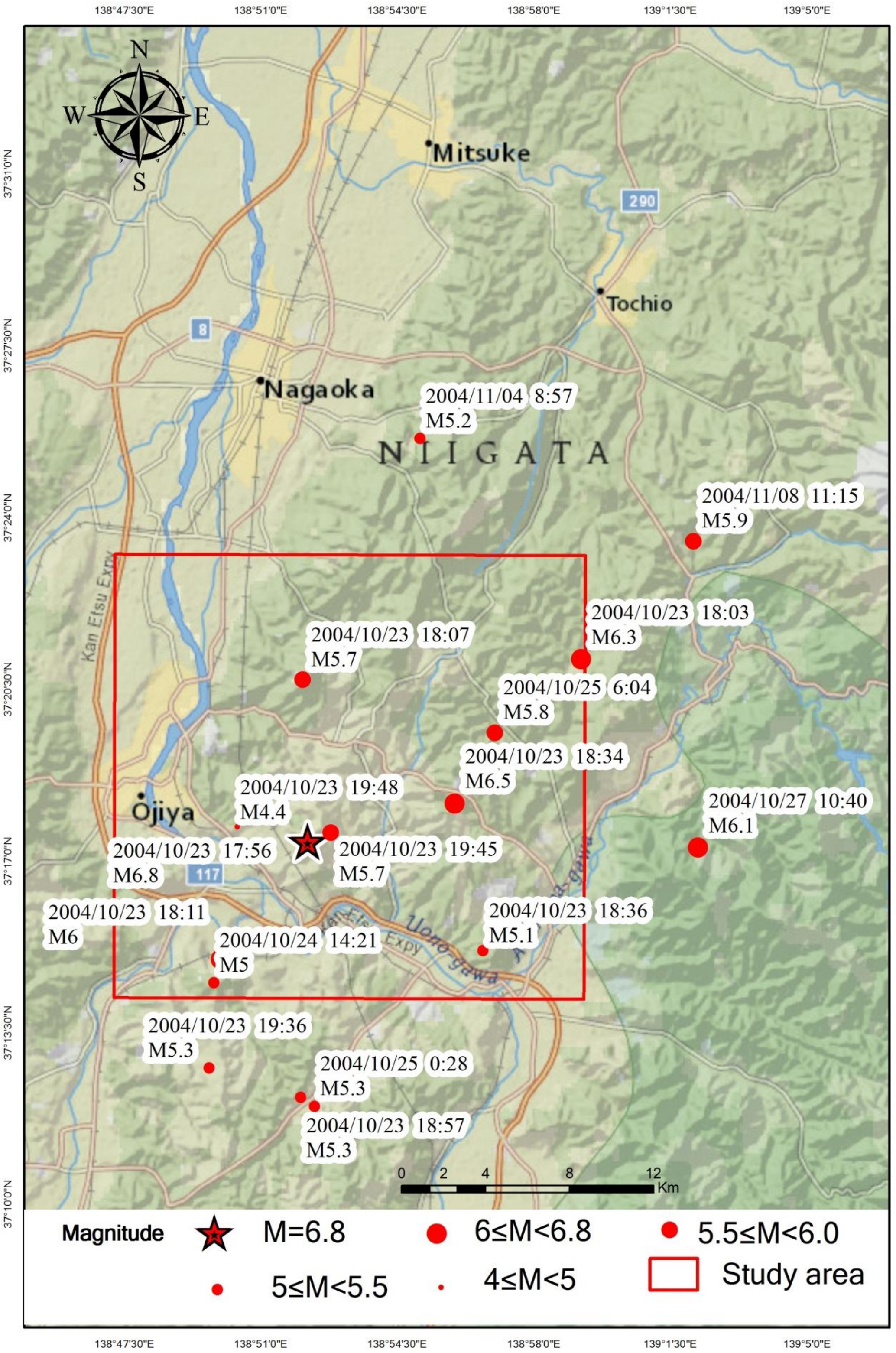

Figure 2-4 Epicentral distribution of major earthquakes (23 Oct – 10 Nov, 2004) in and around the mid-Niigata prefecture region. Map prepared using the data from Japan Meteorological Society, 2004.

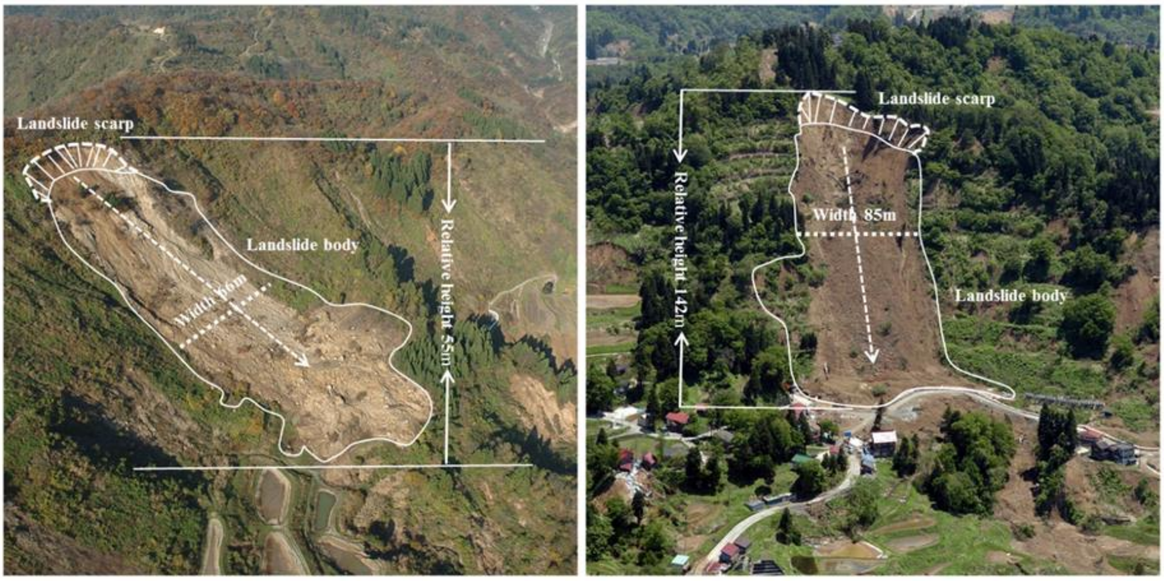

Figure 2-5 A shallow landslide (left) in the southwest of Nigorizawa, Nagaoka City; and a deep-seated (right) landslide in the western entrance of the Haguro tunnel, Nagaoka City. Width and relative height for the deep-seated landslide is larger than those of the shallow landslides. (Images provided by NIED).

2.3. Sado Island, Japan

The study area in Sado Island (Figure 2-6) is located in Niigata Prefecture, Japan, between longitudes 138 °14' - 138 °32' E, and latitudes 37 °57' - 38 °20' N. It covers an area of nearly 400 km2. The elevation varies from 0 to 1172 m with a mean of 333 m. The peak of the island is the Mt. Kimpoku in the Osado Mountains. The geology is composed of Neogene and marine volcanics, such as dacitic and andesitic sediments, associated with pyroclastics and rhyolitic intrusives in green tuffs. Some coastal slopes involve lately formed semi-consolidated and unconsolidated sand deposits and gravel. This area is highly prone to landslides and subjected to tectonic movements that are evidenced by thrust up benches and active faults. In the study area, the landslide types are mostly deep-seated, translational and rotational slides (Figure 2-7). Ayalew et al. (2005) noted that rock falls seldom occurred in this region. Most of the deep-seated landslides are inventoried in the rhyolitic and dacitic lithologies. The landslide susceptibility may also be influenced by rainfall, snow melting, geology, slope aspect, and slope angle.

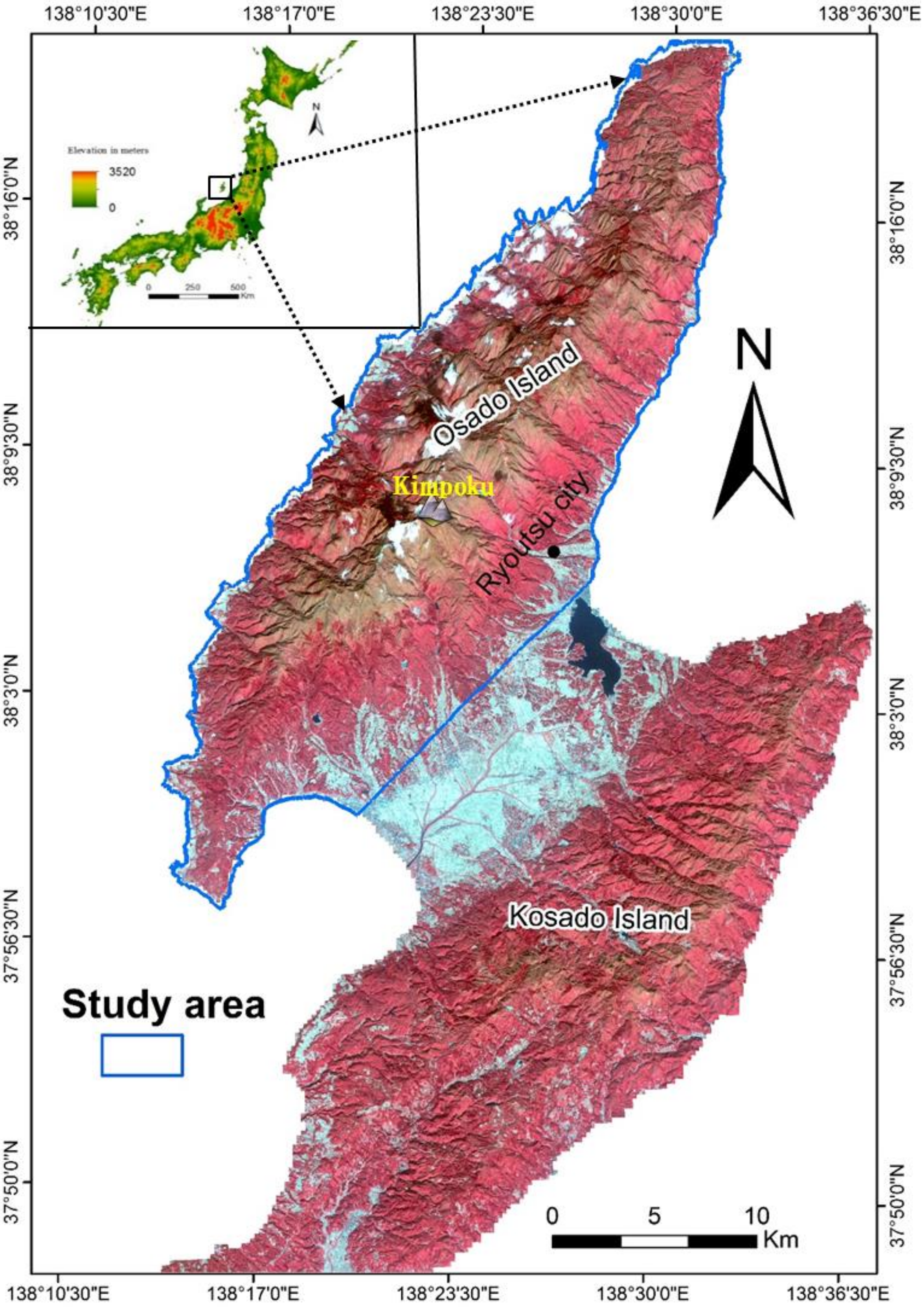

Figure 2-6 Satellite image of the study area. The three bands (Near-IR, red and green) false composite ALOS image was provided by The Japan Aerospace Exploration Agency (JAXA). Vegetation in a red color reflects high reflectance for the Near-IR.

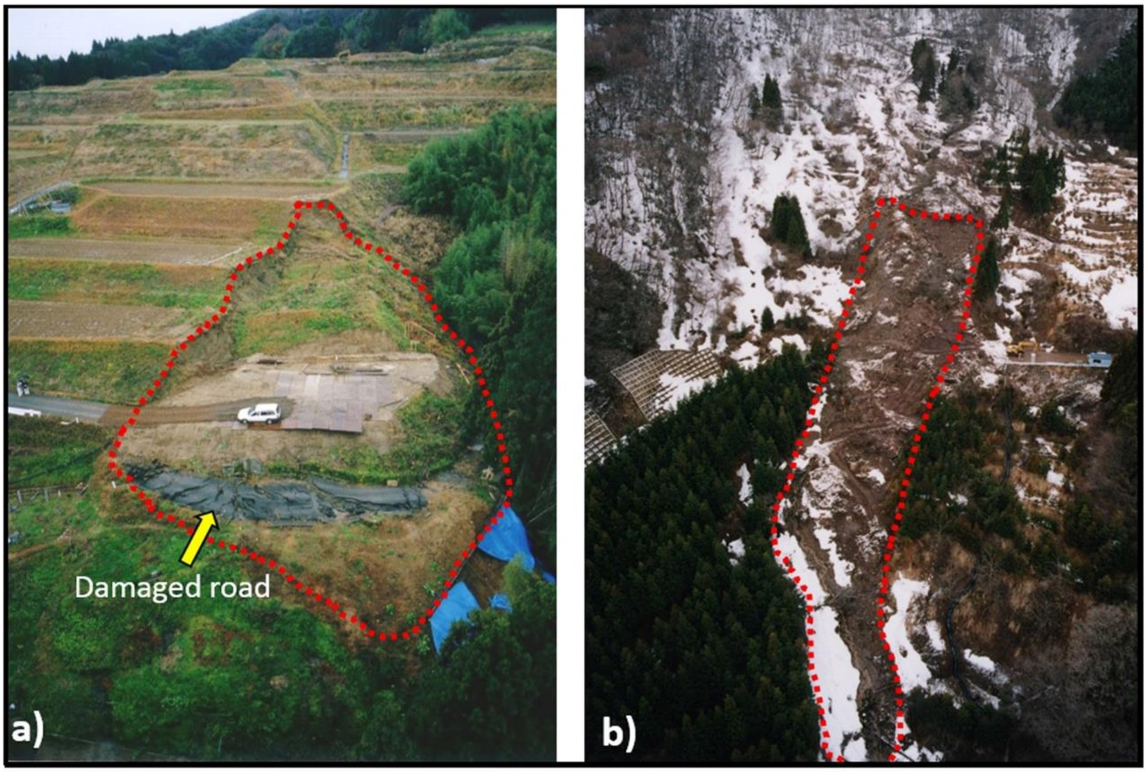

Figure 2-7 Examples of identified major landslide types in Sado Island. a) Rotational slide that severely damaged a road. b) Translational slide in the hilly terrains that carved the dense vegetation.

2.4. Dongjiang Reservoir, China

The Dongjiang Reservoir study area located in the southeast of Hunan Province, China (Figure 2-8 a and b). The Dongjiang Reservoir is susceptible to heavy precipitation during tropical cyclone seasons. The elevation of the area ranges from 78 to 1868 m with a mean of 540 m. It is mainly composed of three distinct geomorphological units: hilly plains, hilly valleys and the Luoxiao Mountains near the eastern and southern boundaries. The area is geologically composed mostly of Paleozoic sedimentary and metamorphic rocks such as sandstone, sandy slate, and limestone, which were intruded by granitic rocks in places. The granitic rocks are deeply weathered. The weathered soils are mostly composed of highly oxidized laterite, prone to erosion. Land use/cover in the study area is characterized by settlements, small-scale agro-industrial activities such as paddy farming, and plantation. The study area belongs to the humid sub-tropical monsoon climate. The annual precipitation is 1,538 mm, mostly affected by typhoons.

The Dongjiang Reservoir is the largest reservoir in the south of Hunan Province, covering the water area of 160 km^2 and has the capacity of 8.12×10^9 m^3. Due to the heavy rainfall caused by the Typhoon BILIS in 2006, thousands of sediment-related disasters, including numerous slope failures (shallow landslides) and debris flows occurred, and were inventoried through the Quickbird images (0.6 m in resolution), CBERS images (20 m) and field work (Figure 2-9). The torrential precipitation event associated with Typhoon BILIS caused 246 deaths, 95 missing and more than 300 million US dollars of economical loss just in and around Zixing City. Damages of destroyed or buried buildings by debris flows were serious. There were also considerable slope failure disasters where precipitation was intense. Flash floods also inundated short and steep rivers in the hilly areas.

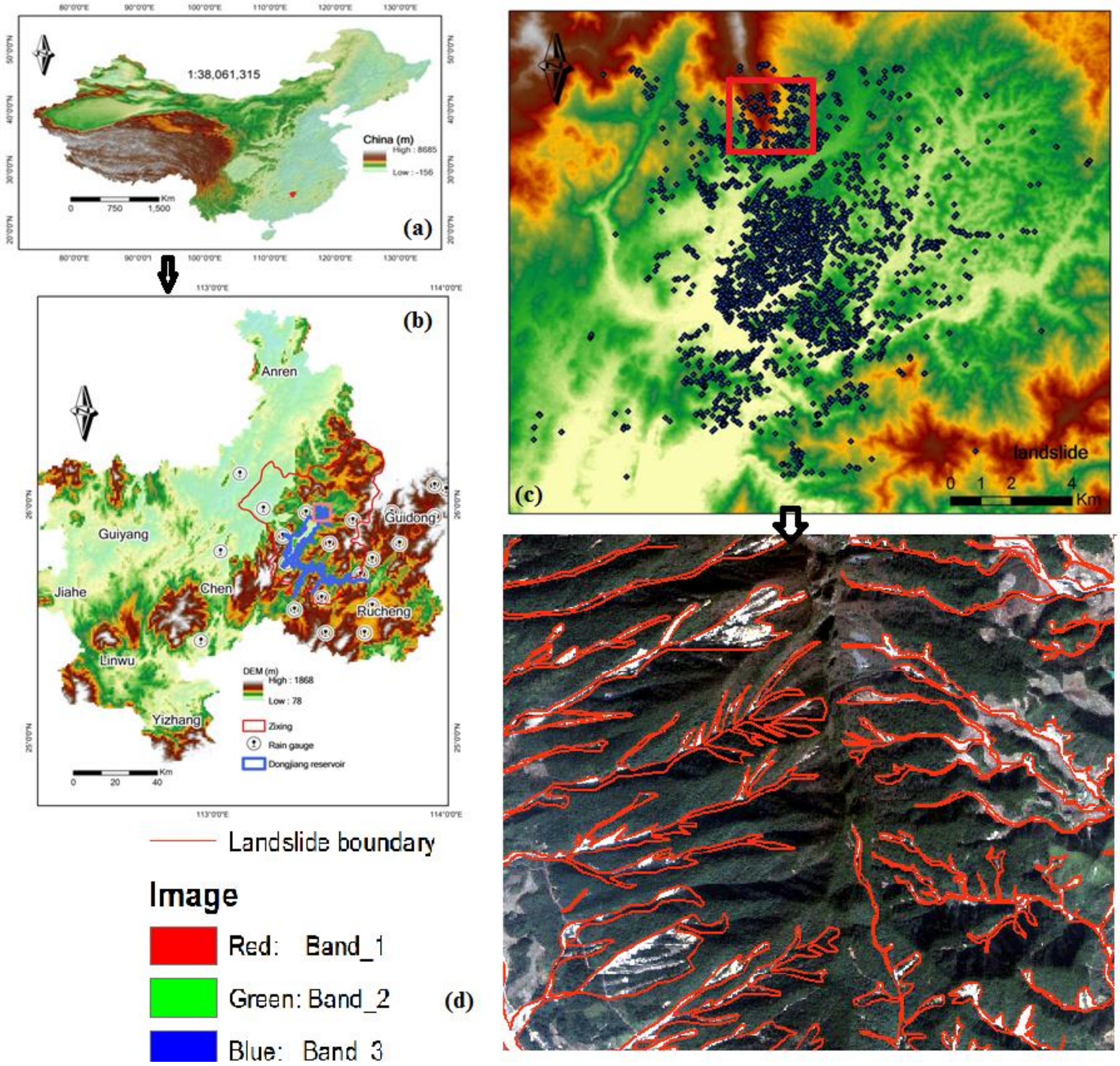

Figure 2-8 Dongjiang Reservoir study area. a) Location map of China. b) Map of the study area with rain gauge distribution. c) Distribution of shallow landslides on the elevation map derived from a 30 m DEM. d) The lower map is the enlarged area of showing the landslide boundary.

Figure 2-9 Rainfall-induced landslides by the typhoon BILIS. (a) Example of shallow landslide (dots on the scar indicate vector points) and associated with debris flows. (b) Landslide scar and threatened property. (c) Many shallow landslides (black arrows).

CHAPTER 3 DATA

3.1. Data for the Conghua district, China

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 a 1 m DEM were used as ancillary data. Table 3-1 lists the specifications and details of the satellite images used. We also conducted field surveys in the same season of the year as the image acquisition periods to support and validate our results.

Table 3-1 Spectral and spatial resolution of QuickBird and SPOT5 satellite image data.

Satellite	Pan (μm)	Red (μm)	Green (μm)	Blue (μm)	Near IR (μm)	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

3.2. Chuetsu area, Japan

Landslides in the past are keys to predict the distribution of future landslides (Guzzetti et al. 1999). We used a landslide inventory prepared by NIED, Japan, based on the interpretation of aerial photographs. The inventory contains 895 shallow landslides and 330 deep-seated landslides, with an average area of 187 and 9,600 m^2 respectively (Figure 3-1).

Their minimum areas are 42 and 271 m^2, and maximum areas are 28,178 and 205,461 m^2, respectively. The inventory contains landslides represented by polygons; however, for this study, the landslide polygons were changed into points each at the centroid of each landslide polygon, using ArcGIS 10.1, a GIS package from ESRI.

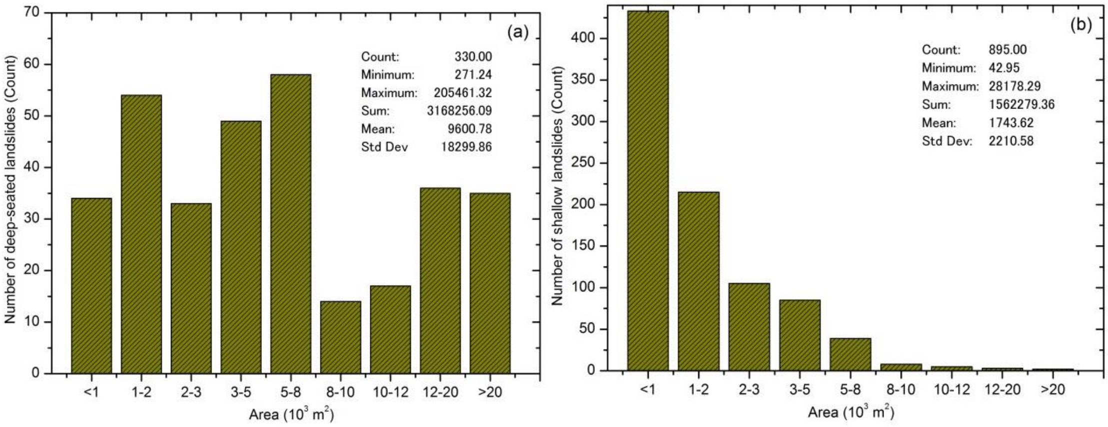

Figure 3-1 Histograms showing characteristics of landslide types: (a) area of deep-seated landslides, (b) area of shallow landslides.

Landslides can be classified into deep and shallow in relation to material and movement mechanism (Dai et al. 2011) (Figure 3-2). NIED used the depth of the sliding plane to differentiate the two types: depth < 10 m = shallow, and else deep-seated. This scheme of landslide differentiation is also used by Roering et al. (2003). Shallow landslides with the movement of the surface soil mantle are smaller in volume than the deep-seated landslides with the movement of both surface mantle and underlying weathered bedrock. Deep-seated landslides more likely cause large scale debris flows and landslide dams, with more disastrous consequences.

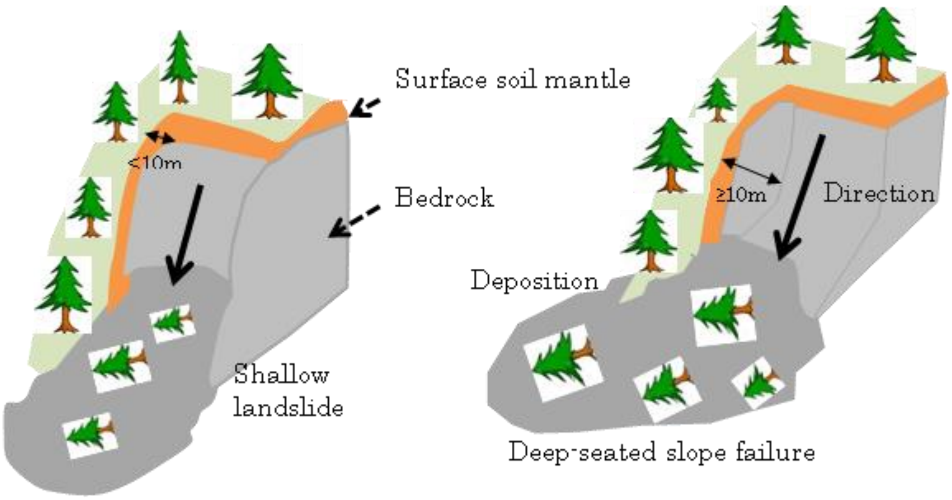

Figure 3-2 Sketch of shallow landslides (left) and deep-seated landslides (right).

The thematic data used are summarized in Table 3-2. An airborne Lidar DEM with a spatial resolution of 2 m was provided by GSI in 2005 (Geographical Survey Institute 2007). The geological information used is based on the geological maps provided by GSJ (Takeuchi and Yanagisawa, 2004). The lithology is mainly classified into three groups: (1) gravel, andesite, dacite lava, andesitic pyroclastic rock, tuffaceous sandstone, and rhyolite tuff; (2) gravel, sand, silt, and mudstone; (3) sand, silt, sandstone, and massive mudstone. All conditioning factors used are continuous except the categorical lithology data.

Table 3-2 Thematic datasets used in the study.

Classification	Sub-classification	GIS data type	Scale or resolution	Classes	Source of data
Landslide inventory map	Landslide	Polygon coverage	1:50,000	Continuous	NIED
Geological map	Lithology Geological boundary	Polygon coverage Line coverage	1:50,000	Categorical Continuous	GSJ
Topographic map	DEM	ARC/INFO Grid	2×2 m	Continuous	GSI

3.3. Sado Island, Japan

According to Guzzetti et al. (1999), landslides which occurred in the past and present are keys to predict future landslides. The first step is to compile a landslide inventory map. The details of the data for Sado Island are itemized in Table 3-3. In this study, a total of 825 known landslides (Figure 3-3) were first obtained for the model development; these landslides were interpreted by the landslide experts at NIED. NIED has been producing this landslide inventory since the year 2000 from the repeated acquisition of multiple aerial photographs. The landslides are depicted as boundary polygons in the GIS shape file format and are available at the NIED archives (http://lsweb1.ess.bosai.go.jp/gis-data/index.html). The archived landslide inventory data were also used to successfully produce landslide hazard map by a logistic regression model in other study regions (Wang et al. 2013). It is observed from the landslide inventory map that most landslide areas are greater than 0.01 km^2. The minimum area observed is 0.0006 km^2, whereas the largest landslide covers an area of about 1.65 km^2. The total area of landslides is about 57 km^2, and accounts for approximately 15% of the study area. The areas of all the mapped landslide display a frequency distribution that can be described by a power law for approximately three orders of area with a good fit ($R^2 = 0.95$) (Figure 3-4). In this study, the landslide distribution was examined only for medium to large landslides (2.34×10^3 to 8.48×10^5 m^2) with a significant rollover value (1.13×10^4 m^2). The power law can be used to examine the dominant landslide areas (Guzzetti et al. 2002). The exponent of the power law (1.6) indicates that the large landslide areas are dominant.

Different sampling strategies are available to construct the reliable landslide susceptibility maps. Several previous researches preferred to use 'points' to represent the spatial location of landslides (Neuhäuser et al. 2011; Tien Bui et al. 2012b). Dai and Lee

(2003) delineated only the source areas during the landslide susceptibility assessments and excluded both the transport and deposition of areas existing landslides. Few other studies preferred to use the landslide area with depletion and accumulation zones like “seed cells” to represent pre-failure conditions (Süzen and Doyuran 2004; Bai et al. 2010; Wang et al. 2013). “Seed cells” are the zones that are regarded to represent the undisturbed morphological condition (Süzen and Doyuran 2004). Comparisons of these sampling strategies are however beyond the scope of this study. Here, we adopted one of the most popular methods, the use of the polygon of each landslide to represent its location (Yalcin et al. 2011; Peng et al. 2014). For building models, the landslide inventory was randomly partitioned into two groups: a training dataset (70%, 578 landslides) and a validation dataset (30%, 247 landslides).

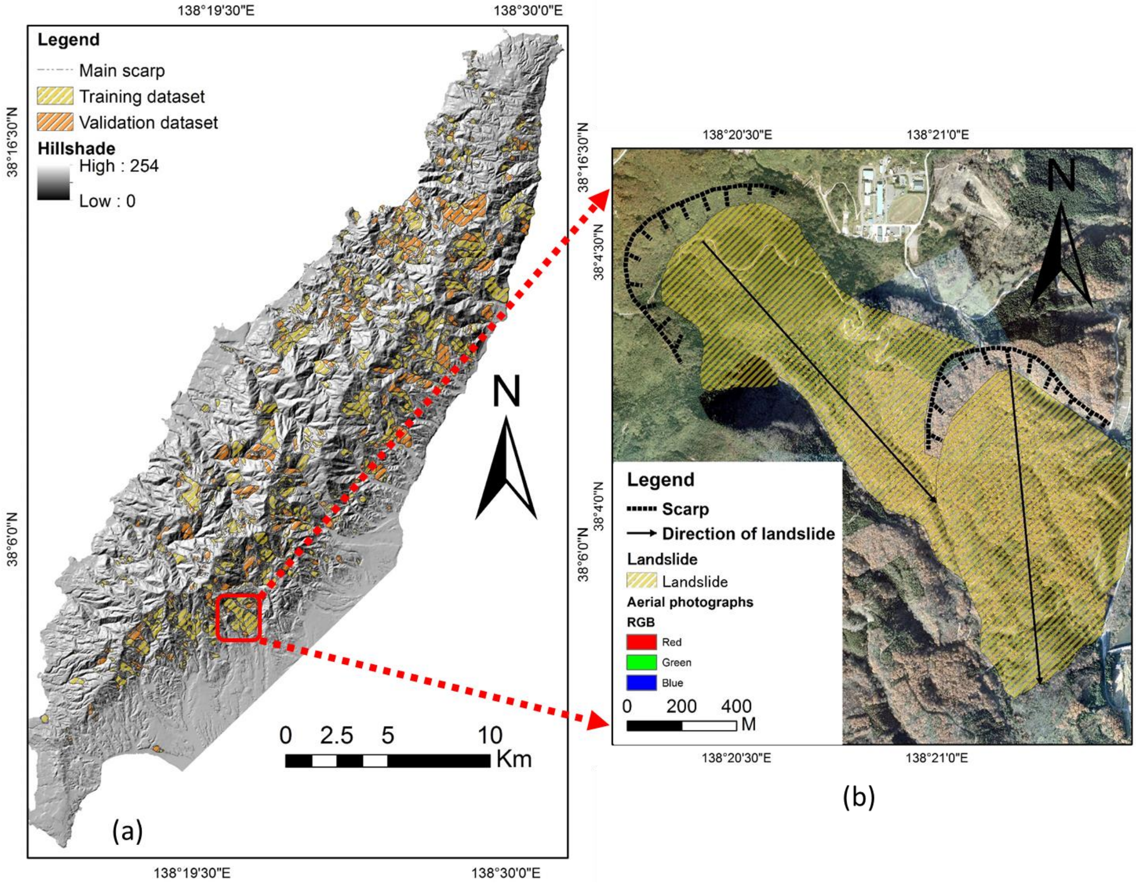

Figure 3-3 Landslide data for Sado Island. (a) Landslide inventory map for the study area randomly divided into two groups (training and validation samples) with a shaded relief map from a 10 m DEM. (b) Enlargement of a landslide location with an aerial photograph from the Midori Niigata and Sado City (acquired in 2005).

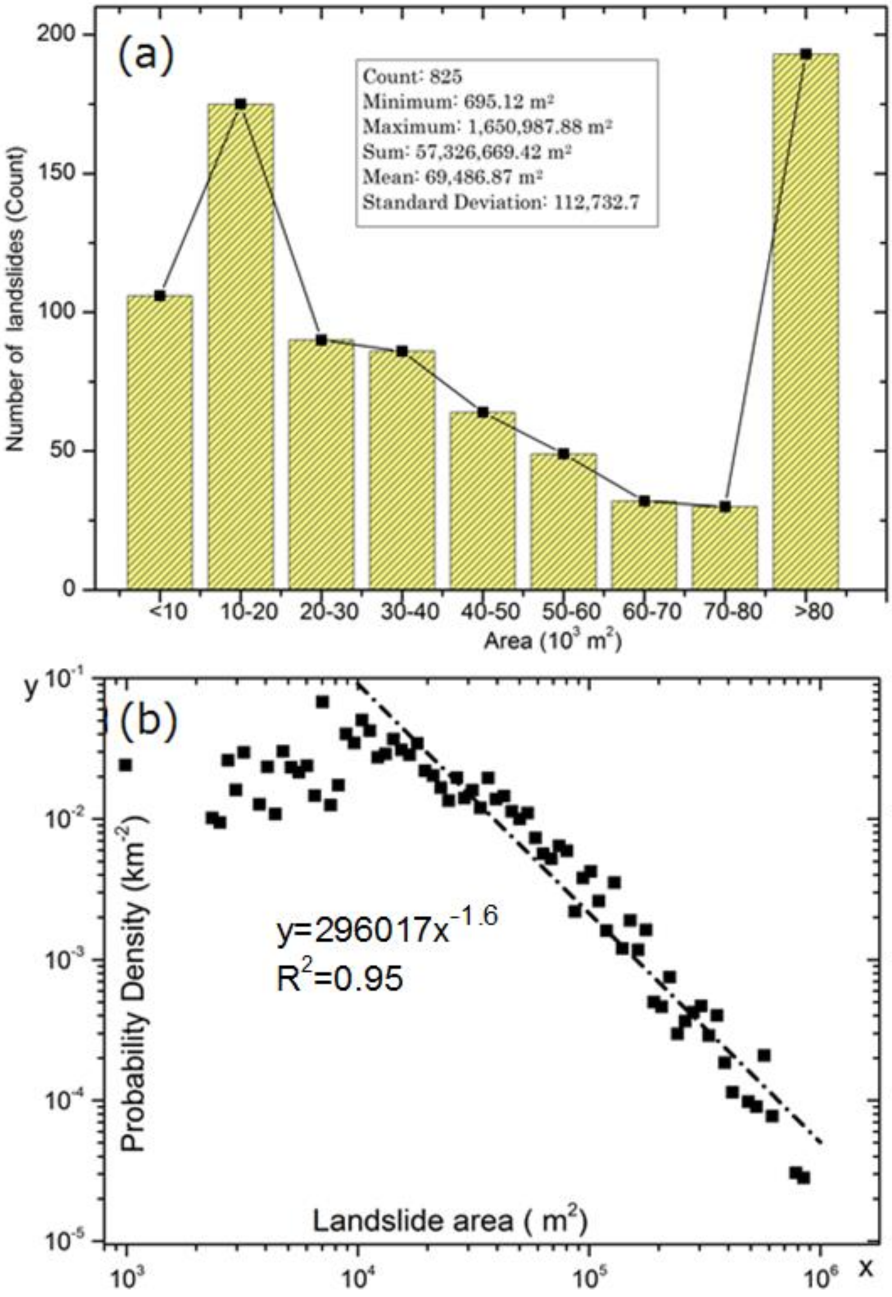

Figure 3-4 Landslide characteristics. (a) Histograms showing the distribution of landslide sizes; (b) Probability distribution of landslide areas in the Sado Island..

Table 3-3 List of data used.

Spatial database	Causative factor	GIS/RS data type	Scale or resolution	Source of data
Landslide inventory map	Landslide	Polygon coverage	1:50,000	NIED
Geological map	Lithology Faults Geological boundary	Polygon coverage Line coverage	1:200,000	GSJ
Topographic map	Morphometric factors	ARC/INFO Grid	10×10 m	GSI
ALOS	NDVI	Raster	10×10 m	JAXA
Aerial photographs		Raster	0.25×0.25 m	Midori Niigata and Sado city

Landslides occurrence are influenced by the interaction of topographic, hydrological and geological factors (Costanzo et al. 2012; Dou et al. 2014a); therefore, the selection of the causative factors is considered to be a fundamental step in landslide susceptibility modeling. In this study, based on analysis of the landslide inventory map and the underlying geo-morphometric conditions (Conoscenti et al. 2015; Dou et al. 2015a), a total of 15 landslide-causative factors (Figure 3-5) commonly found in literature were firstly derived. These factors were extracted from their respective spatial databases (Table 3-3). The source data for the factors may vary in their scale, which affect the accuracy of landslide susceptibility models (Lee et al. 2004b). To be commensurate with the diversity of the data sources and difference in the scales, we converted all the factors to a raster format with a resolution of 10 m that corresponds the DEM resolution.

The 10 m digital elevation model (DEM) obtained from GSI was used to derive elevation, slope angle, slope aspect, total curvature, profile curvature, plan curvature, *CTI* and *SPI* using ArcGIS 10.2 software. The detailed classes and maps of these factors are shown in Figure 3-5 and Table 3-3.

Elevation is widely used for the assessment of landslide susceptibility. The variation in elevation may be related to different environmental settings such as vegetation types and rainfall (Catani et al. 2013). Slope angle is typically considered to be an influential factor because it controls the shear forces acting on hillslopes (Dou et al. 2009a; Tien Bui et al. 2011). Slope aspect, related to sunlight exposure and drying winds control on soil moisture, were also considered an important factor (Magliulo et al. 2008). Total curvature is defined as the change in slope along a small arc of the curve. The profile curvature is the curvature in the downslope direction, while the plan curvature is the curvature of the topographic contours. All of them were found to influence the triggering of landslides (Dou et al. 2015g). Profile curvature influences the driving and resisting stresses within a landslide in the direction of motion, and controls the change in velocity of mass movement flowing down the slope; whereas, the plan curvature controls the convergence or divergence of landslide material and water in the direction of the landslide motion (Ohlmacher 2007). *CTI* and *SPI* are hydrological factors frequently used for the assessment of landslides (Beven and Kirkby 1979; Gessler et al. 1995; Jebur et al. 2014).

Lithology is considered one of the most influential factors in landslide susceptibility mapping because of it represents geo-mechanical characteristics of rocks (Costanzo et al. 2012). In this study, the lithology and faults were derived from the geology map at 1:200,000 scale published by GSJ. A total of 10 lithological units were constructed: metamorphic, plutonic and intrusives, sedimentary (mudstone, sandstone, and slate + sandstone), volcanic (andesite lava, basalt, dacite lava, dacite, and rhyolite lava).

It was found that geologic boundaries often relates to the rock strength. A high density of geologic boundary means lower stability and may lead to increase in landslide occurrences

(Dou et al. 2014a). Therefore the distance to geological boundaries also considered as a factor in this study. The data of geological boundaries are provided by the Geological Survey of Japan (GSJ). The geological boundaries represent the boundary between different geological units or rock types, and are represented as lines. The distance to these lines was defined as the distance from the center of the landslide to the closest point of the geological boundaries. Faults have been regarded as a critical factor in triggering landslide in tectonically active areas (Tien Bui et al. 2011). Additionally, the strength of fracturing and shearing stresses crucially influence the slope instability. Hence, distance to faults was also considered in this study to investigate the relationship between lineaments and landslide occurrence.

The vegetation cover and the land use patterns are often found to affect landslide occurrence, because they are related to the anthropogenic interference on hillslopes (Pradhan and Lee 2010; Zhu et al. 2010). The normalized difference vegetation index (NDVI) was generated from the available cloud free ALOS satellite images (10 m resolution) acquired on November 5th, 2006. $NDVI$ is an indicator that reflects the amount of green vegetation (Pettorelli et al. 2005) and can be computed using the following equation:

$$NDVI = (NIR - RED)/(NIR + RED) \tag{3-1}$$

where NIR and RED are the spectral reflectance of near infrared and red bands of the electromagnetic spectrum, respectively. The values of $NDVI$ vary from -1 to 1 and a higher value implies a denser green vegetation whereas lower values indicate sparse vegetation. High $NDVI$ values are due to high concentration of chlorophyll that cause a relatively lower reflectance in the red band implying high stacking of leaves (Pradhan and Lee 2010).

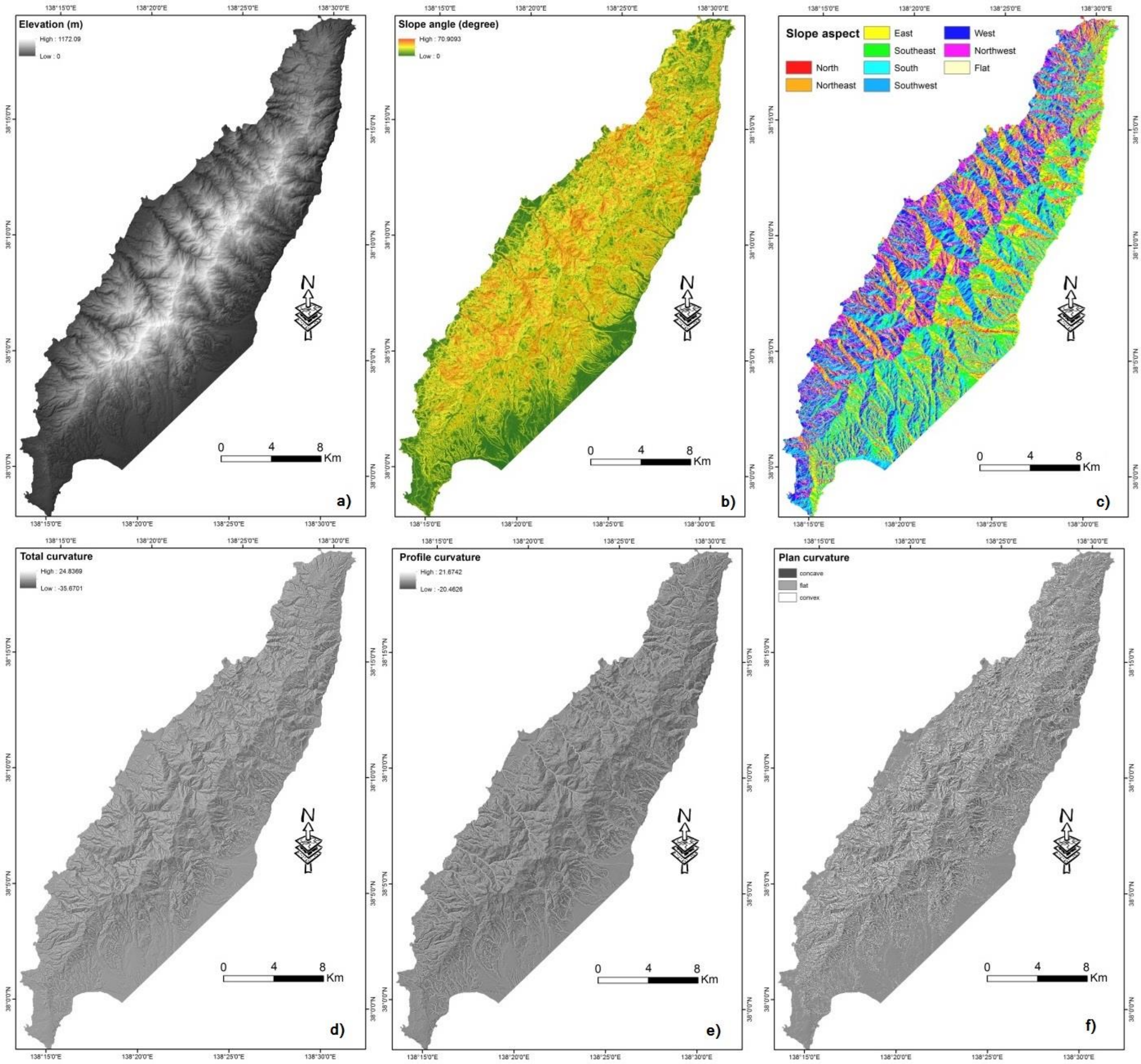

Figure 3-5 Landslide causative factors: a) elevation, b) slope angle, c) slope aspect, d) total curvature, e) profile curvature, f) plan curvature, g) *CTI*, h) *SPI*, i) drainage density (m^{-1}), (j) distance from drainage networks, k) lithology, l) density of geological boundaries, m) distance to geological boundaries, n) distance to faults, and o) *NDVI*.

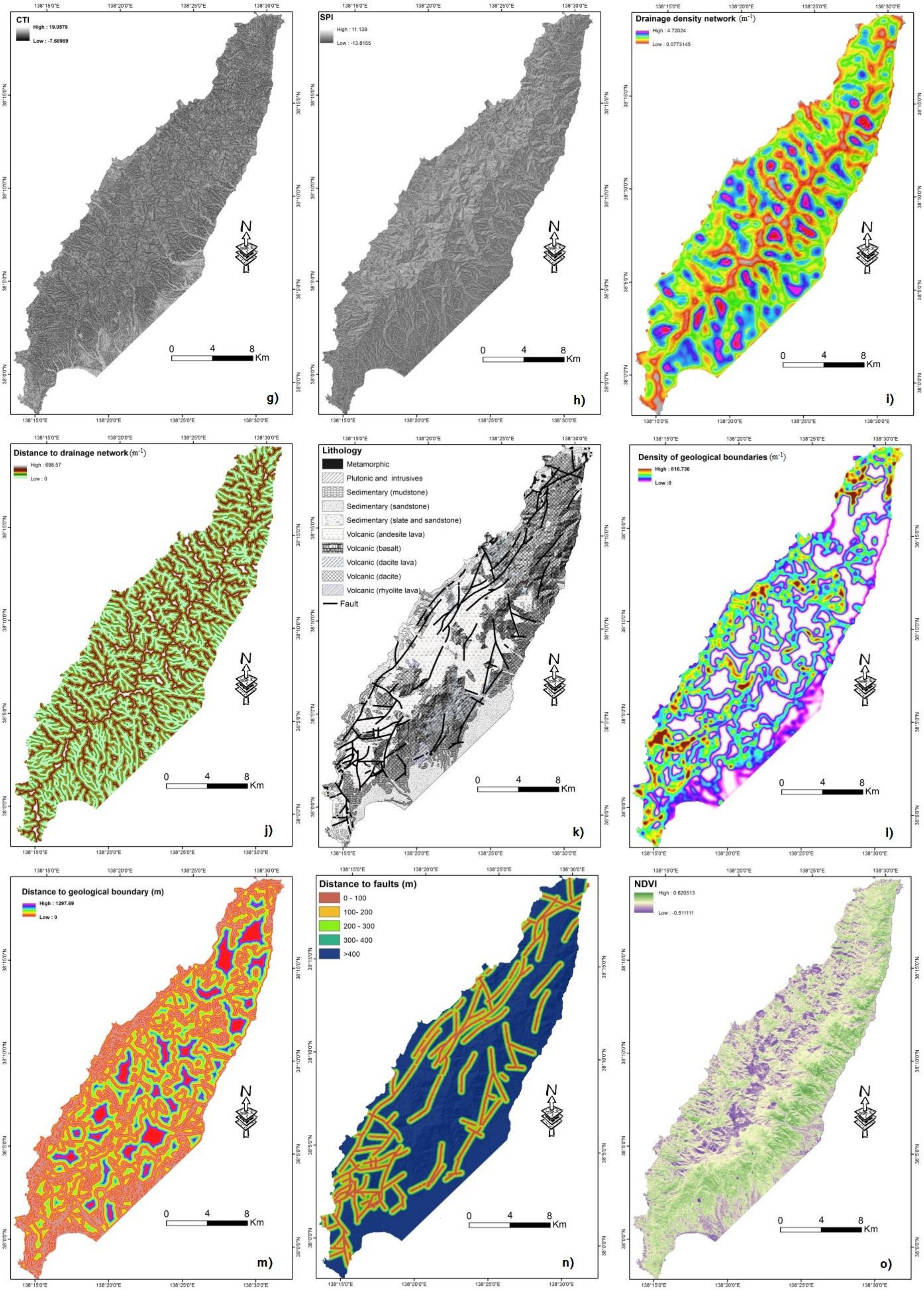

Figure 3-5 Continued.

3.4. Dongjiang Reservoir area, China

Rainfall data from the records of 21 rain gauges in and around the Dongjiang Reservoir area were used to analyze the rainfall characteristics of the major storm. Typhoon BILIS was a strong tropical storm with severe precipitation in a short duration, whose trail was shown in Figure 3-6 and it landed on the coast of Fujian Province, China, on July 14th, 2006, with the maximum wind speed of 108 km/h. Then it weakened into a tropical storm and moved westward and north-westward at the speed of 10-15 km/h until July 16th, 2006 when it disappeared in Hunan Province.

The rainfall observation data from the rain gauge networks around the reservoir on 14-15th July are displayed in Figure 3-7. The Longxi rain gauge shows the maximum rainfall with a total 36-h rainfall of 507 mm and a total monthly rainfall of 826 mm. We also selected two rain gauges, Xingnin and Lianping, for detailed presentation (Figure 3-8). In 48 hours, the cumulative rainfall in Xingnin and Lianping is more than 400 and 285 mm, respectively. In particular the incremental rainfall of Xingnin at 15-18 UTC was approximately 180 mm/3h. Figure 3-9 shows the rainfall contour diagram of the Dongjiang reservoir area in 36 h on July 14th-16th. The reservoir watershed area totally received a rainfall amount of around 6, 6,000, 0000 m^3, leading to a reservoir depth increase of 4.66 m. The reservoir was severely affected by the heavy rainfall in a short time.

Figure 3-6 Trail of the Typhoon BILIS (data from Japan National Institute of Informatics).

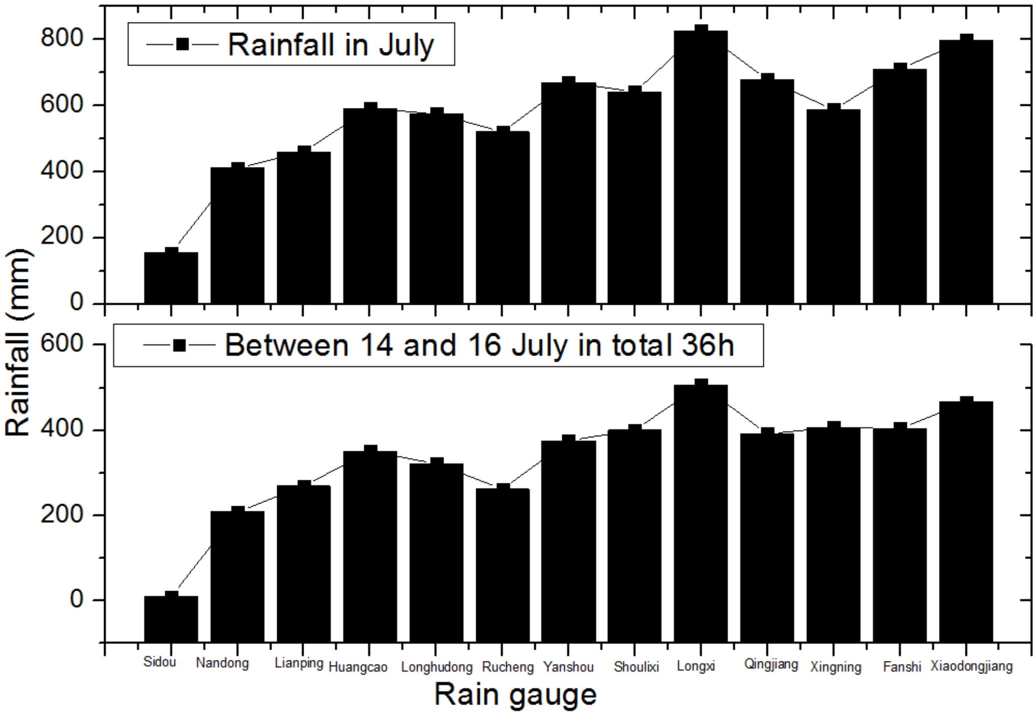

Figure 3-7 Graphs showing rain gauge precipitation in the Dongjiang Reservoir area in July (top) and on 14-16th (36 hour) rainfall (bottom). The depth of water of the reservoir

increased 7.73 m during 14-19th, July, 2006.

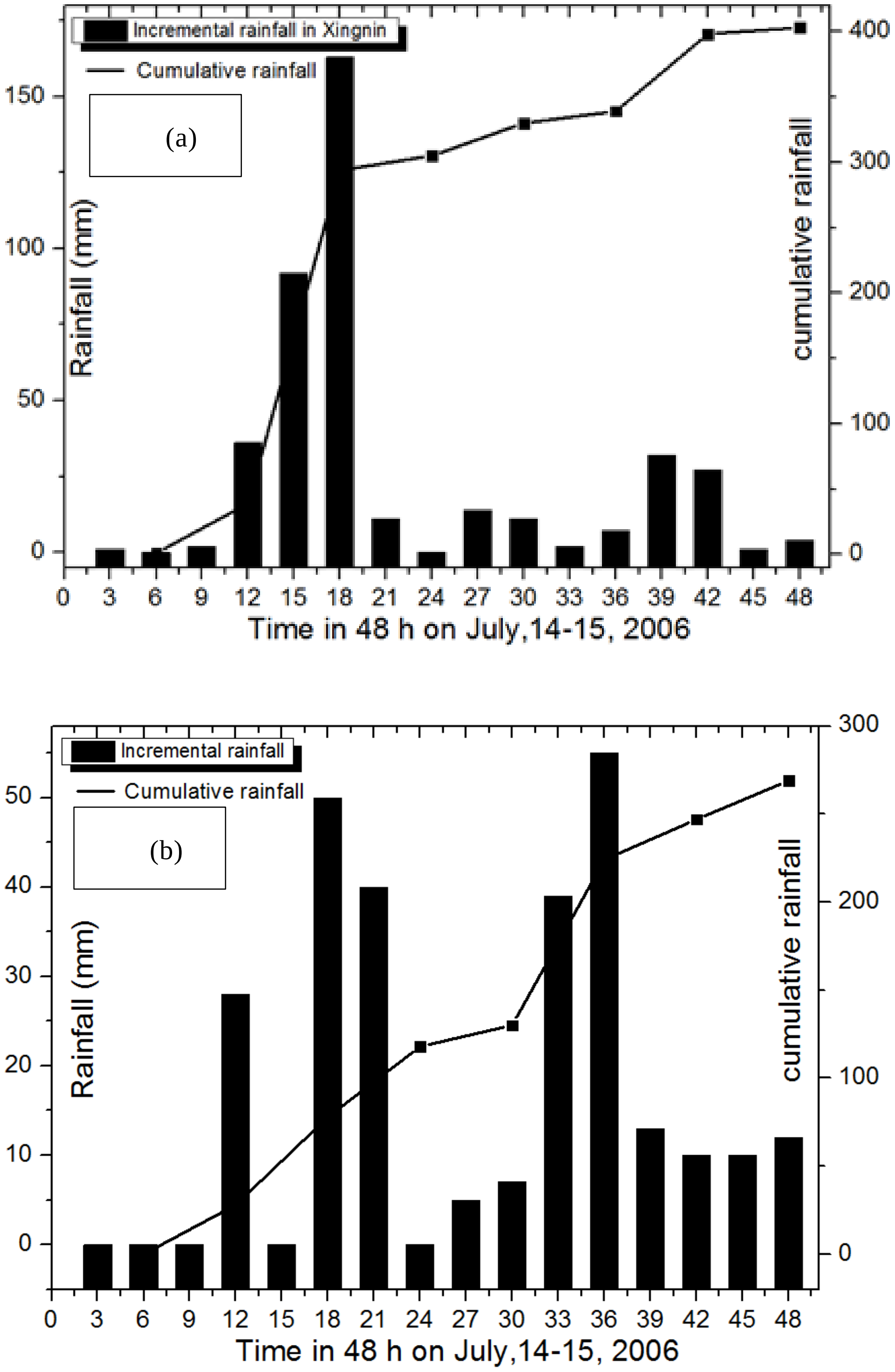

Figure 3-8 Incremental and cumulative rainfall rain gauges in Xingnin (a) and Lianping (b) around the Dongjiang reservoir.

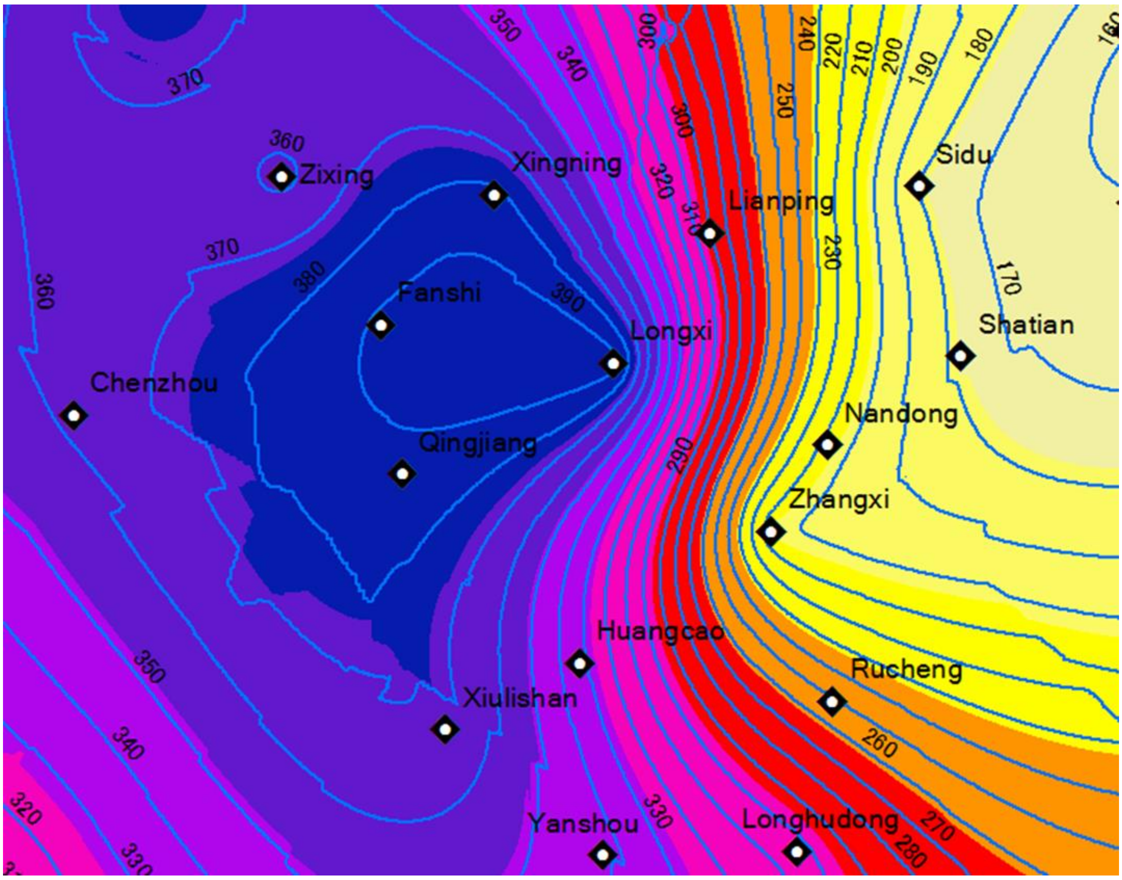

Figure 3-9 Rainfall contour diagram of the Dongjiang Reservoir area in 36 h on 14th-16th, July, 2006. The reservoir area suffered from the total rainfall of around 6, 6,000,0000 m^3, leading to a reservoir depth increase of 4.66 m.

Moreover, using satellite images before and after the event including Quickbirds (0.6 m in resolution) and CBERS (20 m) as listed in Table 3-4, we inventoried 2,407 landslide sites as points each of which corresponds to the center of the landslide scar. The landslide density is approximately 8.2/km^2. Topographic data for analyses were also derived from the 30 m ASTER GDEM (version 2).

Table 3-4 Collected images in the study area

Serial number	Date	Path	Satellite and sensor
1	2000-03-26	373-71	CBERS01-CCD
2	2003-12-26	373-71	CBERS02-CCD
3	2005-11-27	373-71	CBERS02-CCD
4	2005-12-23	373-71	CBERS02-CCD
5	2006-10-05	373-71	CBERS02-CCD
6	2006-10-31	373-71	CBERS02-CCD
7	2006-12-22	373-71	CBERS02-CCD
8	2007-10-27	1-70	CBERS02-CCD
9	2008-11-11	373-71	CBERS02B-CCD
10	2008-05-13	373-71	CBERS02B-CCD
11	2009-01-02	373-71	CBERS02B-CCD
12	2009-10-24	372-71-A-2	CBERS02B-HR
13	2009-10-24	373-71-A-4	CBERS02B-HR
14	2009-01-02	373-71-B-1	CBERS02B-HR
15	2009-01-02	373-71-B-2	CBERS02B-HR
16	2009-01-02	373-71-B-3	CBERS02B-HR
17	2009-01-02	373-71-B-4	CBERS02B-HR
18	2009-01-02	373-71-B-5	CBERS02B-HR
19	2008-01-04	373-71-C-4	CBERS02B-HR
20	2008-01-04	373-71-C-5	CBERS02B-HR
21	2007-12 -18		Quick bird

CHAPTER 4 LANDSLIDE INVENTORY MAPPING

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 (Guzzetti et al. 2012; Dou et al. 2014b).

For a successful landslide hazard evaluation, creating a historical landslide-event inventory is principally crucial for pre-disaster and post-disaster analyses (Guzzetti et al. 2005; Lu et al. 2011; Dou et al. 2015b). Up until now, landslide inventory maps have largely been generated via visual interpretation of aerial photos or satellite images combined with extensive field surveys. Nonetheless, 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 (McKean and Roering 2004a; Dou et al. 2015d). 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 (Dou et al. 2009a; Lu et al. 2011). Experiments based on the emerging technique of expert-based knowledge systems (EKS) 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 (Myint et al. 2011). This research has demonstrated that EKS performance is higher than that of traditional per-pixel approaches in classifying land cover types. The advantage of EKS 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 (Li and Yeh 2004; Dou et al. 2009b).

With the advancement in the spatial resolution of commercial satellite images (e.g., SPOT 5, ALOS, IKONOS, QuickBird, and Wordview1), 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 (Dou et al. 2010). 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 (Qian et al. 2007; Anders et al. 2011; Chang et al. 2012). 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 (Martha et al. 2010; Lu et al. 2011).

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) in this research. 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 (Yang 2007; Vancoillie et al. 2007). 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 EKS, which require high levels of computing competence in cognitive tasks. Although EKS 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 (Ahn et al. 2006; Qian et al. 2007; Du et al. 2013; Minor et al. 2014). 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 Conghua district, PRD, China. According to a literature review, CBR integrated with OOIA, and GA has never been used to detect landslides because they are relatively new techniques.

Figure 4-1 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.

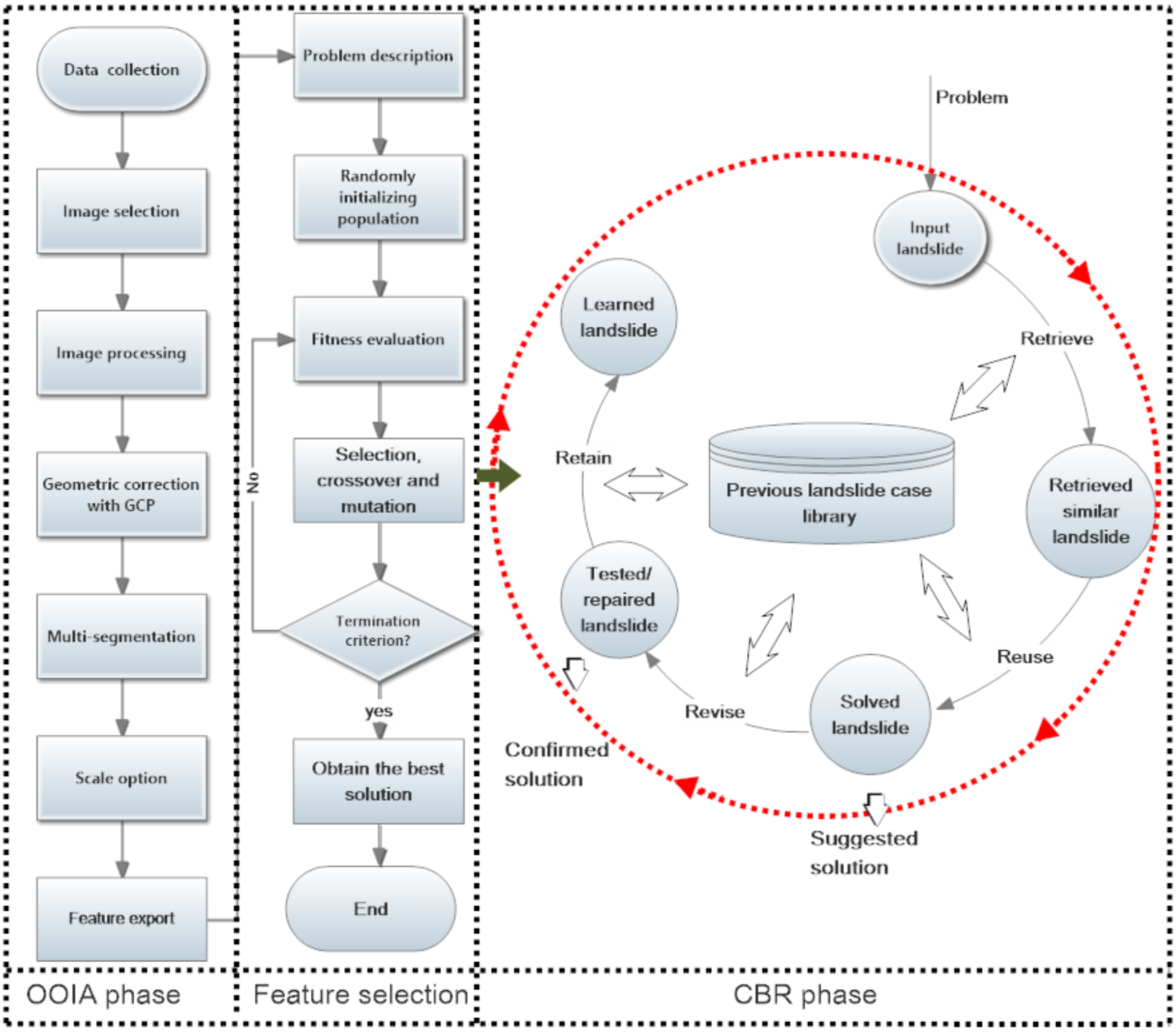

Figure 4-1 Integrated structure for automatic landslide detection, comprising three processes: 1) multi-segmentation by OOIA, 2) feature selection by GA, and 3) detection by CBR and validation using field work data.

4.1. Multi resolution segmentation

Multiresolution segmentation is a bottom-up segmentation algorithm based on a pairwise region-merging technique (Baatz et al. 2004). 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) (Benz et al. 2004). 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:

$$H = \sum_{c} w_c \left(n_{\text{merge}} . \sigma_c^{\text{merge}} - \left(n_{\text{object}_1} . \sigma_c^{\text{object}_1} + n_{\text{object}_2} . \sigma_c^{\text{object}_2} \right) \right) \quad 4\text{-}1$$

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

To remove the image distortions in the different data set, 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 considering that the location is an undulating terrain. This step was crucial before characterizing the attributes of features interest (Martha et al. 2010). Figure 4-2 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 (Baatz et al. 2004). 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 (Drăguţ et al. 2010). In this study, after performing visual interpretation through trial and error, four levels (150, 100, 50, and 30) of the segmentation scale were selected at first. 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 4-2, 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 4-3); 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.

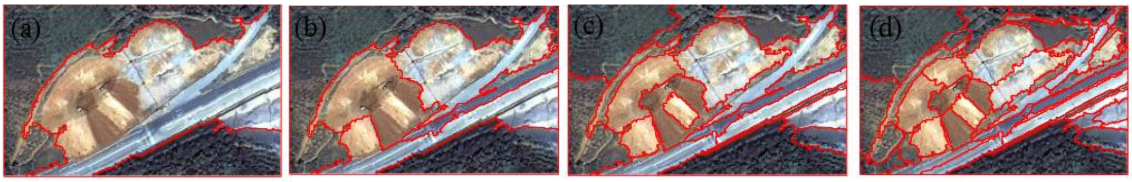

Figure 4-2 Details of QuickBird image segmentation including four scales: (a) 150, (b) 100, (c) 50, and (d) 30.

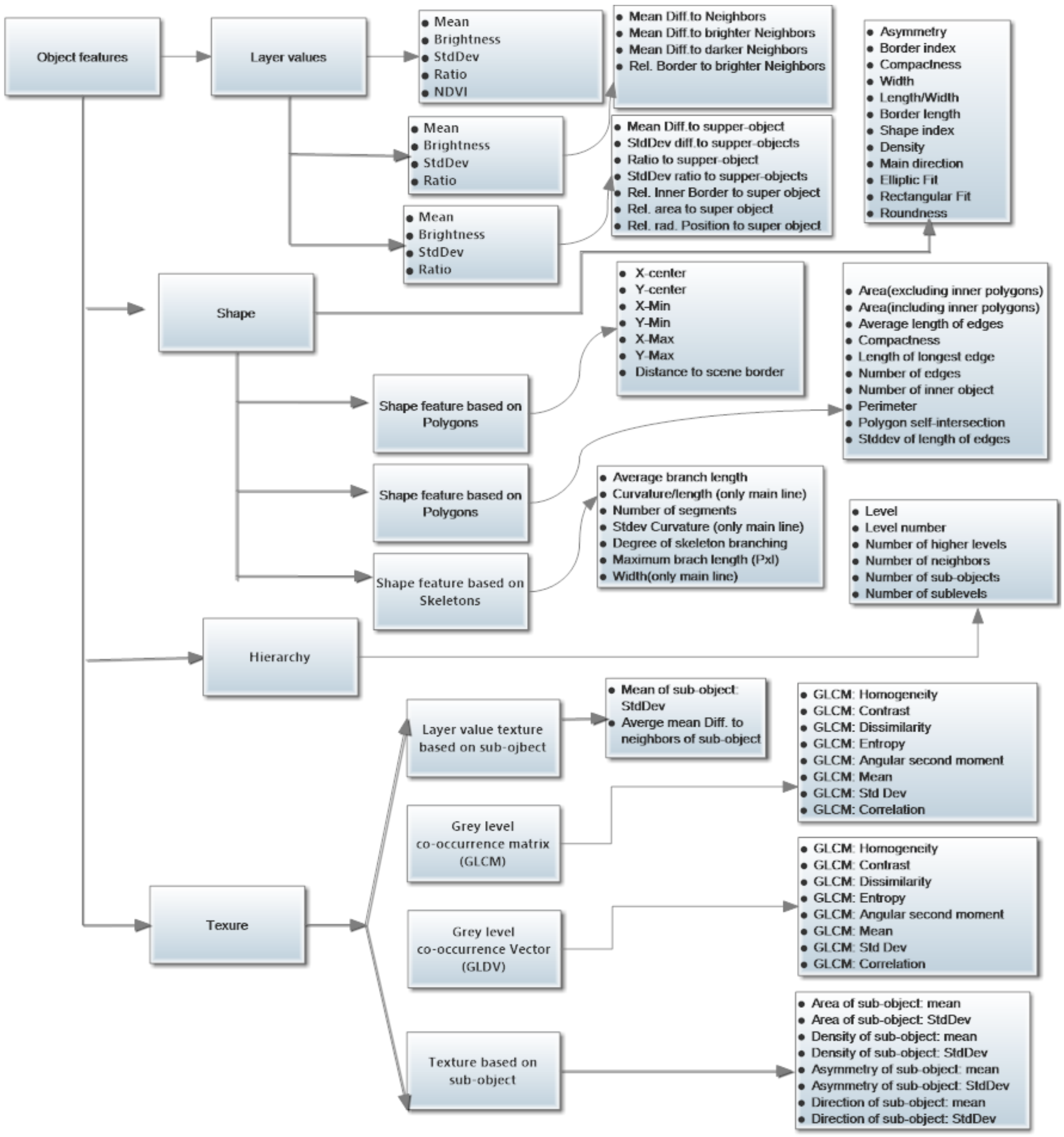

Figure 4-3 Object attributes exported in the OOIA analysis.

4.2. GA based optimization of feature selection

GAs are a class of stochastic search and optimization techniques based on natural selection and evolutionary principles (Addis et al. 2011). This algorithm has been proven to be robust and effective in searching large spaces for a wide range of applications (Tang et al. 2005; Iovine et al. 2005). 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 (Kudo and Sklansky 2000). 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 4-4 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 decrease in search effectiveness (Tang et al. 2005). It facilitates assigning the optimal fitness value for each chromosome. The fitness function $f(x)$ can be expressed as follows:

$$f(x)=\frac{\sum_{i=1}^{n}\sum_{j=1(\omega_i=\omega_j)}^{n}\delta\left(x^i,x^j\right)}{\sum_{i=1}^{n}\sum_{j=1(\omega_i\neq\omega_j)}^{n}\delta\left(x^i,x^j\right)} \tag{4-2}$$

$$\delta\left(x^{i}, x^{j}\right)=\sqrt{\sum_{k=1}^{n} \omega_{k}\left(x_{k}{}^{i}-x_{k}{}^{j}\right)} \qquad (4\text{-}3)$$

here x^i is an n-dimensional feature vector of image object i, $x^i = (x_1^i, x_2^i, \ldots x_n^i)$ is the segmented object, and $\delta\left(x^i, x^j\right)$ are the Euclidean distance between vectors x^i and $x_k{}^i$, which is k-th feature value of i, ω_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.

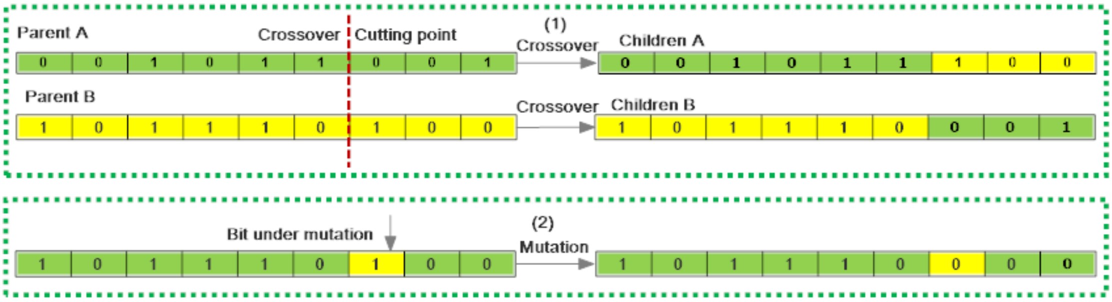

Figure 4-4 Illustration of GA: (top-1) an example of the single point crossover for a binary GA, randomly setting the parents and obtaining the offspring; (bottom-2) the mutation for a binary GA, in which the bits are randomly chosen and the allele's values are altered.

4.3. 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 (Aamodt and Plaza 1994; Jonassen and Hernandez-Serrano 2002; Minor et al. 2014). 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 maximum likelihood classification (MLC) is that it can allow the use of both numeric and nonnumeric data and does not require a normal distribution form (Li and Yeh 2004). 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 4-5 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:

$$X = \left(X_1, X_2 ..., X_n; T_k \right) \tag{4-4}$$

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.

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, and k-nearest neighbor) to calculate *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 when the k-nearest neighbor or Manhattan distance technique is use. 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 (1989) and other studies (Drǎguţ et al. 2010). Hence, this research used the GRA method to calculate similarity because of its advantage of global comparisons. The *SCD* value of GRA is calculated using the following equation:

$$SCD_i(n) = \frac{\underset{i}{Min}\,\underset{n}{Min}\,|X_0(n) - X_i(n)| + \zeta\,\underset{i}{Max}\,\underset{n}{Max}\,|X_0(n) - X_i(n)|}{|X_0(n) - X_i(n)| + \zeta\,\underset{i}{Max}\,\underset{n}{Max}\,|X_0(n) - X_i(n)|} \tag{4-5}$$

where $SCD_i(n)$ is the related coefficient of case *I* at point n, with the value ranging [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 ζ is the identification coefficient. The *SCD* assessment includes multiple values to prevent dispersion in a given system (Deng 1982); thus, the related grade (*RG*) between sequences can be defined by dividing the related coefficient by its average value:

$$RG_i = \frac{1}{N}\sum_{k=1}^{N} SCD_i(n) \tag{4-6}$$

where *N* is the number of features. The value of *RG* ranges from 0 to 1.

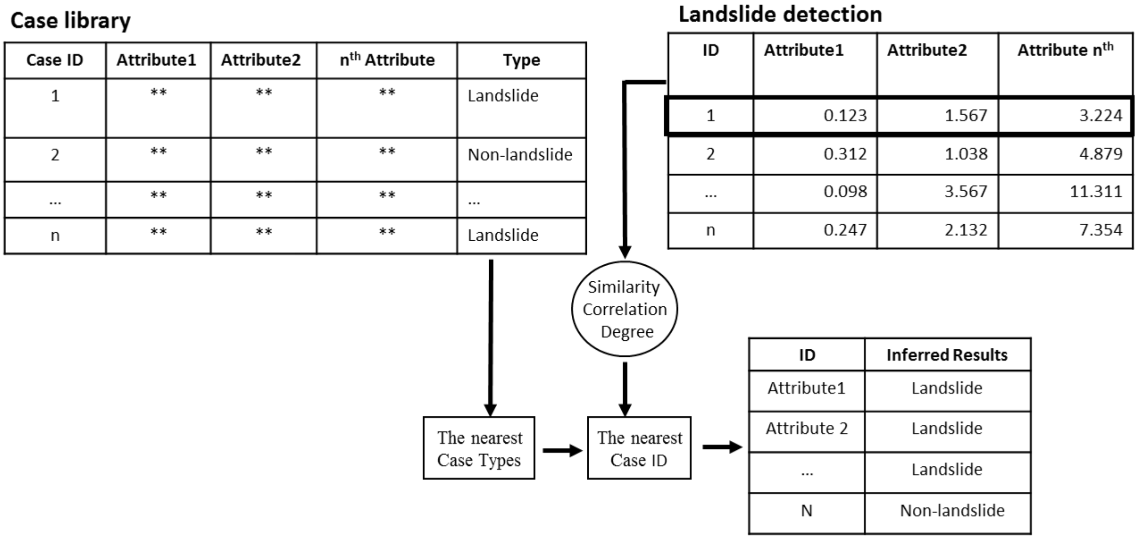

Figure 4-5 Illustration of landslide detection using the CBR method in this study.

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

$$Y = \frac{X - X_{Min}}{X_{Max} - X_{Min}} \tag{4-7}$$

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

Figure 4-6 shows examples of landslides 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.

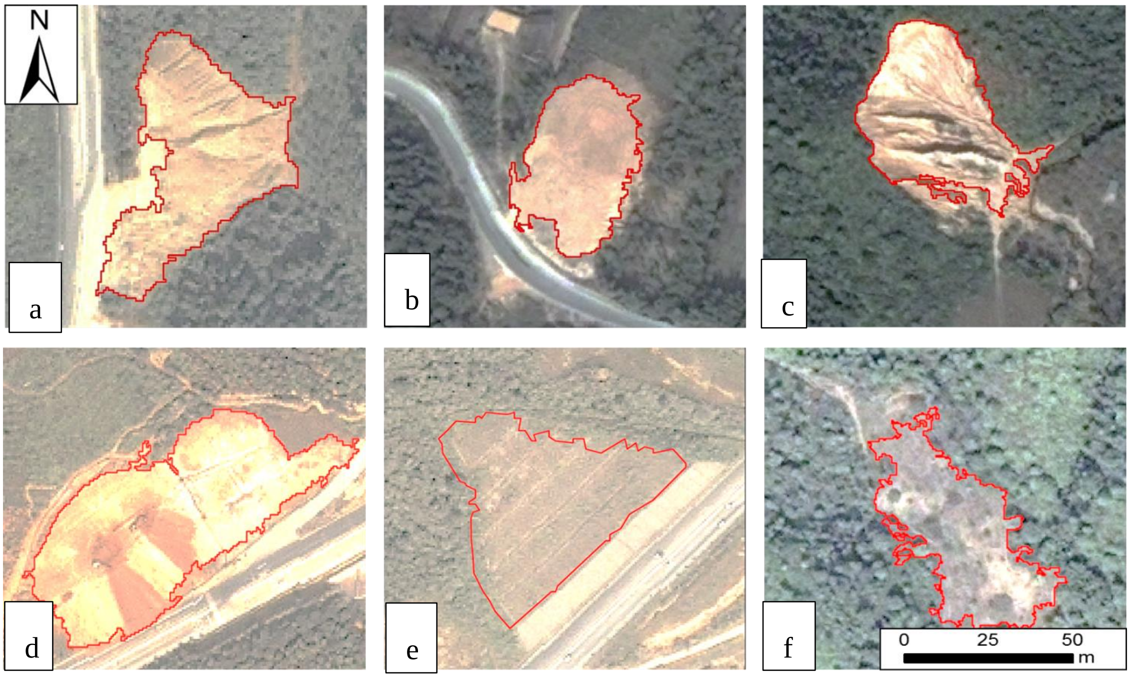

Figure 4-6 Examples showing the preparation of reference landslides in the case library, in the form of Quickbird images: a, b, c, and d are young landslides (bright area with no vegetation cover); e and f are old landslides (less bright areas covered with concrete structures and vegetation).

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 (Matlab 2010).

4.4. Accuracy estimation

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 (Congalton 1991; Radoux and Bogaert 2014). 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

(Congalton 1991; Yang et al. 2013). 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 (Pontius and Millones 2011). Radoux and Bogaert (2014) recommended the use of sample polygons to assess thematic accuracy when spatial objects are identified on a map as polygons. Mondini et al. (2011) 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. (2013) successfully applied overall accuracy to assess the landslide identification in the 2008 Wenchuan earthquake area in China. However, thus far, no universal optimal fit method exists for evaluating map accuracies (Radoux and Bogaert 2014).

In this study, the *ROC* index was employed for validating the model. *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 (Swets 1988). Although it is widely accepted in binary calculations, it is less frequently 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 under the curve (*AUC)* 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 4-7 shows the schematic diagram of *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 (Fawcett 2006).

$$\mathrm{Y} = \text{Sensitivity } = TNR/(TNR + FPR) \tag{4-8}$$

$$\text{Specifity } = \frac{\sum \text{True negative}}{\sum \text{Condition negative}} = \mathrm{TNR}/(\mathrm{TPR} + \mathrm{FNR}) \tag{4-9}$$

$$\text{X=1- Specifity} = 1 - \frac{\sum \text{True negative}}{\sum \text{Condition negative}} = 1 - \mathrm{TNR}/(\mathrm{TPR} + \mathrm{FNR}) \tag{4-10}$$

$$AUC = \int_0^1 ROC(X)dX \tag{4-11}$$

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

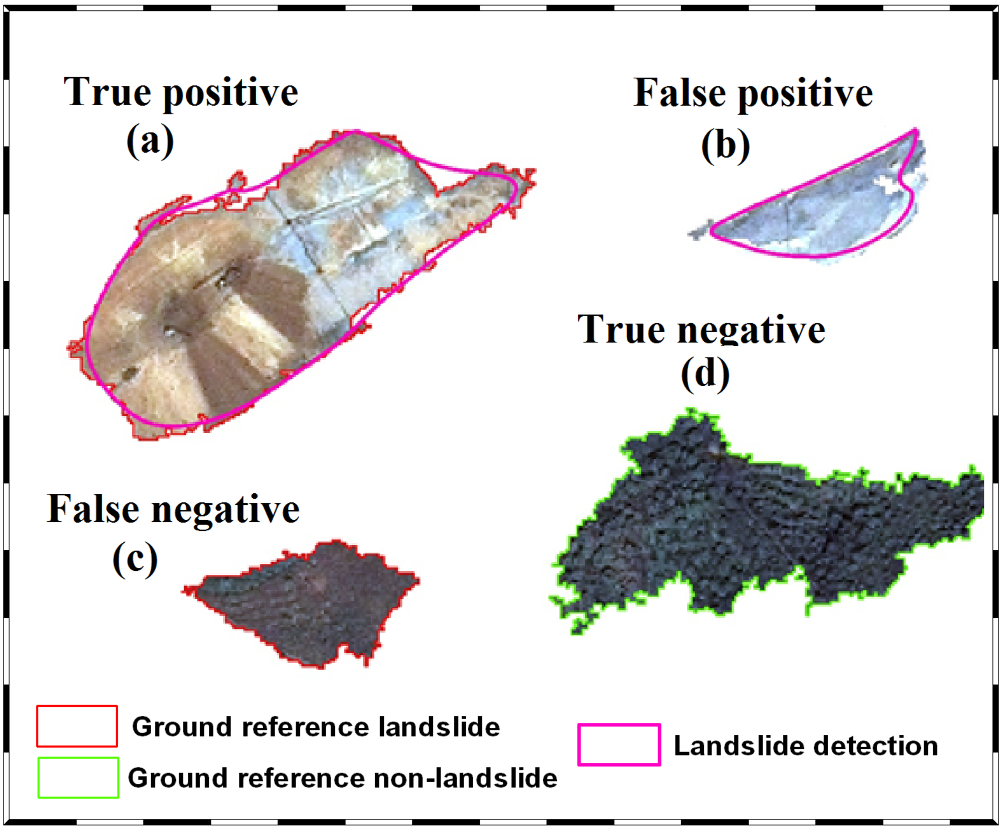

Figure 4-7 Schematic diagram showing the computation of true positives, false positives, false negative, and true negatives for verifying the accuracy of the model. Herein, (a) true positive: actual landslides that were correctly classified as landslides, (b) false positive: non-landslides that were incorrectly classified as landslides, (c) false negative: landslides that were incorrectly classified as non-landslides, and (d) true negative: non-landslides that were correctly classified as non-landslides.

4.5. Application

4.5.1 CBR-based multi-scale landslide detection

Based on previous studies and objects that were recognizable and interpretable from the satellite images (Van Den Eeckhaut et al. 2010), three classes of objects were categorized in this study: old landslides, young landslides, and non-landslides. Old landslides appear as large crescentic amphitheaters 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 2-2 c, e, and 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. 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 4-8 illustrates the relationship between fitness value and the feature number, indicating that the number decreases with the feature number for the first 11 features. After the 11th feature, the fitness value increases (Figure 4-8). Based on these settings, the first 11 features were selected by the GA (Table 4-1). The selected features included spectral features such as the layer mean and ratio, textural features such as gray level co-occurrence matrix (GLCM is a measurement of the variation intensity in the pixels of interest, which is tabulation of how often different arrangements of gray levels co-occur in an image or image section) and grey level difference vector (GLDV is the sum of the diagonals of the GLCM) (Vancoillie et al. 2007; Blaschke 2010), 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 selection was not related to the scale sizes.

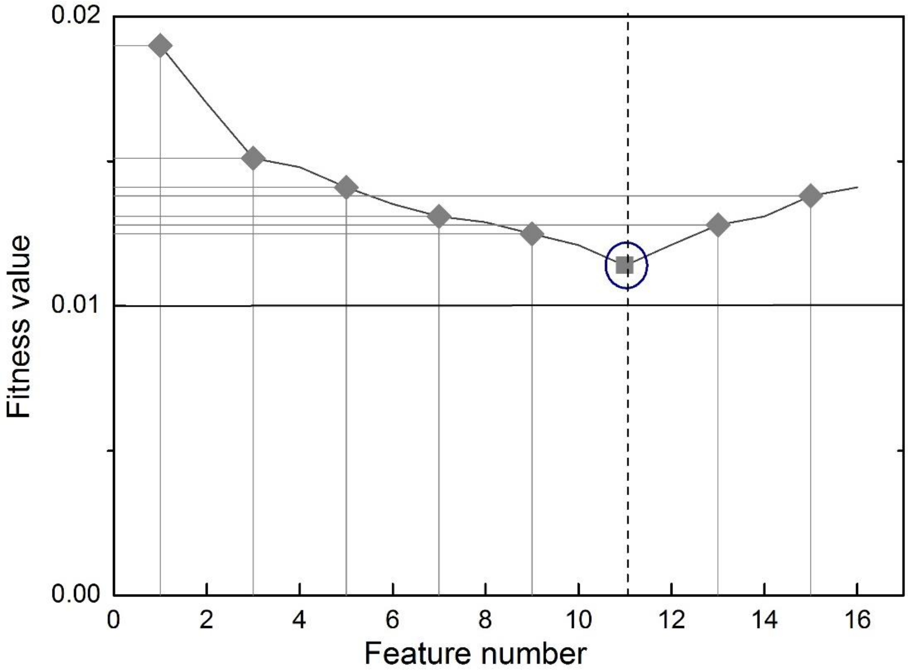

Figure 4-8 Relationship between the feature number and the fitness value in the GA process.

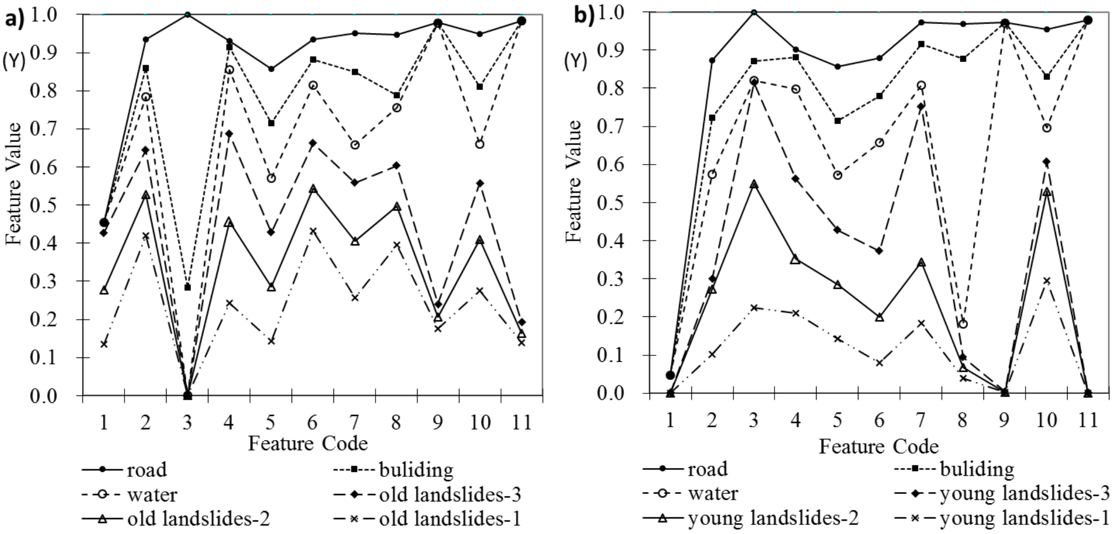

Figure 4-9 Plots of old (a) and young (b) landslide curves. The curves have similar periodic

trend changes.

Table 4-1Feature selection optimized by the GA.

Code	Feature	Depiction
1	Mean difference 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	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 (GLDC). 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 at comparatively high elevation.
11	Slope	Slope = Raise/Run, [0, 90°], affectsing the stability of slope failure.

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 4-9). This implies that the *SCD* value of the GA can be used to compare the experimental cases with the cases in the library.

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 4-10). The 800 cases in the library were then prepared from the three scales of segmentation. 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^2. The identified landslides varied in size between 59 and 32700 m^2. The results also indicate 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.

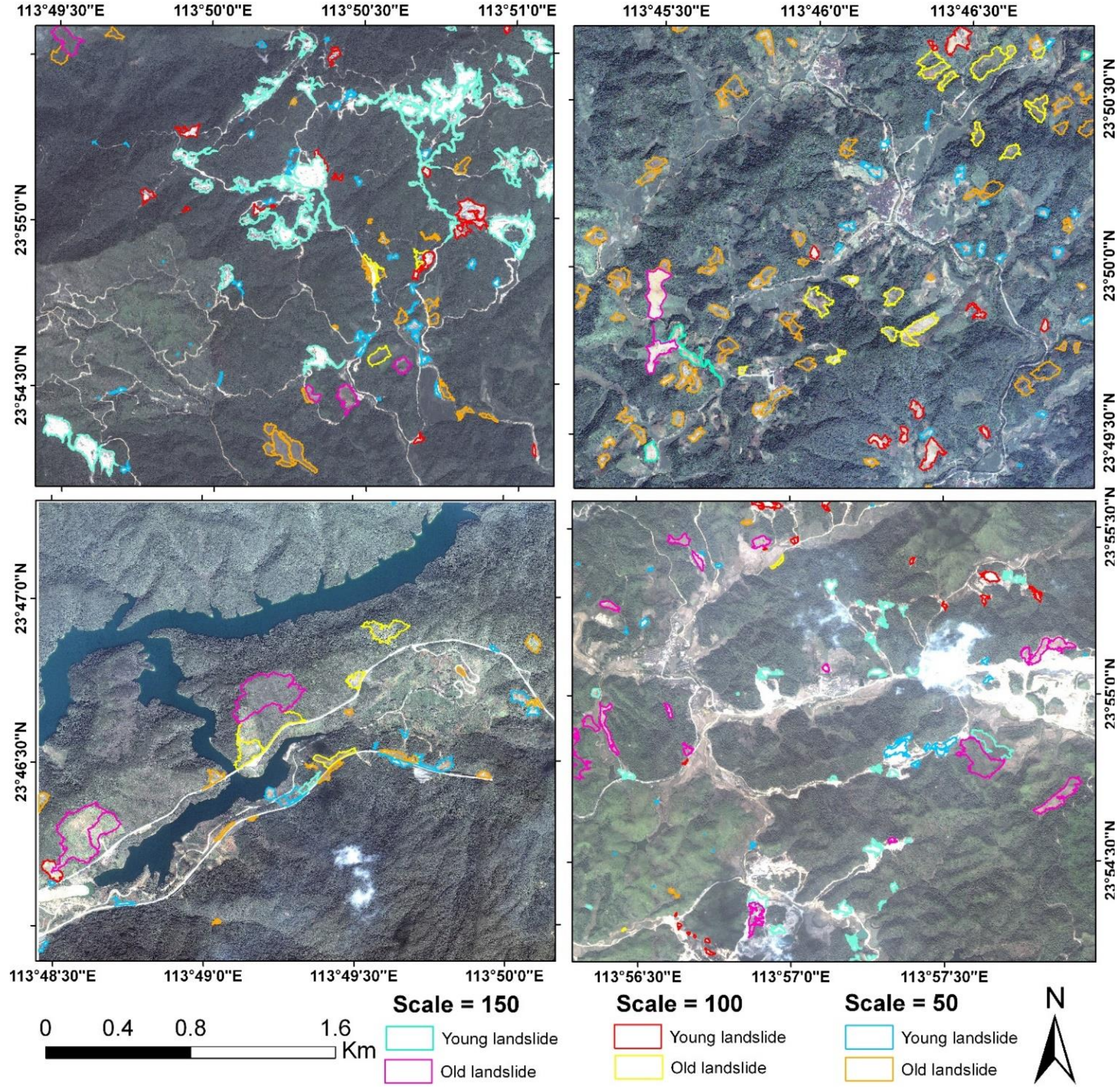

Figure 4-10 Results of landslide detection at different scales (150, 100, and 50).

4.5.2 Validation

Accuracy assessment of the results was verified using field investigation, GPS control points, and a QuickBird image acquired in 2007 (Figure 4-11). The accuracy of the proposed methodology was evaluated with the second set of samples separated from the 800 cases. ROC curves for the two types of landslide mentioned in Section 4.1 are displayed in. *AUC* values for the young landslides (0.9) are higher than those for the old landslides (0.82). Table

4-2 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 indicate 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.

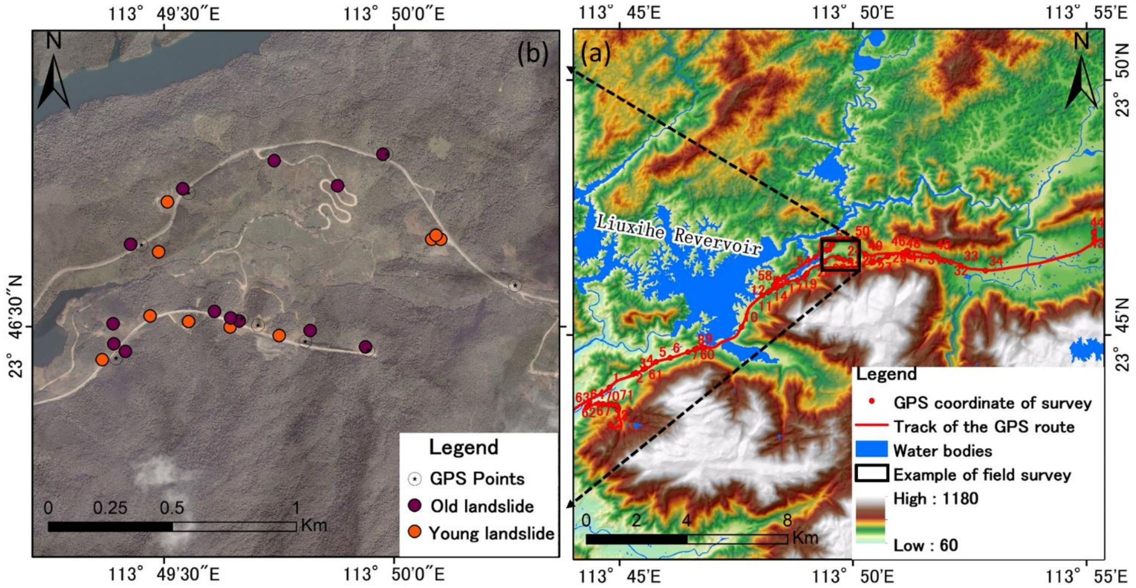

Figure 4-11 Data for validation: (a) track of GPS route for field survey; (b) examples of landslide detection points overlying the QuickBird image, using CBR.

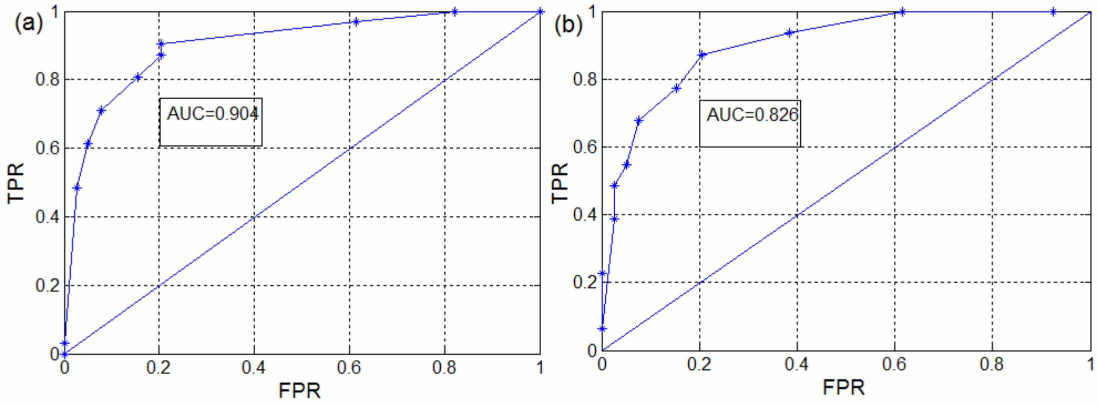

Figure 4-12 Prediction rate of ROC curve for each landslide class using CBR: (a) young landslides; (b) old landslides.

Table 4-2 Comparison of accuracy in the CBR method

Landslide type	User's accuracy	Producer's accuracy	Omission error	Commission error	Overall accuracy
Old landslides	0.82	0.85	0.15	0.18	
Young landslides	0.86	0.89	0.11	0.14	0.87

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 (σ) 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 4-13 a and b show that the *AUC* values of the OOIA method for old and young landslides were 0.7 and 0.74, respectively.

Table 4-3 shows the results of the OOIA method, indicating an overall accuracy of 0.75. Additionally, other comparisons were made by applying a standard traditional per-pixel classification, viz., and the supervised MLC for classifying the same images with the same training and testing samples, using the Erdas software. This method was derived from the Bayes theorem, which expresses a posteriori distribution based on spectral data. The *AUC* value of the MLC method for old and young landslides were 0.64 and 0.67, respectively

(Figure 4-13).

Table 4-4 shows that the maximum likelihood method of overall accuracy was 0.68. Both results suggest 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.

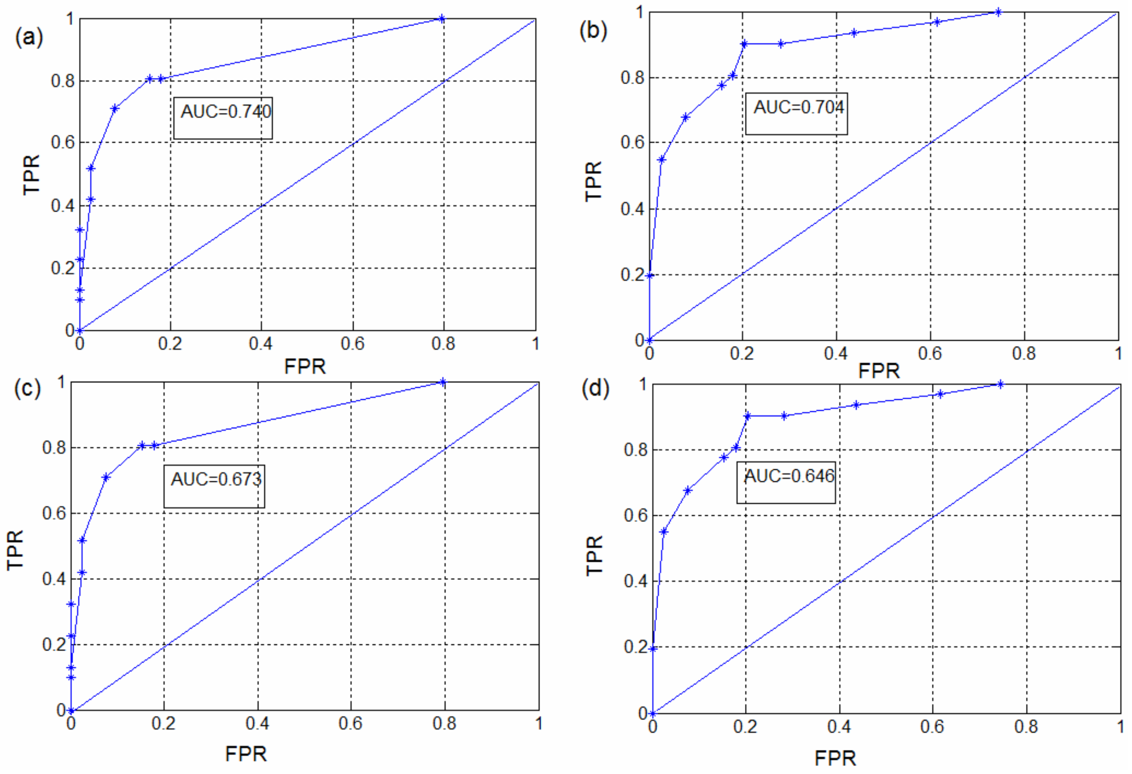

Figure 4-13 Prediction rate of the ROC curve for each landslide class: (a) young landslides, using OOIA; (b) old landslides, using OOIA; (c) young landslides, using MLC; and (d) old landslides, using MLC.

Table 4-3 Comparison of accuracy in the OOIA method.

Landslide type	User's accuracy	Producer's accuracy	Omission error	Commission error	Overall accuracy
Old landslides	0.82	0.85	0.15	0.18	
Young landslides	0.86	0.89	0.11	0.14	0.75

Table 4-4 Comparison of accuracy in the supervised maximum likelihood method.

Landslide type	User's accuracy	Producer's accuracy	Omission error	Commission error	Overall accuracy
Old landslides	0.65	0.62	0.38	0.35	0.68
Young landslides	0.67	0.63	0.37	0.33	

4.6. Summary

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.

CHAPTER 5 CLASSIFICATION OF LANDSLIDE INVENTORY

Shallow failures and erosion into bedrock play important roles in shaping landscapes in mountainous areas (Oguchi 1996). However, previous studies tended to focus on the spatial prediction of only a single type of landslides (Chang and Chao 2006; Cheng et al. 2010; Lee and Tsai 2008). Therefore, few studies differentiated the probabilities of shallow and deep-seated landslides.

Shallow and deep-seated landslides innately differ in their size, extent and the risk posed (Zêzere et al. 2005). Such landslide types reflect a variety of environmental and geological factors (Turner and Schuster 1996; Schmidt et al. 2001; Roering et al. 2005). Differentiating the two landslide types is helpful in evaluating the geomorphic hazards contributing to the soil erosion and sediment discharge for the protection of human settlements and infrastructures (Dramis and Sorriso-Valvo 1994; Korup 2005a; Korup 2006; Larsen et al. 2010; Chang and Lin 2013). Some studies focused on factors controlling the occurrence of deep-seated landslides. Roering et al. (2005) applied an algorithm developed from the relationship between hillslope angle and curvature to differentiate large, deep-seated landslides from debris flows and shallow slope failures. May (2007) developed an automated algorithm that granted the identification and mapping of deep-seated landslides over a wide area.

Landslides are regarded as a nonlinear system and therefore a sophisticated mathematical approach is required for their analysis. The landslide prediction methods developed in recent years (Chang et al., 2014; Dou et al., 2014; Guzzetti et al., 1999, 2006; Hoopes, 2014) may not always maintain their stability and reliability when used with a smaller training dataset (Crowther and Cox 2006). On the contrary, support vector machines (SVMs) have been known to work well even with smaller training datasets (Chi et al. 2008). Huang et al. (2002) found that SVMs with smaller training data was more persistently accurate and stable than the MLC, decision tree classification (DTC), and artificial neural network (ANN) classification with larger training data. Therefore, SVMs have been widely applied in various fields including remote sensing, computer science, pattern recognition and economics (Marjanovic et al. 2009; Tien Bui et al. 2012b).

Fine-resolution topographic data are necessary for geomorphological analyses of landslides (Glenn et al. 2006). McKean and Roering (2004) used high-resolution topographic data from airborne laser altimetry to identify earth flows by contrasting surface roughness and surface texture. We have used a 2 m airborne Light detection and ranging (Lidar) DEM for our analyses. Lidar data have been used to create the detailed geomorphic maps that differentiate the types of landforms and characterize landforms including landslides (Ardizzone et al. 2007; Schulz 2007; Van Den Eeckhaut et al. 2009a; Pulko et al. 2012).

Intense earthquakes are important as preparatory and triggering factors of landslides (Keefer 2000; Harp et al. 2011). In 2004, Niigata Prefecture in central Japan experienced an unprecedented number of landslides, including shallow and deep-seated landslides, triggered by the Chuetsu earthquake. Several studies have been made on this event (Chigira and Yagi 2005; Kieffer et al. 2006), and most of them focused on the contribution of geologic and geomorphic factors to landslide occurrence. This study incorporates the topographic and geological variables to predict the spatial differentiation of landslide types that may occur by future earthquakes using SVMs and relatively few training samples.

Selection of landslide causative factors is the fundamental step in predicting landslides. This study assumes that factors previously used to study landslide susceptibility are equally useful in predicting the probable landslide types. Therefore we considered landslide susceptibility studies by Caniani et al. (2007), Dou et al. (2009), Guzzetti et al. (1999), (2006), Lee and Tsai (2008), Lee et al. (2008) and Lee and Sambath (2006) for selecting factors summarized in Table 5-1. The factors include several DEM derivatives: elevation, slope angle, aspect, curvature, distance from drainage network, *CTI*, and *SPI*, as well as lithology, distance from the nearest geological boundary and density of geological boundaries. The density of geological boundaries was computed within a circle of 200 m radius based on Kawabata and Bandibas (2009) who studied the same area. The factors used were calculated using ArcGIS and the results are shown in Figure 5-1.

Table 5-1 Landslide causative factors used in the study.

Source dataset	Conditioning factors	Description or definition	Significance
DEM	Elevation	Height above the mean sea level	Vegetation, climate, solar energy.
	Slope	Rate of change in elevation for each cell	Overland and sub-surface flow velocity, runoff rate, rainfall, vegetation, geomorphology, soil water content.
	Aspect	Downslope direction of the maximum rate of change in value	Evapotranspiration, distribution of flora and fauna.
	Curvature	Curvature of the line parallel or perpendicular to the direction of the maximum slope	Erosion or deposition.
	Distance from drainage networks	The minimum distance from the closest drainage network	Erosion, ground water condition and relative stability.
	Compound topographic index (CTI)	$CTI = \ln(As/\tan\beta)$ with As specific catchment area per unit channel width orthogonal to the flow direction and β the slope angle	Also known as the topographic wetness index (TWI); it correlates with soil moisture.
	Stream power index (SPI)	$SPI = As \times \tan\beta$	Erosive power of overland flows, thickness of soil horizons.
Geological map	Lithology	Lithological information as types	Strength of the surface and direct control over most of the factors.
	Distance from the nearest geological boundary	The minimum distance from the boundary of the nearest geological unit	Stress, cohesion.
	Density of geological boundaries	Number of geological boundaries per unit area	Stress, cohesion, tectonic activity.

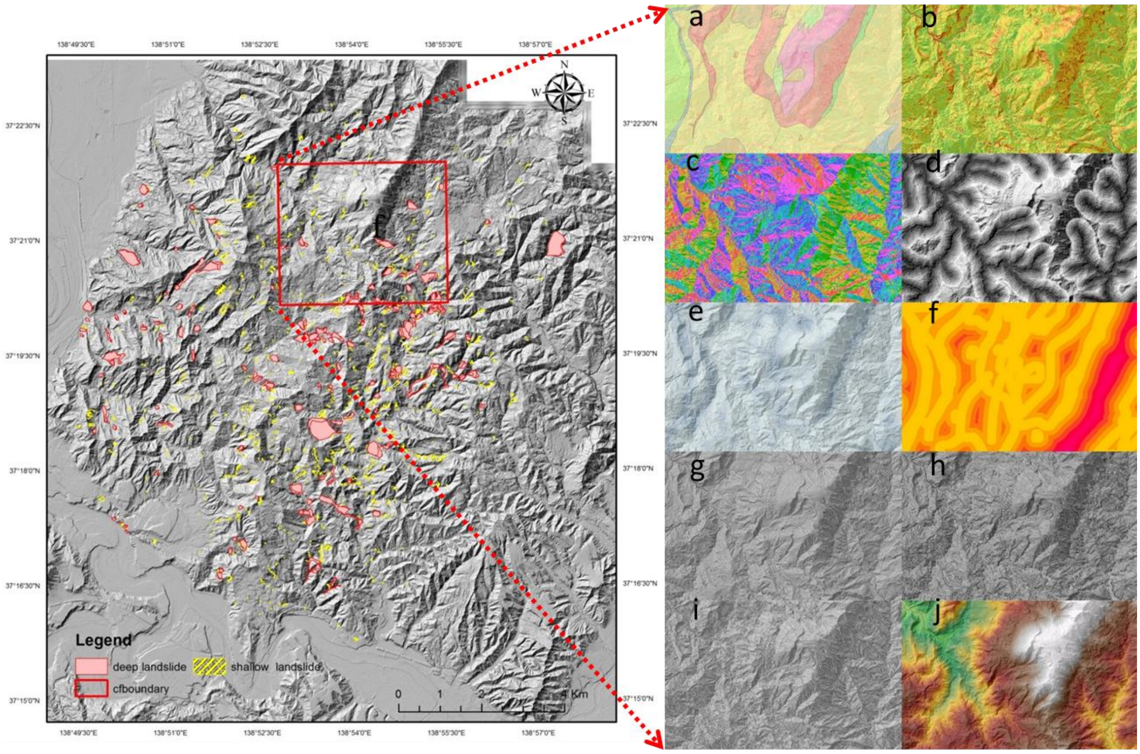

Figure 5-1 Factors maps: a) lithology, b) slope angle, c) aspect, d) distance from drainage network, e) density of geologic boundaries, f) distance from the nearest geologic boundary, g) curvature, h) compound topographic index (*CTI*), i) stream power index (*SPI*), and j) elevation.

Elevation greatly influences precipitation and vegetation due to its orographic effect. Slope angle is also an important factor that influences slope stability (Lee et al. 2008). Aspect can be an indirect measure of hydro-meteorological influences on vegetation and weathering and thus the resistance of slope material (Kawabata and Bandibas 2009; Dou et al. 2014b). Curvature controls hydraulic flow in relation to convergence and divergence, and hence landslide occurrence (Dai et al. 2011). Differentiation of shallow and deep-seated landslides may depend on lithology (Wieczorek and Jäger 1996) that affects the thickness of weak beds. Groundwater condition and soil moisture in relation to topography affect landslides (Zinko et al., 2005). To describe these, several topographic indices have been proposed. *CTI* and *SPI*, developed by Beven and Kirkby (1979) and Gessler et al. (1995), respectively, are used in this study. CTI and SPI could be calculated as follows:

$$\text{CTI} = \ln(\text{As}/\tan\beta) \quad (5\text{-}1)$$

$$\text{SPI} = \text{As} \times \tan\beta \quad (5\text{-}2)$$

where As is the specific catchment area per unit channel width orthogonal to the flow direction (m^2/m) and β is the slope angle expressed in degrees. CTI is strongly correlated with soil moisture and SPI takes into account both slope and flow accumulation and hence is correlated with erosion potential. These indices also provide information on soil depth and soil constituents (Moore et al. 1991; Florinsky 2012) suggesting that they can be invaluable factors in predicting the landslide types (Wieczorek and Jäger 1996). Density of geological boundaries reflects the density of geological discontinuities or relatively weaker zones which may contribute to slope instability (Dou et al. 2014b). The data of geological boundaries between different geological rock types were obtained from the Geological Survey of Japan (GSJ).

5.1. SVM model in landslide type classification

5.1.1 SVM model

SVMs provide supervised learning models with associated algorithms based on the concept of optimal separating hyperplane and statistical learning theory (Vapnik 1998). SVMs are useful non-linear classifiers whose goal is to find the widest margin between two classes in a feature space. Figure 5-2 illustrates this concept: ovals and squares are two kinds of samples and the separating hyperplane (H) is one of possible planes which can separate the two classes; and the distance between the two dotted lines in Figure 5-2 is called margin. The vectors which constrain the width of the margin are called the support vectors. Although SVMs are often considered easier to use than neural networks, inappropriate parameter setting often leads to unsatisfactory results (Chang and Lin 2011).

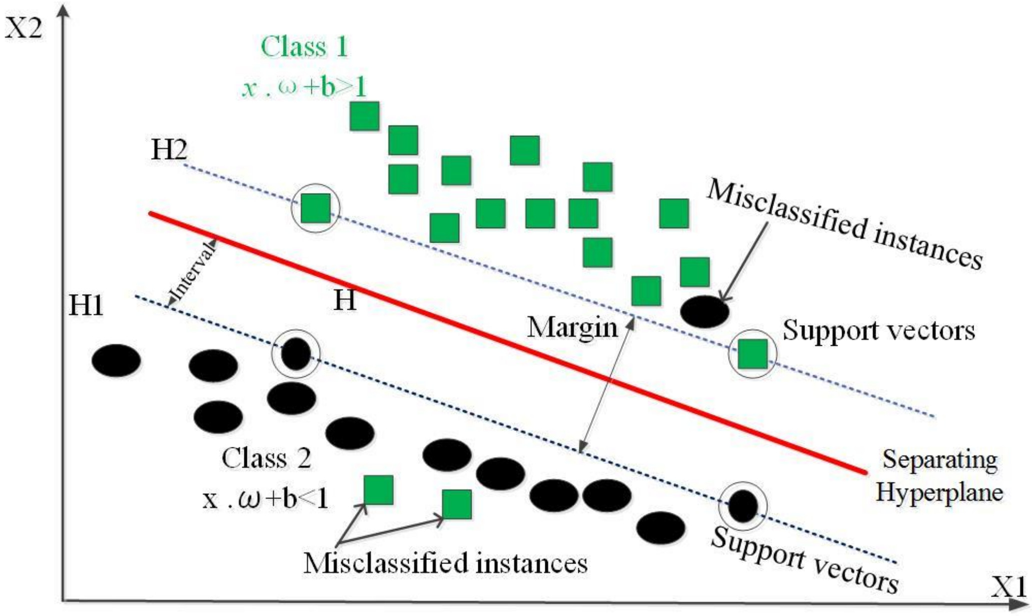

Figure 5-2 Illustration of the optimal separating hyperplane.

SVMs involve a training phase using a training dataset. SVMs are not relatively sensitive to the size of training samples and may successfully perform with a limited number of training samples (Mantero et al. 2005; Foody and Mathur 2006). Foody and Mathur (2004) demonstrated that only a quarter of the entire training data set was sufficient for high accuracy classification.

For a set of linear separable training vectors x_i (i = 1, 2…n), consisting of two classes represented as $y_i = \pm 1$, SVMs try to obtain an optimal hyperplane by differentiating the two classes by solving the following optimization function (Vapnik 1998):

$$\underset{w,b,\xi}{Min} : \quad \frac{1}{2} w^T w + c \sum_{i=1}^{1} \xi_i \tag{5-3}$$

Subjected to the constrains of the following equation:

$$\begin{gathered} y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{gathered} \tag{5-4}$$

where w is a coefficient vector, b is the offset of the hyperplane from the origin, ξ_i is the positive slack variable, *c* (> 0) is the penalty parameter of the error term; and the kernel function is:

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{5-5}$$

Normally four basic kernel functions, linear (LF), polynomial (PF), radial basis (RBF) and sigmoid (SF) functions, are used in the SVMs. Table 5-2 shows their formulas. LF is the simplest one; PF is non-stationary and well suited when all training data are normalized; SF is from the field of neural networks; and RBF depends on the distance from the origin.

Table 5-2 SVMs kernel functions, their parameters, and their overall accuracy.

Kernel	Formula	Kernel parameters	Accuracy of prediction	
			Training	Testing
Linear function (LF)	$k(x_i, x_j) = x_i^T x_j$		69.76%	59.23%
Polynomial function (PF)	$k(x_i, x_j) = (\gamma x_i^T x_j + \Upsilon)^d, \Upsilon > 0$	γ, Υ	95.26%	72.81%
Radial basis function (RBF)	$k(x_i, x_j) = \exp(-\gamma \lvert x_i - x_j \rvert^2),\ \Upsilon > 0$	γ	94.38%	84.31%
Sigmoid function (SF)	$k(x_i, x_j) = \tan h(\gamma x_i^T x_j) +\ \Upsilon,\ \Upsilon > 0$	Υ	71.25%	62.37%

In this study, the four kernel functions were employed. The 10 landslide controlling factors were normalized into 0 to 1 to limit the dominating effect of large values:

$$y = \frac{x - Min(x)}{Max(x) - Min(x)} \tag{5-6}$$

where *y* is the normalized data value and *x* is the original data value. Partial input and targets for SVMs training samples after normalization are listed in Table 5-3.

In this study, 1225 landslides were randomly divided into two groups: training and testing datasets. Varying training sample size (30%, 40%, and 50%) was used to test the effect of the size. The shallow and deep-seated landslides were assigned values of 1 and 0, respectively. Prior to the calculations, the penalty parameter (*c*) was obtained using the cross-validation technique detailed in LIBSVM. LIBSVM is an integrated software used for support vector classification, distribution estimation (one-class SVM and multi-class classification) and regression (Chang and Lin 2011). This led to lesser support vectors and significantly reduced time of calculation. To test the stability and reliability of the model, the process was iterated 10 times as done by other scholars (Chang and Chao 2006; Pradhan et al. 2010b; Seppelt et al. 2012) Each time a random set of landslides was selected for training, and the remaining data are used as the test samples. The SVMs model was operated on the platforms of Matlab 2012a and LIBSVM.

Two separate SVMs were used in this study: one using only the landslide types as input to differentiate them, and the other using not only landslide types but also data for points in non-landslide areas as input. The non-landslide points were randomly selected from areas with no landslides.

Table 5-3 Partial inputs and targets for SVMs training samples.

Type	Elevation	Slope	Aspect	Curvature	Density of geological boundary	Distance from drainage network	Landslides (target values)		
							(1,0, 0) = deep (0,1,0) = shallow (0, 0, 1) = non-landslide		
non-landslide	0.19	0.17	0.75	0.23	0.22	0.27	0	0	1
non-landslide	0.20	0.30	0.49	0.25	0.78	0.04	0	0	1
non-landslide	0.28	0.09	0.28	0.25	0.28	0.33	0	0	1
deep	0.02	0.34	0.05	0.25	0.41	0.05	1	0	0
deep	0.06	0.51	0.85	0.25	0.47	0.13	1	0	0
deep	0.07	0.48	0.85	0.25	0.46	0.06	1	0	0
shallow	0.43	0.01	0.75	0.26	0.12	0.01	0	1	0
shallow	0.41	0.07	0.98	0.26	0.12	0.01	0	1	0
shallow	0.45	0.41	0.19	0.25	0.15	0.01	0	1	0
deep	0.21	0.50	0.96	0.26	0.18	0.48	1	0	0
deep	0.22	0.19	0.83	0.25	0.20	0.37	1	0	0
shallow	0.48	0.51	0.22	0.25	0.23	0.01	0	1	0
shallow	0.42	0.12	0.73	0.25	0.12	0.01	0	1	0
shallow	0.43	0.01	0.75	0.26	0.12	0.01	0	1	0
shallow	0.41	0.07	0.98	0.26	0.12	0.01	0	1	0
shallow	0.45	0.41	0.19	0.25	0.15	0.01	0	1	0
shallow	0.43	0.31	0.28	0.25	0.21	0.01	0	1	0
shallow	0.45	0.28	0.71	0.27	0.37	0.01	0	1	0
shallow	0.43	0.05	0.36	0.26	0.35	0.01	0	1	0
non-landslide	0.33	0.12	0.86	0.21	0.23	0.51	0	0	1
non-landslide	0.34	0.09	0.28	0.20	0.36	0.33	0	0	1
deep	0.02	0.30	0.99	0.24	0.43	0.04	1	0	0
deep	0.18	0.38	0.88	0.26	0.28	0.32	1	0	0
deep	0.16	0.31	0.85	0.26	0.28	0.29	1	0	0
deep	0.06	0.39	0.98	0.27	0.44	0.15	1	0	0
deep	0.05	0.45	0.84	0.26	0.45	0.11	1	0	0

5.1.2 Back-propagation for a feedforward neural network

An ANN can be regarded as a quantitative black-box model approach that emulates human pattern recognition functions (Aleotti and Chowdhury 1999). Moreover, earth science's non-linearity analysis and prediction can be studied applying this efficient tool. ANN has also been successfully implemented for evaluating landslide susceptibility by several researchers (Arora et al. 2004; Falaschi et al. 2009; Pradhan et al. 2010a; Zare et al. 2013a) . Our study focus on a particular type of ANN model, known as a back-propagation neural network. A BPNN algorithm is used in the feedforward ANN. This is typically used to train the network among the different types of ANN models including RBF, general regression neural networks (GRNN), and probabilistic neural networks (PNN). The BPNN algorithm is simply a gradient-descent algorithm (also called a generalized delta rule) that uses to minimize the total error or mean error of target computed by the neural network. This algorithm is a neural network that is composed of three layers, input, hidden, and output. The structure of a typical three layer BPNN is displayed in Figure 5-3. The input layer propagates components of a special input vector after weighting synaptic weights to each node in the hidden layer. Each hidden layer computes outputs corresponding to these weighted sums through a non-linear/linear function, e.g., log-sigmoid, purelin, or tan-sigomid (Yesilnacar and Topal 2005; Prasad et al. 2012). The BPNN algorithm comprises of two paths, feed forward and backward paths. The feed forward path is expressed as follows (Rumelhart et al. 1986):

$$y_j = F(X_j) = F(W_{oj} + \sum_{i=1}^{I} W_{ij} x_i) \tag{5-7}$$

$$Z_k = F(Y_k) = F(W_{ok} + \sum_{j=1}^{J} W_{jk} y_i) \tag{5-8}$$

where x_i, y_j, and Z_k represent the input, hidden, and output layers, respectively, W_{oj} and W_{ok} are the bias weights for setting the threshold values, X_j and Y_k represent temporarily computing results before using the activation function, and F is the activation function applied in the hidden and output layers. In this study, a sigmoid function or logistic function is chosen as the activation function. Thus, the output y_j and Z_k can be expressed as:

$$y_j = F(X_j) = F(\frac{1}{1+e^{-X_j}}) \tag{5-9}$$

$$Z_k = F(Y_k) = F(\frac{1}{1+e^{-Y_k}}) \tag{5-10}$$

The value of F ranges from 0 to 1.

For error back propagation weight training, the error function can be defined as (Rumelhart et al. 1986):

$$E = \frac{1}{2}\sum_{k=1}^{K} e_k^2 = \frac{1}{2}\sum_{k=1}^{K}(t_k - z_k)^2 \tag{5-11}$$

where, t_k and e_k are the predefined target value and error in each output node, respectively. The goal is to minimize E, the error between the desired and actual output values of the network. To adjust the weight, a gradient-descent strategy was used. The weight between the hidden and output layers can be expressed as follows:

$$\frac{\partial E}{\partial w_{jk}} = -e_k \frac{\partial F(Y_k)}{\partial Y_k} y_j = -e_k F^{'}(Y_k) y_j = -\delta_k y_j \tag{5-12}$$

$$\delta_k = e_k F^{'}(Y_k) = (t_k - z_k) F^{'}(Y_k) \tag{5-13}$$

Therefore, the weight adjustment in the link can be computed by:

$$\Delta w_{jk} = \eta y_j \delta_k \tag{5-14}$$

Where, η is the learning rate with value ranges between 0 and 1. If the learning rate is relatively small, the BPNN is slow to converge the network. Conversely, a learning rate that is overly large can lead to a widely oscillating network. Thus, it is preferable to choose a single value throughout the experiment. The new weight herein is updated by the following equation:

$$w_{jk}(n+1) = w_{jk}(n) + \Delta w_{jk}(n) \tag{5-15}$$

Here n is the number of iterations in the network.

Similarly, the error gradient in links between the input and hidden layers can be derived from the partial derivative with respect to w_{ij} ,

$$\frac{\partial E}{\partial w_{ij}} = \left[\sum_{k=1}^{K} \frac{\partial E}{\partial z_k} \frac{\partial z}{\partial Y_k} \frac{\partial Y_k}{\partial y_j} \right] * \left(\frac{\partial y_j}{\partial X_j} \right) * \left(\frac{\partial X_j}{\partial w_{ij}} \right) = -\Delta_j x_i \tag{5-16}$$

$$\Delta_j = F^{'}(X_j) \sum_{k=1}^{K} \delta_k w_{jk} \tag{5-17}$$

The new weight in the hidden and input links can be updated as:

$$\Delta w_{ij} = \eta x_i \Delta_j \tag{5-18}$$

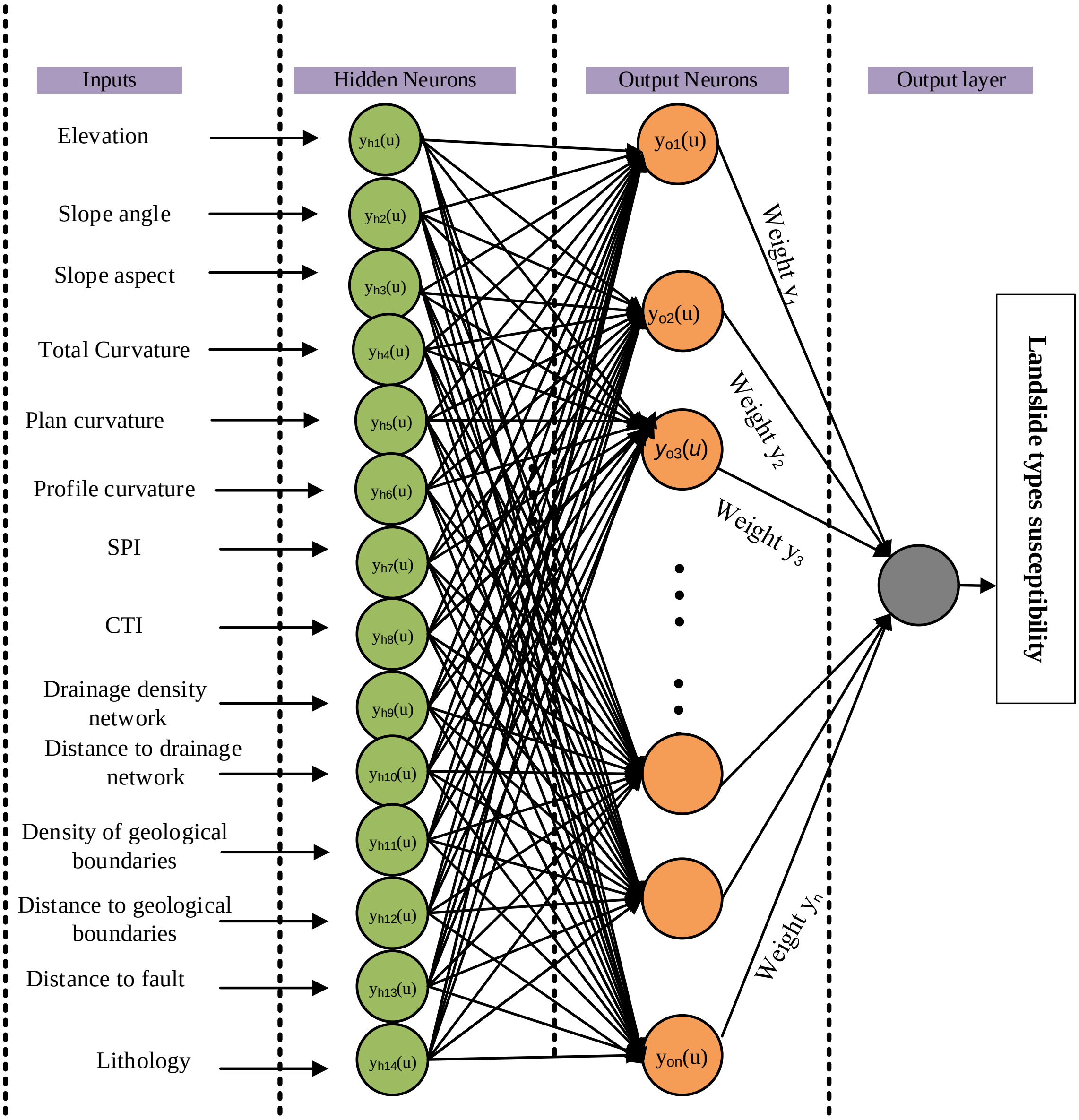

Figure 5-3 Structure of a typical three-layer feed forward BPNN (multi-layer perception) for landslide types susceptibility analysis.

The initial weights were automatically assigned to the random values between 0.1 and 0.25. The parameters were adjusted as follows: 1) initial learning rate (*ILR*) for influencing the convergence of the network: 0.1, 2) number of epochs: 1500 iterations, 3) momentum parameters: 0.9 (to prevent instabilities caused by an excessively high *ILR* value), 4) activation (transfer) function for layers: transig for hidden layer, purelin for output layer, and 5) training function of networks: variable *ILR* with momentum (traingdx). The value of *RMSE* for terminating the criterion was set to 0.001.

5.1.3 Resultant landslide classification using SVM

The performance of the SVMs model is directly related to the selection of the kernel function and parameters. Each kernel was trained and tested, and the results of the prediction of the two landslide types from the 10 factors for each kernel (Table 5-2) show that the accuracy of the training samples of PF is the highest (95.26%) followed by that of RBF (94.38%); however, RBF outperformed for the testing samples (84.31%) and was hence selected as the SVMs kernel for this study. This classifier selection may be a limitation of SVMs, because only one constraint is active at a time (Burges 1998; Kavzoglu et al. 2013a).

Average accuracies of the backbone propagation (BP) technique using artificial neural network (ANN) applied to the same dataset, from an unpublished work from the same authors, show that for models trained with 50% of the data, the average training and testing accuracies obtained from SVMs (89.24% and 77.78%) are higher than those from BP (Table 5-4). The low standard deviation values across the iterations (< 5%) suggest the stability of the method. The results (not detailed here) show that deep-seated landslides were classified more accurately (88.18%) than shallow landslides (76.99%), as visually represented in Figure 5-4. This indicates the strong morphological signatures cast by deep-seated landslides while the imprints of shallow landslides were not correctly captured by the DEM used (Korup 2005b; May 2007). Although we used the 2 m DEM, the use of higher resolution topographical information seems to be necessary to study shallow landslides in detail.

Table 5-4 Accuracy of the SVMs and BP model with the data equally (50%) divided into training and testing samples.

	Accuracy (%) of SVMs		Accuracy (%) of BP	
Iteration number	Training (%)	Testing (%)	Training (%)	Testing (%)
1	90.46	75.87	87.24	69.21
2	85.59	79.38	86.09	62.10
3	86.38	73.74	79.31	65.13
4	88.00	74.22	84.41	59.23
5	93.82	76.87	81.64	65.12
6	94.38	77.91	83.16	61.02
7	84.15	77.39	80.73	58.39
8	85.38	78.09	81.19	62.34
9	88.00	80.04	82.36	61.04
10	96.27	84.31	78.25	58.37
Min	**84.15**	**73.74**	**78.25**	**58.37**
Max	**96.27**	**84.31**	**87.24**	**69.21**
Standard deviation	**4.04**	**2.89**	**2.71**	**3.27**
Average	**89.24**	**77.78**	**82.44**	**62.20**

For the reduced size of the training dataset (30%, 40% and 50%), the model performed equally well (Table 5-5 and Figure 5-5). The models trained with 30% and 40% of the data yielded the average of overall accuracy of 75.1% and 75.24%, with the standard deviation of 2.93 and 2.43, respectively, which are very similar to the results obtained from the 50% training data. This agrees with Burges (1998), Chi et al., (2008), Huang et al., (2002) and Kavzoglu et al., (2013b) in terms of the stability of SVMs even with fewer training datasets, compared to other models like ANN.

Figure 5-6 is the final map showing the probable landslide types (shallow and deep-seated) and non-landslide areas in the case of an earthquake of a magnitude similar to the Chuetsu earthquake, on the basis of the SVMs and the causative factors. The prediction map has an overall accuracy of 71.75%, which seems to be acceptable as the first trial of this kind in the study area, and may provide a guideline for social preparation for future landslide hazards. From the figure, we can also visualize that most of known-shallow and deep-seated landslides are located in the corresponding probable zones. In the same prediction map, it should also be noted that the area with the probable deep-seated landslides is broader than that with probable shallow landslides. This result agrees with the landslide inventory for the Chuetsu earthquake, including deep-seated landslides fewer in number (330, compared with 895 shallow landslides) but much larger in average area (9600 m^2 compared to 187 m^2 for shallow landslides).

Table 5-5 Accuracy of the SVMs model using 30%, 40%, and 50% of the data to train the model.

Iteration number	Prediction accuracy (%) with the testing samples		
	Model trained with 50% data	Model trained with 40% data	Model trained with 30% data
1	75.87	78.60	74.70
2	79.38	75.10	80.80
3	73.74	73.52	73.74
4	74.22	74.10	77.40
5	76.87	72.31	79.50
6	77.91	80.01	72.30
7	77.39	76.80	72.62
8	78.09	75.13	72.91
9	80.04	74.80	72.30
10	84.31	72.10	74.83
Min	**73.74**	**72.10**	**72.30**
Max	**84.31**	**80.00**	**80.80**
Standard deviation	**2.89**	**2.43**	**2.93**
Average	**77.78**	**75.24**	**75.10**

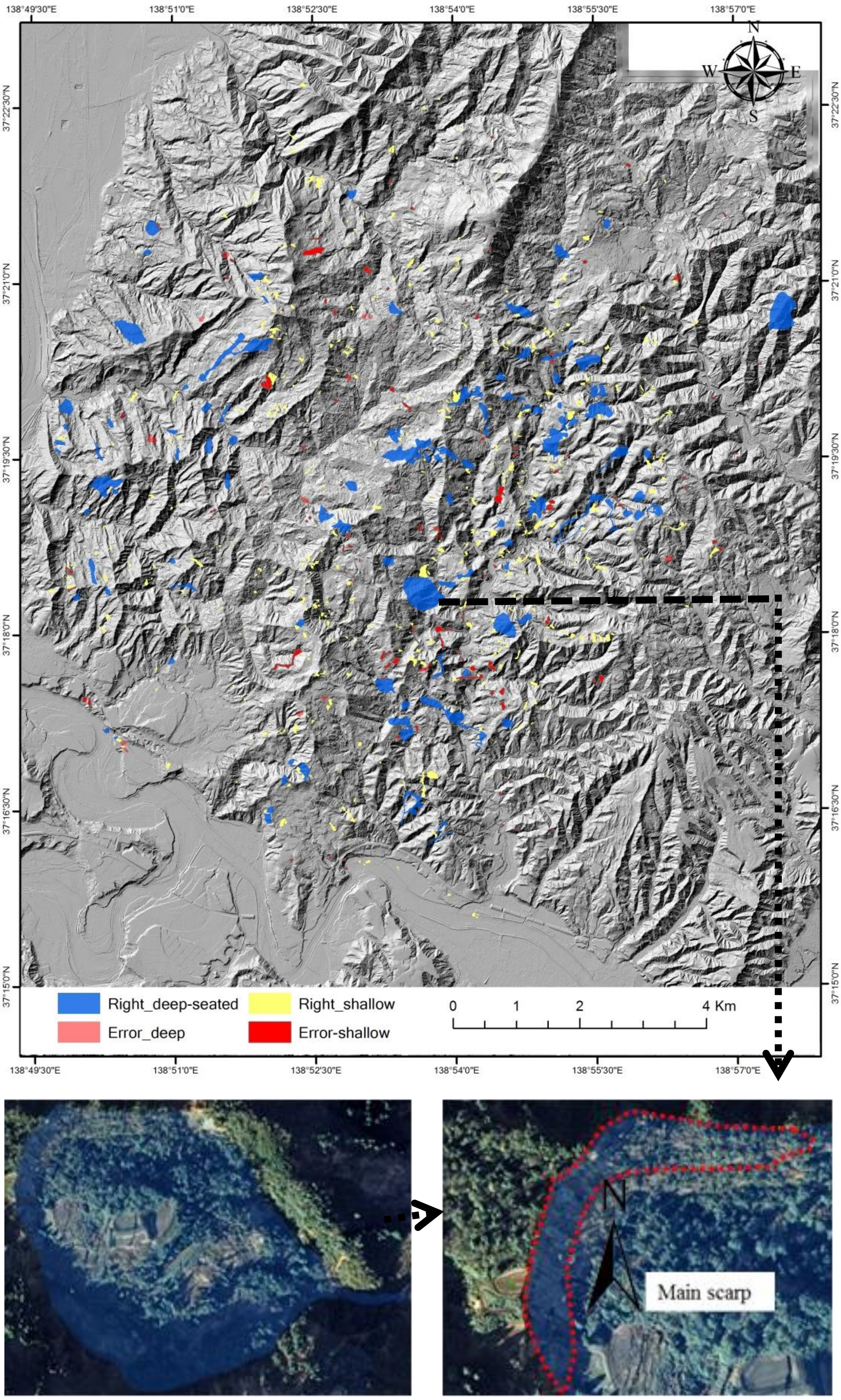

Figure 5-4 Upper: map prepared using the confusion matrix obtained from classification of landslide types using SVM with 50% trainings sample, lower: representational aerial photographs.

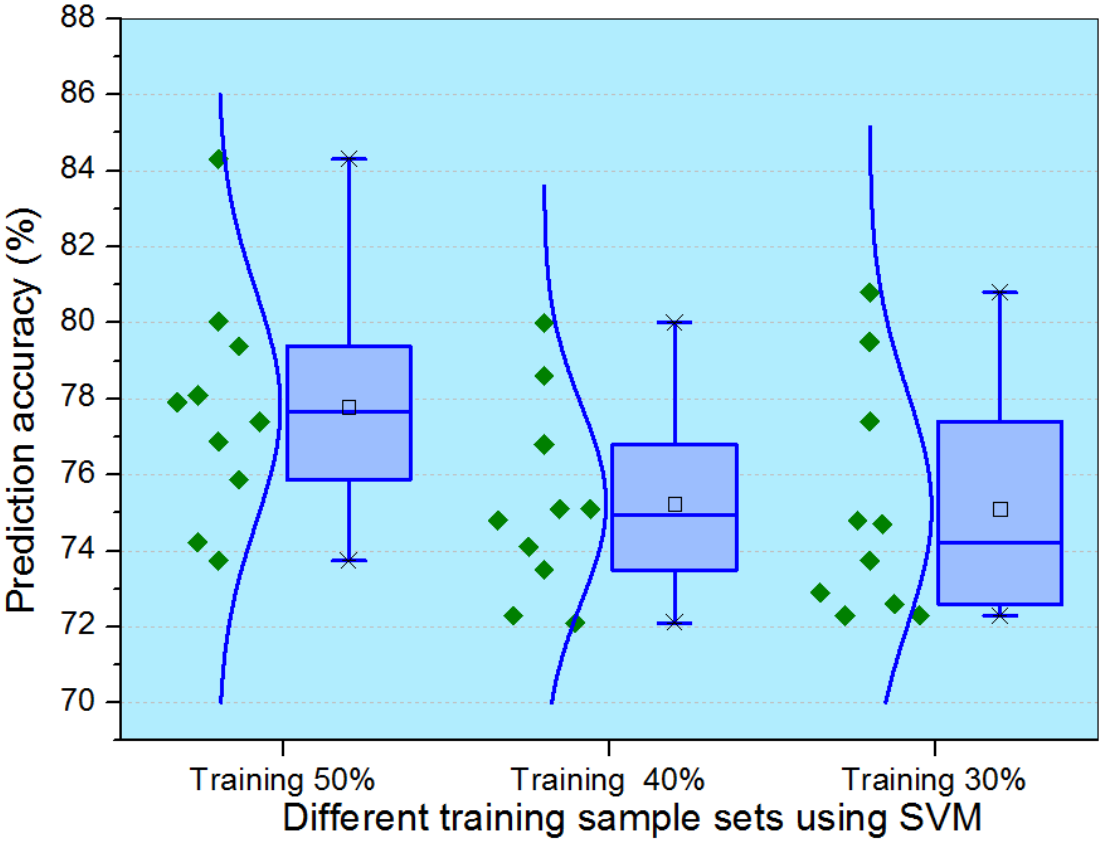

Figure 5-5 Prediction accuracy (testing samples) of the SVM model using 30%, 40% and 50% of the data to train the model.

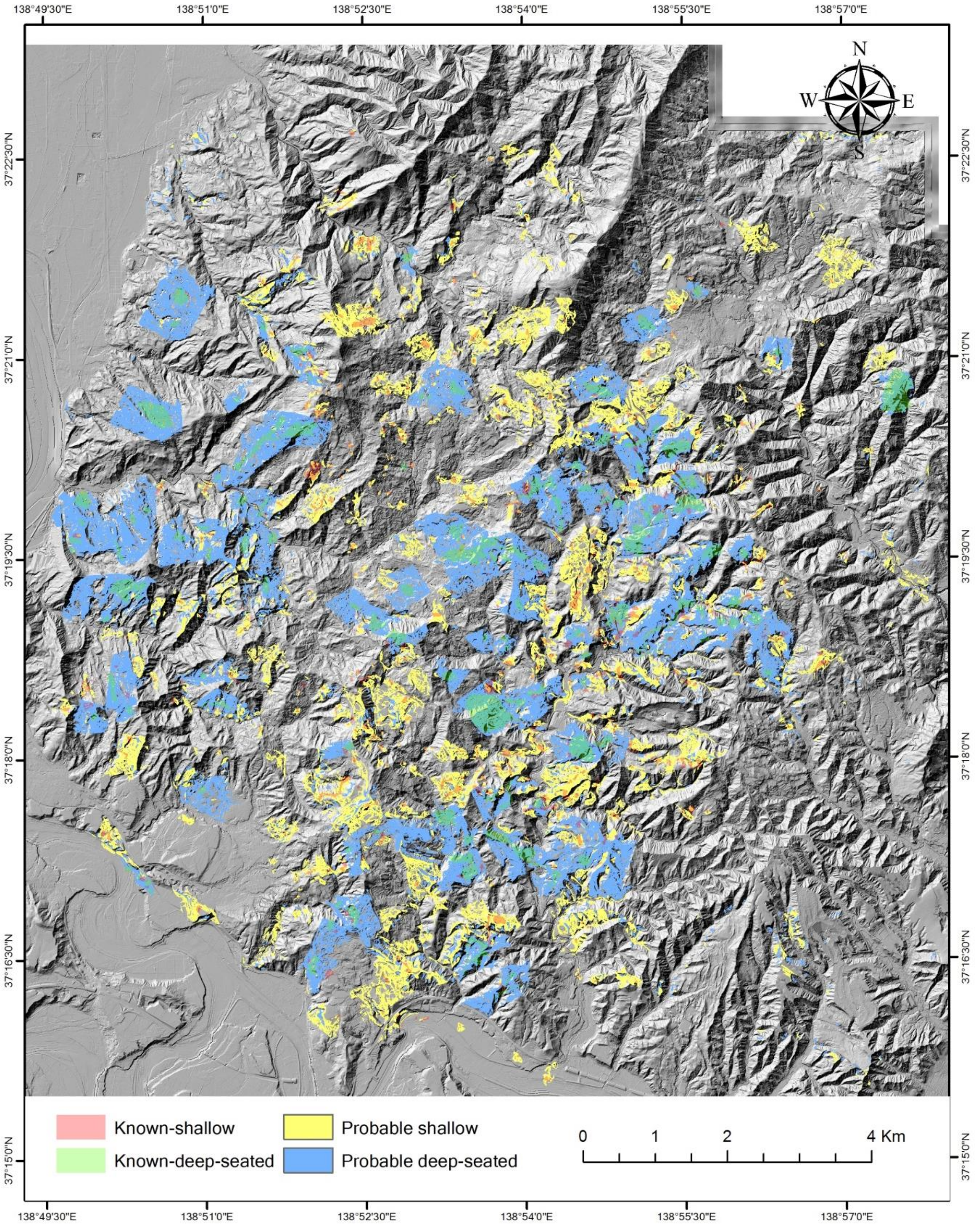

Figure 5-6 Map showing the probable occurrences of shallow and deep-seated landslides in the whole study area.

5.2. Summary

This study assumed that variations in topographic and geographic factors used for evaluating landslide susceptibility, are equally useful in predicting and differentiating shallow and deep-seated landslides. An inventory of landslides triggered by the M 6.8 Chuetsu earthquake and subsequent aftershocks in 2004, a 2 m Lidar DEM, geological data, and SVMs were used. The results with high accuracy suggest that our assumption is valid. The existing landslides matched the predictions in most cases. SVMs also outperformed ANN (BP) in terms of model stability and accuracy. Among the four SVMs kernels, RBF was selected after a comparative test. Moreover, reduction in the size of the training dataset from 50% to 30% of the total dataset did not significantly affect the accuracy of the SVMs model, confirming that SVMs work even with a smaller training dataset. However, we found that a higher resolution DEM is necessary for studying the details of shallow landslides.

Active geological processes like landslides play an important role in reshaping topography. Therefore, differentiating the types of landslides is important for discussing the geomorphological evolution of hillslopes, and also for supporting the local government managing and mitigating local hazards. Further studies using not only a finer DEM but also other detailed information such as the peak ground acceleration (PGA) and volume of landslides are necessary.

CHAPTER 6 LANDSLIDE SUSCEPTIBILITY MAPPING (LSM)

LSM plays a vital role in assisting and managing hazards for land use planning and risk mitigation (Wu et al. 2008; Yalcin et al. 2011; Tofani et al. 2014; Dou et al. 2015c). LSM provide information on the likelihood of landslides occurring in an area given the local terrain conditions (Brabb 1984). Using GIS, various methods for landslide susceptibility mapping have been proposed in the past. These methods can be grouped into qualitative and quantitative, based on the properties they involve (Felicísimo et al. 2012; Peng et al. 2014). Qualitative methods denote susceptibility levels in descriptive terms using expert knowledge (Conoscenti et al. 2015). Such techniques are relatively subjective and were extensively used during the 1970s and 1980s (Aleotti and Chowdhury 1999; Yilmaz et al. 2011). A main limitation of qualitative method is that the accuracy depends on the knowledge of the experts who conducts the research. Quantitative methods, on the other hand investigates the relationship between landslides and causative factors to predict the occurrence probabilities (Neuhäuser et al. 2011; Anbazhagan and Ramesh 2014). Compared to the former one, a more realistic susceptibility map can be obtained from statistical and numerical methods (Yalcin et al. 2011) since they reduce the subjectivity and biases in the process of weighting landslide causative factors.

A wide range of quantitative methods have been successfully used for landslide susceptibility mapping by researchers around the globe (Yalcin et al. 2011; Dou et al. 2014a; Dou et al. 2015g). The widely used methods are bivariate, multivariate (Magliulo et al. 2008; Yalcin et al. 2011), and logistic regression (LR) (Conoscenti et al. 2015), neuro-fuzzy (Tien Bui et al. 2012b; Pourghasemi et al. 2013a), support vector machines (Tien Bui et al. 2012b; Pradhan 2013; Dou et al. 2015e), and probabilistic models using Monte Carlo simulation with GIS (Wang et al. 2008; Komac 2012). The bivariate and multivariate statistical methods estimate landslide probabilities based on correlation analysis between causative factors and historical landslide events, whereas the deterministic methods assess slope failures using the factor of safety (*FoS*) (Bahsan et al. 2014; Jamsawang et al. 2015). In literature, statistical index (*SI*) and LR are considered to be the most commonly used methods for the assessment of probability of occurrence of landslides at medium and regional scales (Shahabi et al. 2013; Meinhardt et al. 2015). In contrast, *FoS* is used widely for the landslide assessment at local scales (Pradhan and Lee 2010; Felicísimo et al. 2012; Dou et al. 2014a). The advantage of LR over other multivariate analysis methods is that it is independent of data distribution and can handle a variety of data sets such as continuous, categorical, and binary data (Tien Bui et al. 2011; Yalcin et al. 2011). However, if a set of irrelevant independent variables are included, the LR model may have little to no predictive value. Owing to such constraints, prediction of landslide susceptibility requires a distributed approach that identifies all the relevant independent aspects of models used. In addition to that, successful landslide susceptibility mapping require optimal causative factors as input to the LSM models. In LSM studies, causative factors are usually selected based on the analysis of the landslide types and the characteristics of the study area (Ayalew et al. 2005a). Commonly used causative factors are elevation, slope angle, slope aspect, plan curvature, and distance to drainage networks (Dou et al. 2015g). However, most researchers arbitrarily and subjectively selected the causative factors such as geological, geomorphological, hydrological and anthropogenic factors. Hence, selection of landslide causative factors and their classes are key points in LSM (Costanzo et al. 2012; Meinhardt et al. 2015).

6.1. Feature selection using a CF model

Landslide causative factors and their classes play a crucial role in LSM (Costanzo et al. 2012; Meinhardt et al. 2015). Landslide causative factors are usually selected based on the analysis of landslide types and characteristics of the study area (Ayalew et al. 2005a). However, most scholars randomly and subjectively selected the causative factors such as geological, geomorphological, hydrological and anthropogenic factors to produce the landslide susceptibility maps. Lee and Talib (2005) noted that the selection of positive factors can improve the prediction accuracy of LSM. This indicates that the optimized factors are significant to LSM. Thus, before building a susceptibility model, predictive abilities of the initial selected factors should be quantified and factors with very low or null predictivity should be removed. This helps to reduce noise and uncertainties and thus the prediction ability of the resulting models will improve (Martínez-Álvarez et al. 2013). For instance, Pradhan and Lee (2010) removed the causative factors with small weights down to four, seven and eleven factors. Their research concluded that seven factors gave the best predicting accuracy. However, it is difficult to decide the threshold of weight to select factors. Although Lee and Talib (2005) used factor analysis to remove the correlated variables which is a time-consuming method. Jebur et al. (2014) followed an optimal technique for detecting best landslide causative factors, and their methods rather requires preparation of two sets of causative factors. Although various other techniques have been proposed such as the linear correlation (Irigaray et al. 2006), Goodman–Kruskal’s gamma (Costanzo et al. 2012), and GIS matrix combination (Cross 1998), no standard guideline is available. As herein, we address this issue by proposing the CF method that has rarely been used for feature selection in landslide studies (Binaghi et al. 1998). CF is an approach using rule-based expert systems to resolve certain problem classes. In the past, the search for the probabilistic interpretation of a CF model has been attracted considerable attention (Binaghi et al. 1998; Lucas 2001; Devkota et al. 2013; Dou et al. 2014b).

In our study, CF is applied for selecting the positive causative factors related to landslide occurrence. Compared with the other methods, CF can be relatively easy to perform when different layers need to be integrated using the combination rule (Binaghi et al. 1998; Devkota et al. 2013; Dou et al. 2014b).

6.2. Methods used for LSM

Figure 6-1 is an overview of the approach that was applied for the landslide susceptibility mapping in the study areas. The flowchart consists of three phases: (i) data preparation and extraction of the landslide causative factors; (ii) selection of the best subset of the causative factors using the CF method; and (iii) landslide susceptibility mapping using the SI and LR method, (iv) model validation and comparison. Each method is presented in the following sections.

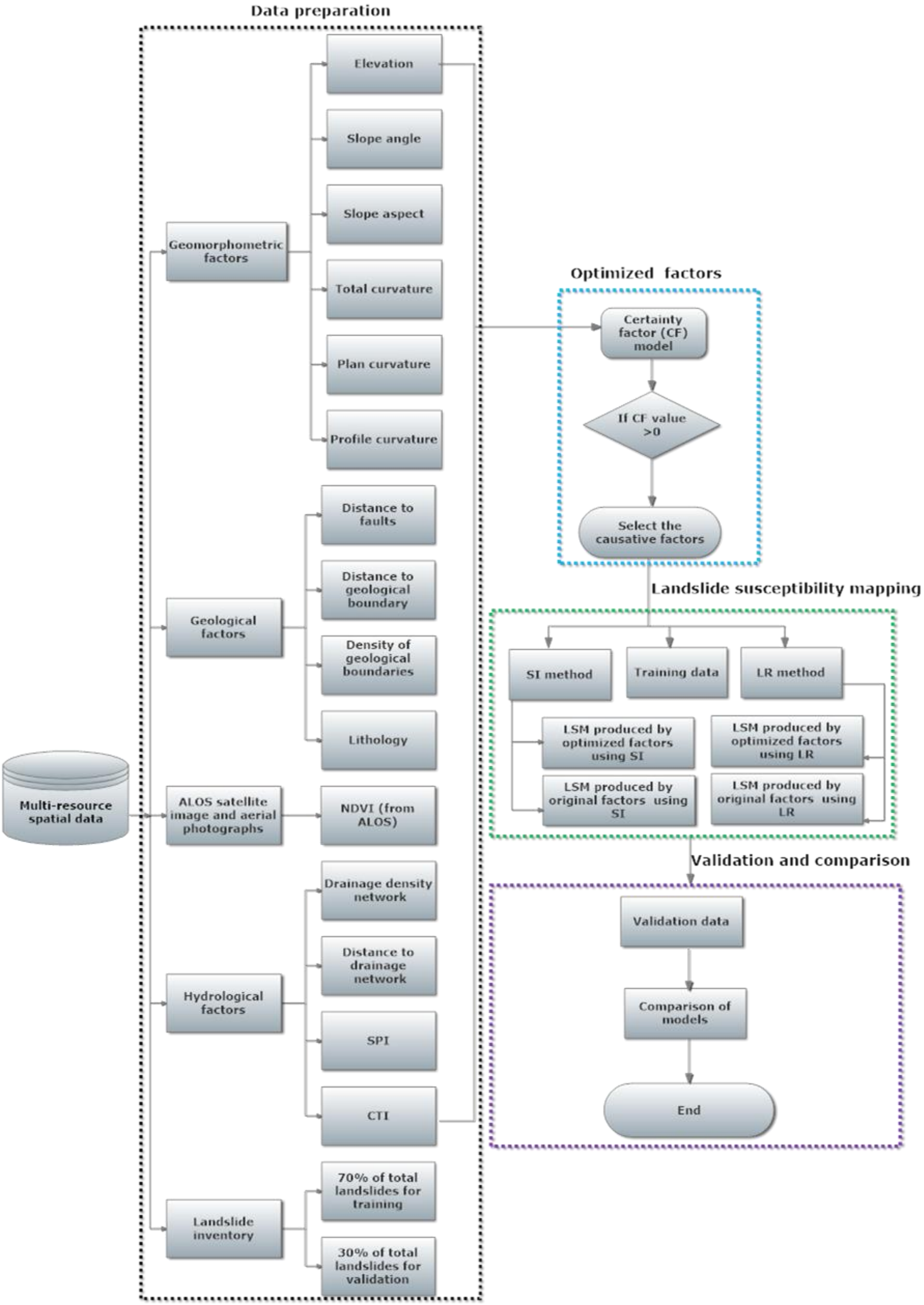

Figure 6-1 Flowchart showing overall methodology adopted for this study.

6.2.1 CF model

CF is a rule based expert system method developed by Shortliffe and Buchanan (1975) for the management of uncertainty in computational studies. CF provides probable favorability functions (FF) for integrating heterogeneous data (Chung and Fabbri 1993). The CF index can be calculated using the following functions:

$$CF = \begin{cases} \dfrac{PP_i - PP_s}{PP_i(1 - PP_s)} & \text{if } PP_i \geq PP_s \\ \dfrac{PP_i - PP_s}{PP_s(1 - PP_i)} & \text{if } PP_i < PP_s \end{cases} \tag{6-1}$$

where PP_i is the conditional probability of landslides in class i and PP_s is the prior probability of total number of landslides in the study area.

The *CF* values range between -1 and 1, and it indicates a measure of belief and disbelief (Lucas 2001). A positive value measures decreasing uncertainty whereas negative values imply an increasing uncertainty of landslide occurrence. If *CF* equals 0, no information on the certainty is indicated. Once the *CF* values for classes of the causative factors are obtained, these factors are then incorporated pairwise using the combination rule (Binaghi et al. 1998) as follows:

$$Z = \begin{cases} CF1 + CF2 - CF1\,CF2 & CF1, CF2 \geq 0 \\ CF1 + CF2 + CF1\,CF2 & CF1, CF2 < 0 \\ \dfrac{CF1 + CF2}{1 - \min(|CF1|,\ |CF2|)} & CF1, CF2\text{, opposite signs} \end{cases} \tag{6-2}$$

where $CF1$ is a value in class 1, and $CF2$ is a value in class 2.

The pairwise combination is carried out until all the *CF* layers are brought together, and the causative factors are optimized by computing the *Z* values. If the *Z* values are positive, we regard those factors have strong relationships with landslide occurrence

Based on the range of *CF* values, feature weights were obtained. The weights are estimated as the sum of the ratio computed relative causative factors that provides a measurement of certainty in forecasting the landslides (Binaghi et al. 1998). Based on the results, CF weights were then categorized into six classes as shown in Table 6-1.

Table 6-1 CF weights classification according to the range of CF values

Code	Range	Description
1	−1.0 - −0.09	Extremely low certainty
2	−0.09 - 0.09	Uncertainty
3	0.09 - 0.2	Low certainty
4	0.2 - 0.5	Medium certainty
5	0.5 - 0.8	High certainty
6	0.8 - 1.0	Extremely high certainty

6.2.2 Statistical index method

The *SI* method proposed by van Westen et al. (1997) is based on the assessment of correlation of a landslide inventory map and causative factors. In SI models, the weight for each class of the landslide causative factors was firstly determined. Landslide susceptibility indexes were then obtained by summing up the weights.

The weight (Wi) of each class i is defined as the natural logarithm of the landslide density in the class over the landslide density in the factor map as follows (van Westen et al. 1997) :

$$W_i = \ln\left(\frac{DensClass}{DensMap}\right) = \ln\left(\frac{Npix(S_i)/Npix(N_i)}{\sum Npix(S_i)/\sum Npix(N_i)}\right) \qquad (6\text{-}3)$$

where W_i is the weight given to a certain parameter class; *DensClass* is the landslide density within the parameter class; *DensMap* is the landslide density of the entire factor map for all classes; $Npix(S_i)$ is the number of landslide pixels in the *i* class; and $Npix(N_i)$ is the total number of pixels in all classes.

6.2.2 Binary LR

Binary LR is one of the most frequently used multivariate analytical methods for creating landslide susceptibility maps. The LR approach is useful to predict the presence or absence of a characteristic outcome from a set of predictor variables (Devkota et al. 2013; Conoscenti et al. 2015). Here, we do not use the ordinary least squares regression (OLS) because of three problems: 1) the error terms are heteroskedastic; 2) the error terms are not normally distributed; 3) the predicted probabilities can be larger than 1 or less than 0. In this study, the purpose of LR is thus to simulate the relationships between a dependent variable and multiple independent parameters (Tien Bui et al. 2011). The merit of LR is that it does not compulsorily require a normal distribution data. Additionally, both continuous and discrete data types can be used as an input for the LR model.

The dependent variable (Y) in the LR method is a function of the probability and can be computed as follows (Lee and Pradhan 2006):

$$P(Y=1|x)=\frac{\exp(\sum bx)}{1+\exp(\sum bx)}x \tag{6-4}$$

where P is the estimated probability of landslide occurrence and ranges from 0 to 1; and

X is the independent variables (landslide causative factors), $X=(x_0,x_1,x_2,...x_n)$, $x_0=1$;

and b is regression coefficient.

To linearize the mentioned method as well as remove the 0/1 boundaries for the original dependent variable, the estimated P is transformed by the following formula:

$$P^{'}=\ln(\frac{P}{1-P}) \tag{6-5}$$

This alteration is referred to as the logit transformation. Theoretically, the logit transformation of binary data can ensure that the dependent variable is continuous and the logit transformation is boundless. Moreover, it can ensure that the probability surface will be continuous within the range [0, 1]. Using the logit transformations, the standard linear regression models can be obtained as follows:

$$P' = \ln(\frac{P}{1-P}) = \mathrm{b}_0 + \mathrm{b}_1\mathrm{x}_1 + \mathrm{b}_2\mathrm{x}_2 + ... + \mathrm{b}_\mathrm{n}\mathrm{x}_\mathrm{n} + \varepsilon \tag{6-6}$$

where, b_0 is the constant or intercept of the formula, $\mathrm{b}_1, \mathrm{b}_2, ...\mathrm{b}_\mathrm{n}$ represents the slope coefficients of the independent parameters, $x_1, x_2, ...x_n$ in the logistic regression and ε is standard error.

The LR model mainly involves five steps in generating LSM models: 1) pre-selection of parameters based on the analysis of the spatial distribution; 2) selection of statistically significant parameters via a *p*-value significance test; 3) significance test to the LR model with these parameters (via the goodness of fit by inputting a parameter or eliminating a parameter); 4) evaluation of the multi-collinearity among the parameters (diagnosis via two indicators, namely, tolerance < 0.1 and variance inflation factor > 5); and 5) assessment of the accuracy of the model.

6.3. LSM for Chuetsu area, Japan

6.3.1 Feature selection

The results of the correlation analysis between the landslide occurrence and causative factors are shown in Table 6-2. Concerning the CF analysis, the *Z* value is positive for slope angle (0.54), slope aspect (0.03), drainage density (0.14), plan curvature (0.17) density of geology boundary (2.89), and lithology (0.3001) (Figure 6-2). These factors have positive relationships with the landslide occurrence. The *Z* value is negative for the other factors. Hence, the six factors are selected for LSM.

A detailed analysis shows that slope angle has the highest influence on slope stability. *CF* values are positive at slopes from 18°–45° (Table 6-2). The percentage of landslide occurrence at the slope class 18 °–24 °, 24 °–30 °, 30 °–35 °, and 35 °–45 ° are 12.64%, 15.57%, 18.36% and 32.87%, respectively. The results indicate that the landslide occurrence increases with an increasing slope angle up to 45 °, and then decreases. Gentler slopes have a relatively low frequency of landslide occurrences because of the lower shear stress corresponding to the low gradient; whereas, very steep slope angles lead to outcropped bedrock, which is less susceptible to landslides.

In the case of slope aspect, landslides mostly occurred along southeast, south, southwest and west facing slopes with positive *CF* values from 0.07 to 0.78. The highest percentage of landslides with the maximum *CF* value (0.78), 23.69% occurred along the southwestern slopes, followed by the south slopes (20.72%). Landslides triggered by an earthquake are distributed mainly along dip direction of geological formations. The northern slope aspect with a negative *CF* value may be related to dipping of bedrock.

The drainage density shows positive *CF* values for the classes 2–4, 4–6, and 6–10. The positive maximum *CF* value of 0.3 is observed with the 4–6 drainage density class. The highest percentage of landslide occurrences is 25.55%. Since the landslides in this study were triggered by an earthquake and the corresponding *CF* values are comparatively very small, a specific cause-reason relation for the distribution of landslides in the drainage density class could not be established.

In the case of density of geological boundary, the *CF* values are always positive. The maximum *CF* value, 0.83, is seen in the class with the most dense geological boundary (>-27) followed by a value of 0.58 in the immediately lower geological density class (20–27). The negative CF values for the geological boundary density classes lower than 7 indicate that geological uniformity affects the stability of the area. A higher density of geological boundaries suggests lower stability, which may lead to landslide occurrence.

Concave plan curvature corresponds to a negative *CF* value (-0.16) while convex curvature to a positive value (0.31). Normally convex areas have a lower *CF* value than concave areas because such slopes retain more water and the increased soil moisture content reduces the stability of the soil. In this study area though, the concavity is not responsible for the landslide occurrences, because the landslides are induced by an earthquake, not by rainfall. Ridges tended to collapse because of higher ground acceleration due to the earthquake.

The results of this study also indicate that the lithology1 and lithology2 classes are positive. These two lithology mostly consist of sandstone and massive mudstone. Their significantly higher *CF* values are 0.75, 0.31, respectively. The highest percentage of landslides among the lithology classes, 93.3%, occurred in lithology1 followed by lithology2 (4.15%). The bedrock in the area of major landsliding consists of a folded sequence of sandstone, mudstone and their interbeddings, and the results point to the occurrence of landslides in the weakly cemented lithological groups.

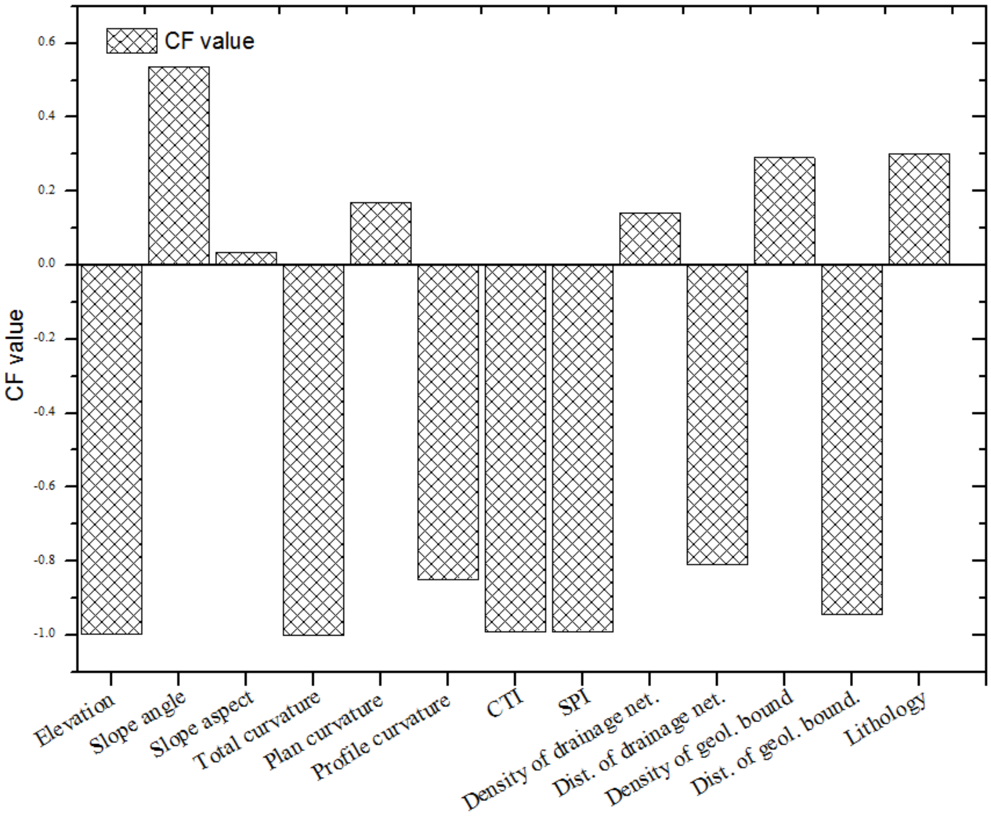

Figure 6-2 Calculation of CF value in the Chuetsu study area

Table 6-2 Spatial relationship between the causative factors and landslide occurrence by the CF method and SI method.

Causative factors	Class	Percentage of domain	No. pixel of landslides	Percentage of landslides	CF	Z	SI
Elevation (m)	<100	28.28	239	3.23	-0.92	-1.00	-2.17
	100-200	25.72	2165	29.24	0.17		0.13
	20-300	23.61	2826	38.16	0.55		0.48
	300-400	14.55	1703	23.00	0.52		0.46
	400-500	6.01	437	5.90	-0.03		-0.02
	500-600	1.48	32	0.43	-0.78		-1.23
	600-700	0.30	2	0.03	-0.94		-2.42
	>700	0.05	1	0.01	-0.78		-1.23
Slope angle (degree)	0-5	29.27	72	0.97	-0.98	0.54	-3.40
	5-10	11.88	314	4.24	-0.72		-1.03
	10-18	12.54	710	9.59	-0.31		-0.27
	18-24	12.49	936	12.64	0.02		0.01
	24-30	9.25	1153	15.57	0.58		0.52
	30-35	8.31	1360	18.37	0.78		0.79
	35-45	10.39	2434	32.87	0.98		1.15
	>45	5.86	426	5.75	-0.03		-0.02
Slope aspect	North	11.90	216	2.92	-0.78	0.03	-1.41
	Northeast	11.63	463	6.25	-0.46		-0.62
	East	14.42	787	10.63	-0.20		-0.30
	Southeast	11.84	1344	18.15	0.58		0.43
	South	12.34	1534	20.72	0.65		0.52
	Southwest	11.37	1754	23.69	0.79		0.73
	West	14.77	1011	13.66	0.07		-0.08
	Northwest	11.74	294	3.97	-0.68		-1.08
Drainage density	<2	51.59	3204	43.27	-0.22	0.14	-0.18
	2-4	13.59	1056	14.26	0.06		0.05
	4-6	19.94	1892	25.55	0.30		0.25
	6-10	11.13	1036	13.99	0.28		0.23
	>10	3.76	217	2.93	-0.29		-0.25
Distance from the drainage network	<50	28.66	2674	36.11	0.28	-0.81	0.23
	50-100	23.64	1902	25.69	0.10		0.08
	100-175	21.29	1372	18.53	-0.19		-0.14
	175-250	15.20	904	12.21	-0.27		-0.22
	250-330	8.68	480	6.48	-0.33		-0.29
	300-700	2.54	73	0.99	-0.70		-0.95
	700-799	0.00	0	0.00	-1.00		$-\infty$
	>800	2.54	73	0.99	-0.70		-0.95

Table 6-2 Continued

Causative factors	Class	Percentage of domain	No. pixel of landslides	Percentage of landslides	CF	Z	SI
SPI	<-8	6.51	38	0.51	-0.94	-0.99	-2.54
	-8--4	12.21	648	8.75	-0.36		-0.33
	-4-0	35.98	1284	17.34	-0.61		-0.73
	0-4	42.13	5109	68.99	0.56		0.49
	4-8	2.79	312	4.21	0.48		0.41
	8-12	0.33	13	0.18	-0.55		-0.62
	12-16	0.04	1	0.01	-0.75		-1.13
CTI	<-4	9.29	647	8.74	-0.08	-0.99	-0.06
	-4-0	11.00	605	8.17	-0.33		-0.30
	0-4	62.43	5715	77.18	0.27		0.21
	4-8	15.28	397	5.36	-0.73		-1.05
	8-12	1.81	39	0.53	-0.78		-1.24
	12-16	0.18	2	0.03	-0.89		-1.88
	16-24	0.02	0	0.00	-1.00		-∞
Curvature	<-224	0.01	0	0.00	-1.00	-1.00	-∞
	-224-0	63.62	4555	61.51	-0.05		-0.03
	0-224	36.37	2850	38.49	0.08		0.06
	224-448	0.01	0	0.00	-1.00		-∞
	>448	0.00	0	0.00	-1.00		-∞
Profile curvature	<-350	0.00	0	0.00	-1.00	-1.00	-∞
	-350--220	0.00	0	0.00	-1.00		-∞
	-220-0	33.30	2679	36.18	0.11		0.08
	0-45	66.52	4713	63.65	-0.06		-0.04
	45-175	0.17	13	0.18	0.05		0.04
	>175	0.01	0	0.00	-1.00		-∞
Plan curvature	Concave	69.74	4546	61.39	-0.16	0.17	-0.13
	Convex	30.26	2859	38.61	0.31		0.24
Distance from the geologic boundary	<35	36.72	3269	44.15	0.23	-0.95	0.18
	35-130	40.24	2888	39.00	-0.05		-0.03
	130-320	18.30	1066	14.40	-0.29		-0.24
	320-600	4.29	176	2.38	-0.54		-0.59
	>600	0.45	6	0.08	-0.87		-1.71
Density of geology boundary	<7	63.20	4128	55.75	-0.17	2.89	-0.13
	7-14	28.83	2472	33.38	0.18		0.15
	14-20	7.06	675	9.12	0.31		0.26
	20-27	0.68	86	1.16	0.59		0.54
	>27	0.24	44	0.59	0.84		0.89
Lithology	1	13.39	189	2.55	-0.72	0.30	-1.66

2	5.99	307	4.15	0.31	-0.37
3	80.62	6909	93.30	0.76	0.15

6.3.2 LSM

The relationship between the landslide occurrence and causative factors using SI is represented in Table 6-2. Two landslide susceptibility maps were generated: (i) using the six selected factors (*CF* > 0) and (ii) using originally selected thirteen factors. The results that indicate the spatial probability of landslide occurrence is shown in Figure 6-3. Based on the natural breaks inherent in the data, the susceptible level is eventually divided into six classes; i.e., extremely low, low, moderate, high, very high and extremely high. It can be noticed from the visual observation that there are much more red color areas in Figure 6-3b, whereas there are more dark blue areas in Figure 6-3a. Figure 6-4 and Table 6-3 shows that 89.18% of the total landslides occurred in the 55.34% of the area classified as high, very high and extremely high susceptibilities when the original 13 factors were used, while 92.7% of the total landslides occurred in the 50.37% of the area classified as high, very high and extremely high susceptibilities if the optimized six factors were used (Figure 6-5 and Table 6-4).

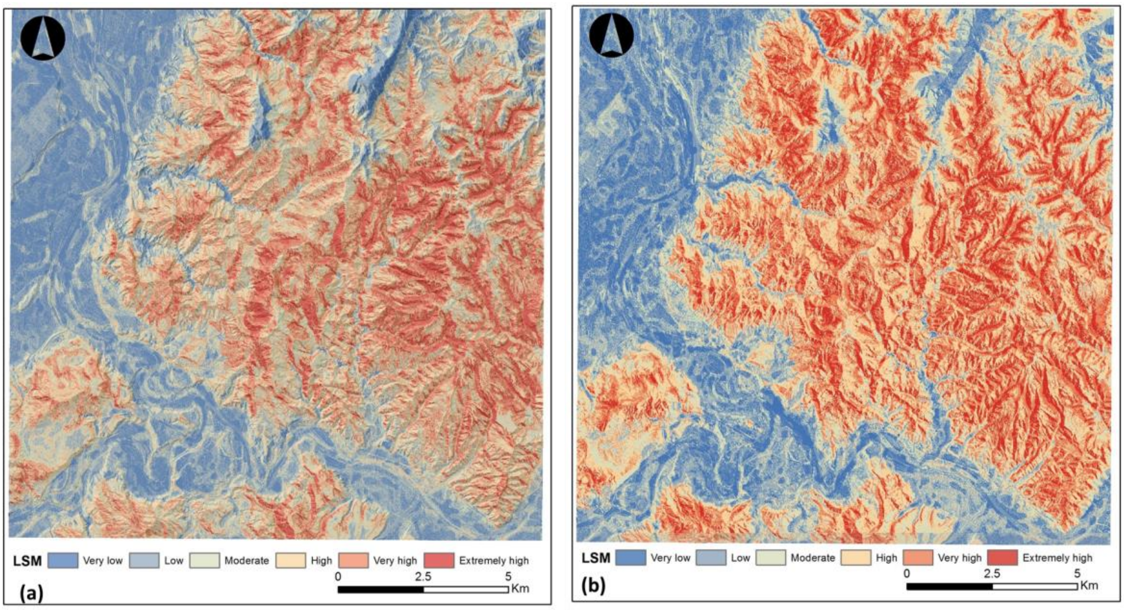

Figure 6-3 LSM maps produced by the SI method: a) selected six factors, and b) original 13 factors. Maps show the spatial probability of landslide occurrence in six classes.

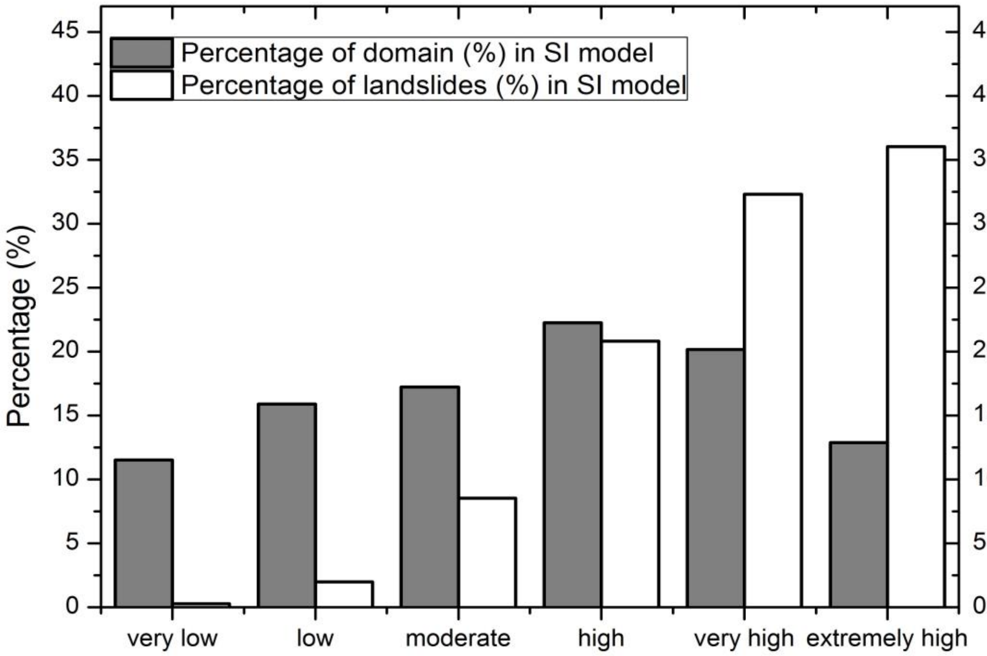

Figure 6-4 Susceptibility class distribution within the study area and the occurrence of landslides according to the classification scheme for LSM using the SI method with the original 13 factors.

Table 6-3 Result of statistical analysis concerning landslide susceptibility from the SI method with the original 13 factors.

Class	Area of each class	Percentage of domain （%）	No. of landslides	Percentage of landslides (%)
Very low	284530	11.53	20	0.27
Low	392251	15.9	148	2
Moderate	425078	17.23	633	8.55
High	549629	22.28	1543	20.84
Very high	497791	20.17	2393	32.32
Extremely high	318130	12.89	2668	36.03

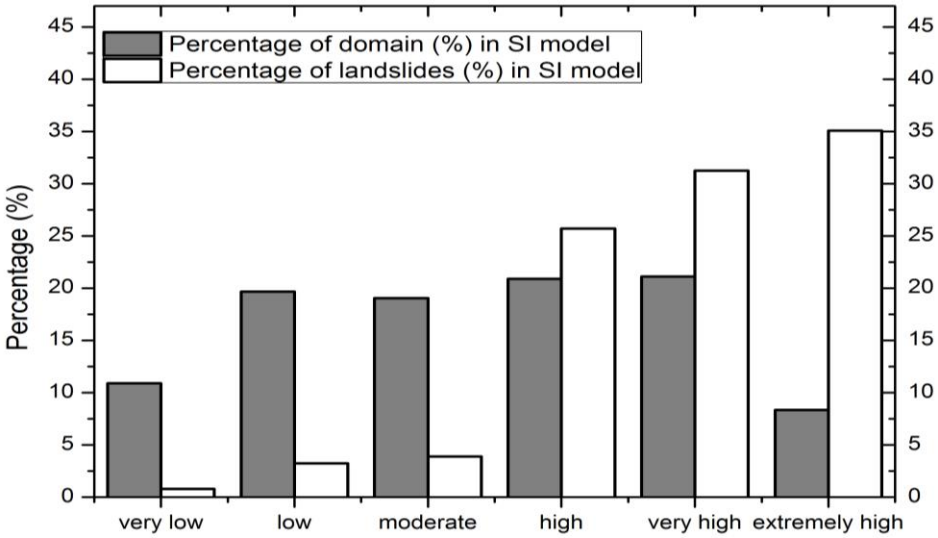

Figure 6-5 Susceptibility class distribution within the study area and the occurrence of landslides according to the classification scheme for LSM using the SI method with the selected six factors.

Table 6-4 Result of statistics analysis concerning landslide susceptibility from the SI method with the selected six factors.

Class	Area of each class	Percentage of domain （%）	No. of landslides	Percentage of landslides (%)
Very low	269170	10.91	59	0.80
Low	485549	19.68	239	3.23
Moderate	469764	19.04	289	3.90
High	516032	20.91	1905	25.73
Very high	521015	21.12	2315	31.26
Extremely high	205879	8.34	2598	35.08

In this study, the forward stepwise LR approach was used to incorporate predictor variables with major contributions to the presence of landslides, using the SPSS 20 software. In the training dataset represented the presence of landslide points and were assigned the value 1. In agreement with the equal proportions of landslides and non-landslides, the same number of non-landslide points were randomly sampled from the landslide-free area and assigned the value of 0.

The result is shown in Table 6-5 with all original factors. Additionally, it is necessary to examine the effect of correlation because LR is sensitive to collinearity among the independent variables. Table 6-6 and shows the multi-collinearity diagnosis indexes for variables used in the LR equation. Tolerance and the variance inflation factor (*VIF*) are two important indexes for multi-collinearity diagnosis. A tolerance of less than 0.20 or 0.1 and/or a *VIF* value of 5 or 10 and above indicates a multi-collinearity problem (O’Brien 2007). The Tolerance value ranges [0.802, 0.999], and *VIF* ranges [1.012, 1.247]; hence there are no distinct multi-collinearity between the optimized six factors. According to Table 6-7, it shows that all the causative factors have a *p*-value less than 0.05, indicating a statistical correlation between factors and the susceptibility of landslides at the 90% confidence level (Tien Bui et al. 2011). The occurrence of landslide probability (*P*) can be computed as mentioned before.

Lastly, the regression coefficients of the predictors were imported to generate the landslide susceptibility map (Figure 6-6) in GIS. The two maps of classes are also both applied the natural break classification to divide the boundaries of each class. Figure 6-7 and Table 6-8 show that 90.2% of the total landslides occurred in 57.28% of the area classified as high, very high and extremely high susceptibilities if the 13 original factors were used, while 96.98% of the total landslides occurred in 51.016% of the area classified as high, very high and extremely high susceptibilities if the optimized six factors were used (Figure 6-8 and Table 6-9).

Table 6-5 Coefficients, statistics of the factors with all the 13 factors used in the LR equation.

Landslide factors	B	S.E.	Wald	Df	Sig.	Exp(B)	95%C.I. for EXP(B)	
							Lower	Upper
Elevation	.28	.03	101.16	1	.00	1.32	1.25	1.39
Slope	.72	.03	495.15	1	.00	2.05	1.93	2.18
Aspect	.27	.03	121.63	1	.00	1.32	1.25	1.38
Density of geology boundary	.08	.03	6.40	1	.01	1.08	1.01	1.15
Drainage density	-.06	.03	3.27	1	.07	.95	.89	1.01
Plan curvature	-.05	.03	2.26	1	.03	.95	.90	1.01
Total curvature	.17	.03	38.01	1	.00	1.18	.166	.03
Lithology (1)	-.2.3	.11	464.70	1	.00	.10	.08	.12
Lithology (2)	-.78	.23	11.15	1	.00	.46	.29	.73
Lithology (3)	-.25	.12	4.62	1	.03	.77	.62	.98
SPI	.08	.03	8.70	1	.00	1.09	1.03	1.15
CTI	.11	.11	7.46	1	.01	1.12	1.03	1.21
Distance to drainage networks	.35	.35	128.23	1	.00	1.42	1.33	1.51
Distance to geological boundary	.04	.03	1.73	1	.18	1.04	.99	1.1
Profile curvature	.02	.04	.38	1	.53	1.02	.96	1.09
Constant	.84	.03	967.79	1	.00	2.32		

Table 6-6 Multi-collinearity diagnosis indexes for the six selected factors.

Landslide factors	B	S.E.	Std. coefficients Beta	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
						Lower Bound	Upper Bound	Tolerance	VIF
Constant	.63	.00		143.71	.00	.62	.64		
Slope	.13	.01	.26	26.47	.00	.12	.13	.86	1.17
Aspect	.05	.00	.10	11.34	.00	.04	.06	.99	1.00
Density of geology boundary	.02	.00	.04	4.17	.00	.01	.03	.98	1.02
Drainage density	.03	.00	.06	6.06	.00	.02	.04	.98	1.01
Plan curvature	.01	.00	.02	1.96	.04	.00	.02	.98	1.00
Lithology	.126	.01	.26	25.81	.00	.12	.14	.99	1.25

Table 6-7 Beta coefficients and test statistics of the variables used in the LR equation after optimization.

Landslide factors	B	S.E.	Wald	Df	Sig.	Exp (B)
Slope	.74	.03	611.63	1	.00	2.10
Aspect	.28	.02	125.57	1	.00	1.31
Density of geology boundary	.1	.02	15.90	1	.00	1.10
Drainage density	.15	.02	34.91	1	.00	1.16
Plan curvature	.04	.02	3.15	1	.00	1.04
Lithology(1)	-2.326	.10	479.29	1	.00	.09
Lithology(2)	-.68	.23	8.47	1	.00	.51
Lithology(2)	-.36	.11	9.86	1	.00	.69
Constant	.83	.02	986.82	1	.00	2.30

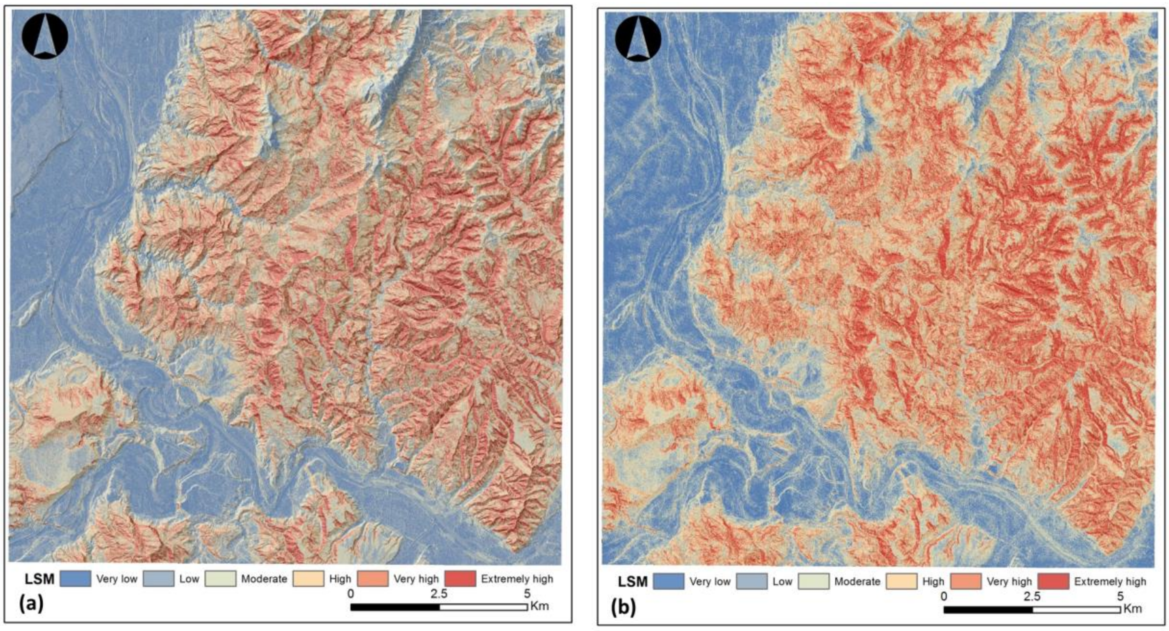

Figure 6-6 LSM maps produced by the LR method: a) six selected factors, and b) original 13 factors. Maps indicate the spatial probability of landslide occurrence in six classes.

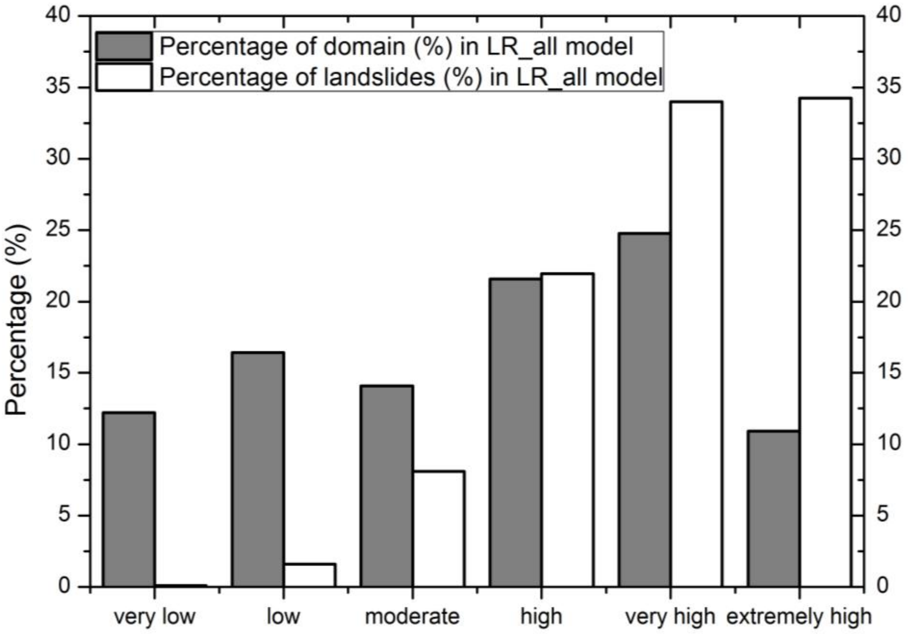

Figure 6-7 Susceptibility class distribution within the study area and the occurrence of landslides according to the classification scheme for LSM using the SI method with the 13 factors.

Table 6-8 Result of statistics analysis concerning landslide susceptibility from the SI method with the 13 factors.

Class	Area of each class	Percentage of domain（%）	No. of landslides	Percentage of landslides (%)
Very low	301311	12.21	8	0.11
Low	405194	16.42	119	1.61
Moderate	347487	14.08	599	8.09
High	532429	21.58	1625	21.94
Very high	611418	24.78	2517	33.99
Extremely high	269570	10.93	2537	34.26

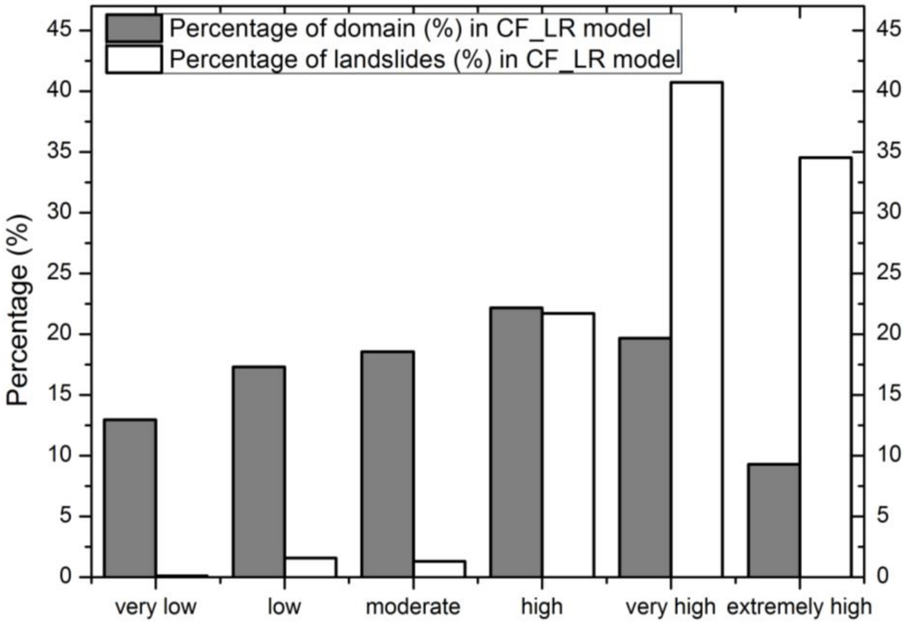

Figure 6-8 Susceptibility class distribution within the study area and the occurrence of landslides according to the classification scheme for LSM using the SI method with the selected six factors.

Table 6-9 Result of statistics analysis concerning landslide susceptibility from the SI method with the selected six factors.

Class	Area of each class	Percentage of domain （%）	No. of landslides	Percentage of landslides (%)
Very low	319706	12.96	9	0.12
Low	427121	17.31	117	1.58
Moderate	458178	18.57	98	1.32
High	547426	22.19	1608	21.72
Very high	485413	19.67	3016	40.73
Extremely high	229565	9.30	2557	34.53

6.3.3 Accuracy assessment

It is essential to verify the accuracy of any prediction model. We verified the accuracy of the *CF* model used for this study and the results using *ROC* and *AUC*.

For the verification, the total landslides were divided into two groups, training data and validation data and an *ROC* plot of sensitivity (true positive rate) and 1-specificity (false positive rate) was made. For the SI method, the *AUC* value for Chuetsu (0.76) is higher for the use of the optimal six factors than 0.67 for the original 13 factors (Figure 6-9). For the LR model, the *AUC* value of the prediction rate curve (78.1%) using the six factors is higher than that from the 13 factors (67.5%) as shown in Figure 6-10. Additionally, the frequency ratio (FR) method is also used to make landslide susceptibility mapping. From the Figure 6-10, six selected factors give higher accuracy than that of using all the original factors in the FR method. In summary, use of the six factors gives higher accuracy than using all the original factors. Additionally, Moreover, LR has a slightly higher accuracy than SI and FR in terms of AUC.

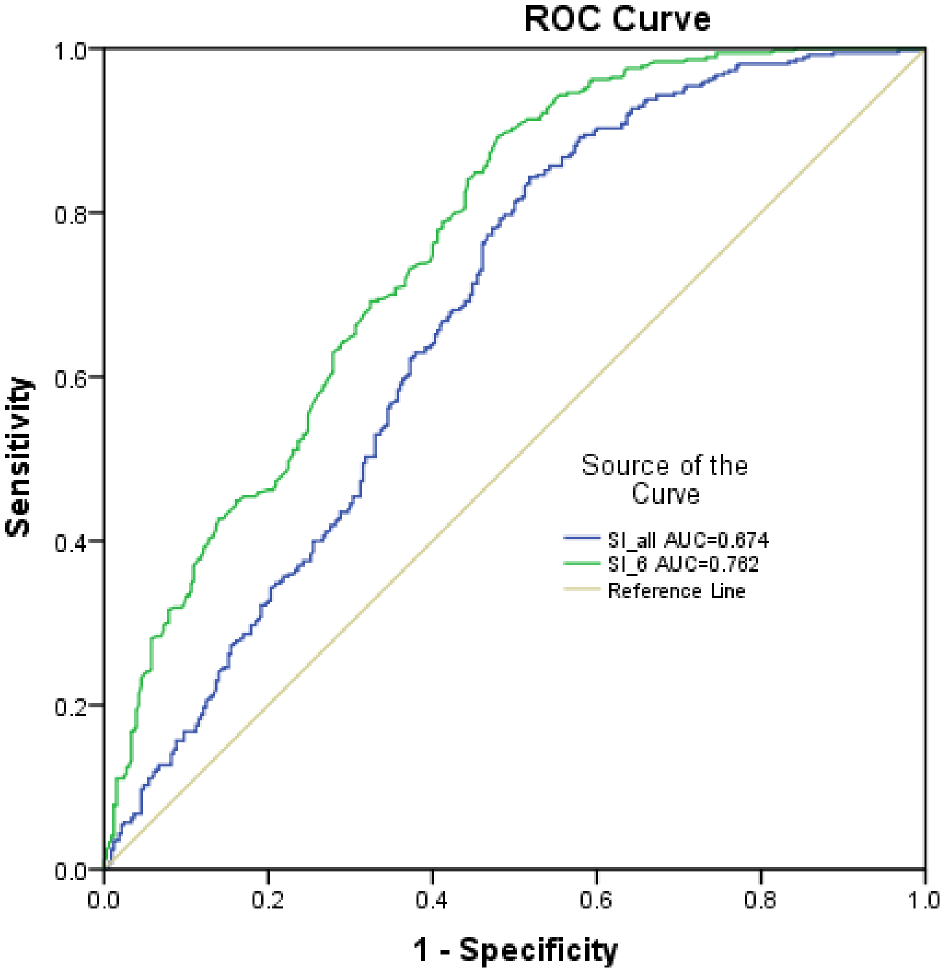

Figure 6-9 ROC curves for landslide susceptibility maps produced using *SI* with the selected six and original 13 factors for Chuetsu.

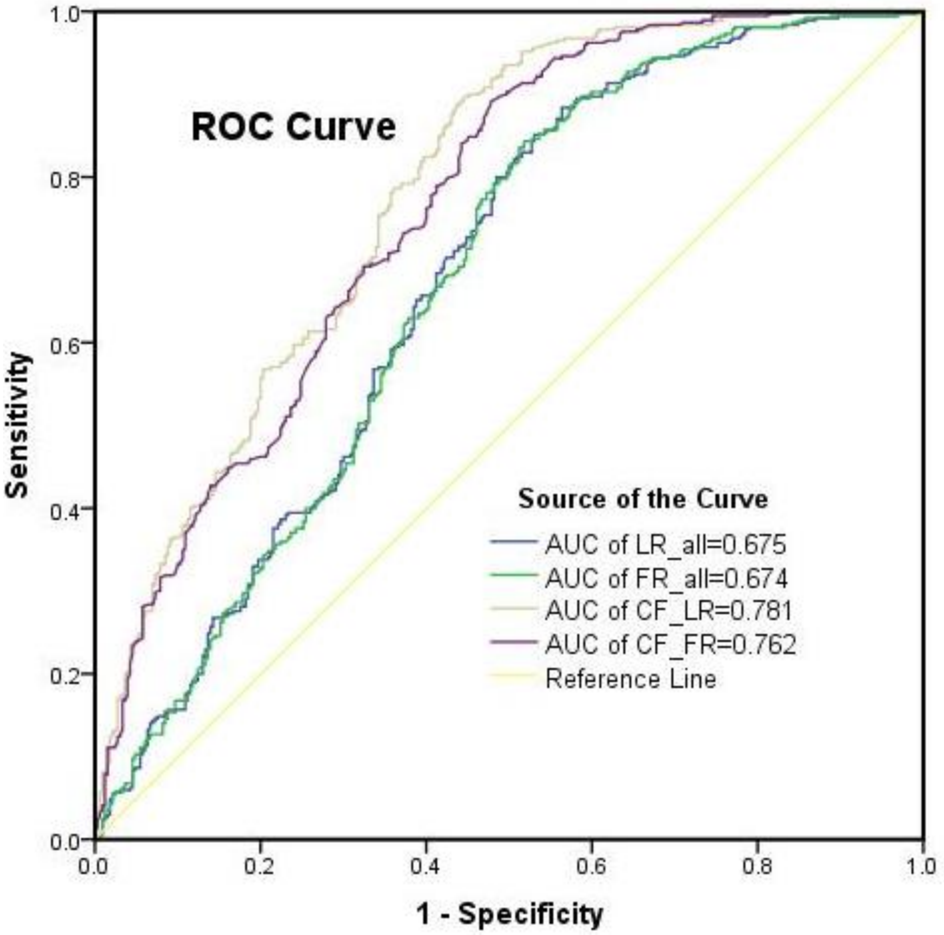

Figure 6-10 ROC curves for landslide susceptibility maps produced using LR with the selected six and original 13 factors for Chuetsu.

6.4. LSM in Sado Island, Japan

6.4.1 Relationship between landslide occurrence and causative factors

Figure 6-11 shows the results of frequency analysis for Sado Island to explore the relationship between the landslide causative factors and landslide occurrence. The frequency of landslides is less than 10% at the elevation less than 100 m due to the gentle terrain characteristics. At the intermediate elevation (100–300 m), the frequency of landslide occurrences tends to increase, as slopes may be prone to sliding due to the cover by thin colluvium deposits. As expected, at higher elevations, the frequency increases. For elevations higher than 600 m, the areal extent of land is low and therefore the frequency of landslide occurrences is also lower. The correlation analysis between landslide occurrence and slope angle is shown in Figure 6-11. The figure indicates that gentle slopes have a low landslide frequency because of the lower shear stress at the slope angles 0–10°. It is obvious that the landslide frequency increases for slope angles 15–35°. However a decrease of landslide occurrences at > 45° slope is also observed.

It is believed that slope angle and aspect may affect vegetation patterns and soil properties, and in turn landslide susceptibility. In the study area the landslide frequency for north-facing slopes is relatively low, and it increases with the orientation angle, reaching the maximum for the south-facing slopes. We also investigated the direction of landslides. The slope aspect from the DEM and the direction of landslides for Sado Island were compared with each other (Figure 6-12). We observe that landslides mostly occur at slope directions of SE, SW, and S (Figure 6-12a) and at SW, NE, S, and SE are major directions of landslide (Figure 6-12b). The results of LSM described were slightly different, but both of them suggest a similar conclusion, for instance, landslides frequently occurred at SW- and S-facing slopes. Because the western part of the ridge (NNE-SSW) of Osado are characterized by less snow cover due to strong western wind, however, much more snow existed in the eastern side of the ridge, and snow melted quickly in the southeast area. This may increase the soil moisture and aggravated the slope instability. Therefore, the likely occurrence of the landslide direction is relatively higher in the southeast part.

Figure 6-11 shows that landslides mostly occurred at the 0–5 category of the total curvature, while for the profile curvature landslides frequently was highest at the -2–0 category followed by the 2–4 category. For the plan curvature, the landslides usually occurred in the concave space because it increases the moisture content of the soil and leads to slope failure. However, in this study, most landslides occurred in the convex space. This may be because the mountain ridges in Sado have got weaker due to local tectonics.

For the hydrological factors *CTI* and *SPI*, landslides mostly occurred at the 0–3 and 0–2 classes, respectively. It may be inferred that dense drainage networks lead to increased occurrences of landslides. However, in this study landslides frequently was highest at the 1–3 m^{-1} drainage density class. For the distance to drainage networks, landslide frequency reached the maximum at 60–120 m, followed by 120–200 m. This may be attributed to more activated gully erosion on slopes that facilitates landslides.

Geological factors such as lithology, density of geological boundaries, distance to geological boundary, and distance to faults are related to rock strength and permeability and in turn slope failure. The results show that landslides mostly occurred with the volcanics (dacite and andesite lava). With respect to the density of geological boundaries, the landslide frequency was highest at the 0–70 m^{-1}, followed by 170–70 m^{-1} suggesting that higher tectonic activity causes slope instability. For the distance to a geological boundary, the weaker boundaries seem to have led to slope instability. The landslide frequency decreases with increasing distance and has the maximum at the <100 m class. Regarding the distance to faults, the results show that the majority of landslides falls into the category of the biggest distance to faults (>400 m). Other than this category, however, the landslide frequency increase in the proximity of fault, suggesting the effect of localize tectonics.

For the vegetation factor, landslide frequency is usually high for lower NDVI values (<0.05) (Ahmed 2014) because the roots of vegetation can retain the slope surface, especially for the shallow landslides. Nevertheless, in Sado, landslides often occurred in areas with thick vegetation cover (*NDVI* > 0.25) because the shallow roots of vegetation seldom influence large landslide occurrence.

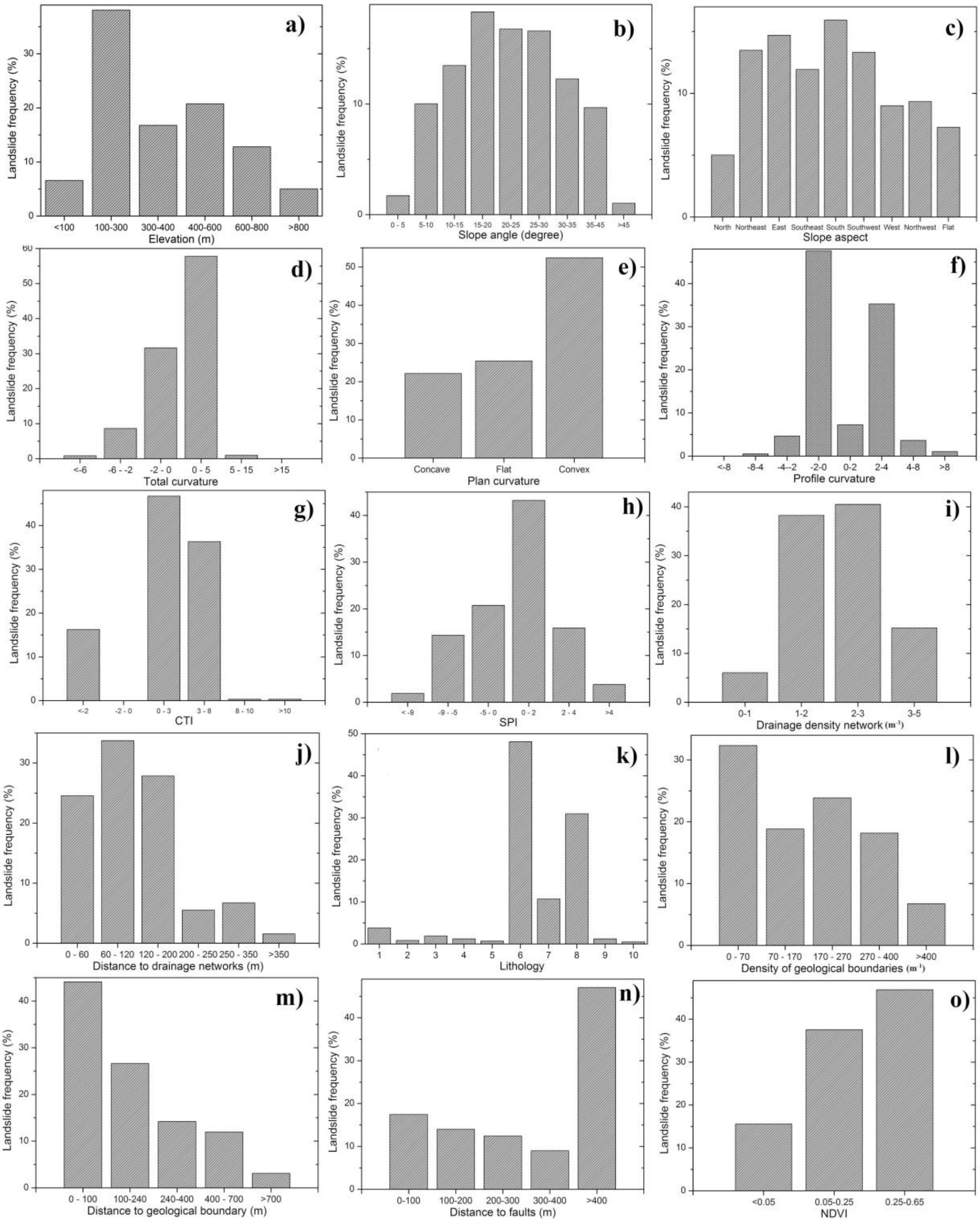

Figure 6-11 Correlations between landslide frequency and the causative factors in Sado Island.

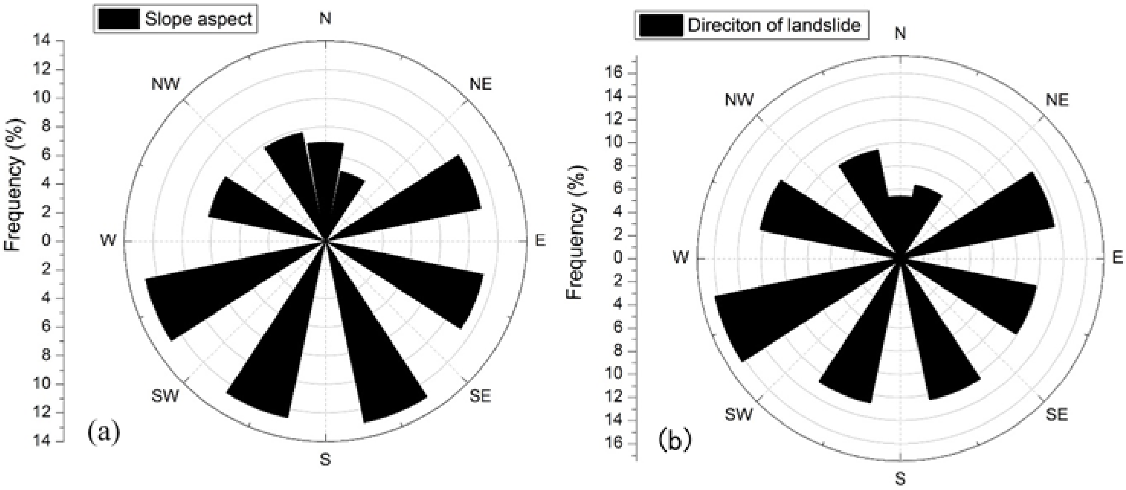

Figure 6-12 Frequency of slope aspect (a), and direction of landslides (b).

6.4.2 Feature selection

The results of the correlation analysis between the landslide occurrence and causative factors are shown in Table 6-10. The *Z* value is positive for slope angle (0.05), slope aspect (0.03), drainage density (0.34), lithology (0.3), distance to geological boundary (0.4) and distance to faults (0.35). The *Z* value is negative for the other factors. Therefore, these six factors are selected for LSM.

A detailed analysis shows that slope angle has the highest influence on the slope stability. *CF* values are positive at slopes from 5–30^{o} (Table 6-10). The percentage of landslide occurrence at the slope classes 10–15^{o}, 15–20^{o}, and 25–30^{o} are 17.82%, 21.79%, and 15.4%, respectively. The landslide frequency increases with an increasing slope angle up to 20 ^{o}, and then decreases. This agrees with the landslide frequency in Figure 6-11c. Concerning slope aspect, landslides mostly occurred on east-, southeast, and south facing slopes with positive *CF* values from 0.09 to 0.15. The highest percentage of landslides with the maximum *CF* value (0.15) was 15.9% for the south-facing slopes, followed by the east-facing slopes (14.2 %). The snow in the study area is normally blown out by the wind from the northwest; therefore, snow accumulates on south to east-facing slopes that may cause landslides during snow melting. Drainage density shows positive *CF* values for the classes

2–3 and 3–5. The maximum positive *CF* value of 0.3 is associated with the 3–5 class. The highest percentage of landslide occurrence is 40.48% at the 2–3 class. Concerning lithology, six lithology classes have positive CF values. The highest percentage of landslides in the lithology class (volcanic-dacite) is 49.31% with a *CF* value of 0.22. More than 50% of the landslides occurred along the margins of dacite and dacite lava. These lavas once covered by ocean and were transformed into pelitic rocks that further changed to materials rich in smectite clay susceptible to sliding. The feature “distance to geological boundary” shows positive *CF* values for classes >100 m; however, the highest percentage of landslides occurred for the class < 100 m. It indicates that the closer to a geological boundary, the more occurrences of landslides. The distance to faults shows positive *CF* values for the classes 0–100, 100–200, and 200–300 m but the *CF* values become negative for 300 m. The maximum CF value is 0.25 at 0–100 m.

Table 6-10 Spatial relationship between the causative factors and landslide occurrence based on the CF and SI methods.

Causative factors	Class	Percentage of domain (%)	No. of landslides	No. of landslide pixels	CF	Z	SI
Elevation (m)	<100	21.33	38	8103	-0.69	-0.42	-1.59
	100-300	30.68	220	62288	0.20		0.08
	300-400	13.09	97	31464	0.22		0.25
	400-600	17.74	120	50504	0.15		0.42
	600-800	11.80	74	29400	0.08		0.29
	>800	5.36	29	5460	-0.07		-0.61
Slope angle (°)	0 - 5	11.77	18	6317	-0.68	0.05	-1.25
	5-10	9.06	58	26860	0.26		0.46
	10-15	8.99	103	27721	0.59		0.32
	15-20	10.35	126	31045	0.61		0.47
	20-25	11.97	86	23115	0.34		0.03
	25-30	13.39	89	28873	0.28		0.14
	30-35	13.98	47	21260	-0.30		-0.21
	35-45	18.05	45	18667	-0.48		-0.59
	>45	2.44	6	3361	-0.49		-0.31
Slope aspect	North	5.74	32	10672	-0.04	0.03	-0.01
	Northeast	12.64	65	25359	-0.11		0.07
	East	12.89	82	25931	0.09		0.07
	Southeast	12.59	80	24548	0.09		0.04
	South	13.41	92	27897	0.16		0.11
	Southwest	14.68	74	23033	-0.13		-0.02
	West	11.84	59	20342	-0.14		-0.09
	Northwest	10.67	54	19635	-0.13		-0.02
	Flat	5.53	40	10802	0.20		0.04
Total curvature	<-6	21.33	5	3321	-0.96	-1.00	-2.49
	-6 - -2	30.68	50	23168	-0.72		-0.91
	-2 - 0	13.09	183	64542	0.60		0.97
	0 - 5	17.74	334	93753	0.70		1.04
	5 - 15	11.80	6	2433	-0.91		-2.21
	>15	5.36	0	2	-1.00		-8.52
Profile curvature	<-8	0.03	0	7	-1.00	-1.00	-1.92
	-8-4	0.87	3	1189	-0.40		-0.31
	-4--2	5.81	27	9310	-0.20		-0.15
	-2-0	40.00	275	76347	0.16		0.02
	0-2	9.04	42	11779	-0.20		-0.36
	2-4	36.74	204	75185	-0.04		0.09
	4-8	5.83	21	10719	-0.38		-0.02

Plan curvature	Concave	8.02	34	15730	-0.27	-0.22	0.05
	Flat	36.24	212	69796	0.01		0.03
	Convex	55.74	332	101693	0.03		-0.03
CTI	<-2	14.48	94	23738	0.11	-1.00	-0.13
	-2 - 0	0.34	0	480	-1.00		-0.29
	0 - 3	52.98	270	94694	-0.12		-0.05
	3 - 8	29.68	210	62790	0.19		0.12
	8 - 10	1.47	2	3866	-0.77		0.34
	>10	1.05	2	1651	-0.67		-0.18
SPI	<-9	3.50	11	3023	-0.46	-0.61	-0.77
	-9 - -5	11.07	83	20789	0.23		0.00
	-5 - 0	26.55	120	39004	-0.22		-0.24
	0 - 2	42.38	250	84561	0.02		0.06
	2 - 4	12.56	92	28670	0.21		0.20
	4 - 12	3.94	22.00	11172	-0.03		0.41

Table 6-10 Continued.

Causative factors	Class	Percentage of domain (%)	No. of landslides	No of landslide pixels	CF	Z	SI
Drainage density (m^{-1})	0-1	10.31	35	10261	-0.42	0.34	-0.63
	1-2	40.98	221	70163	-0.07		-0.09
	2-3	37.92	234	78979	0.06		0.11
	3-5	10.78	88	27907	0.30		0.32
Distance to drainage networks (m)	0 - 60	25.44	142	64448	-0.03	-0.82	0.30
	60 - 120	22.99	208	49926	0.37		0.15
	120 - 200	23.61	148	39592	0.08		-0.11
	200 - 250	11.18	32	15262	-0.51		-0.32
	250 - 350	12.82	39	15250	-0.48		-0.45
	>350	3.96	9	2832	-0.61		-0.96
Lithology	1. Sedimentary (sandstone)	15.41	27	8082	-0.70	0.30	-1.27
	2. Sedimentary (mudstone)	3.95	11	1500	-0.52		-1.60
	3. Plutonic and intrusives	2.23	15	5005	0.14		0.18
	4.Volcanic (basalt)	0.63	6	1207	0.40		0.02
	5. Volcanic (rhyolite lava)	0.63	3	281	-0.18		-1.43
	6. Volcanic (dacite)	38.82	285	95356	0.22		0.27
	7.Volcanic (dacite lava)	5.75	52	12766	0.37		0.17
	8. Volcanic (andesite lava)	30.92	162	60295	-0.09		0.04
	9.Sedimentary (slate and sandstone)	1.05	7	1308	0.13		-0.41
	10.Metamorphic	0.60	10	1430	0.66		0.24
Density of geological boundaries (m^{-1})	0 - 70	30.26	184	65069	0.05	-0.03	0.14
	70 - 170	21.96	112	41448	-0.12		0.01
	170 - 270	23.56	138	48197	0.01		0.09
	270 - 400	17.41	105	23067	0.04		-0.35
	>400	6.81	39	9498	-0.01		-0.29
Distance to	0 - 100	48.57	255	74525	-0.09	0.40	-0.20
	100 - 240	25.68	154	54091	0.04		0.12
	240 - 400	13.79	82	30935	0.03		0.18

geological boundary (m)	400 - 700	9.72	69	21069	0.19		0.15
	>700	2.25	18	6659	0.28		0.00
Distance to faults (m)	0-100	13.19	101	23269	0.25	0.35	-0.06
	100-200	11.82	81	20520	0.16		-0.08
	200-300	10.67	72	16936	0.15		-0.17
	300-400	9.25	52	14905	-0.03		-0.15
	>400	55.06	272	111638	-0.15		0.08
NDVI	<0.05	24.43	89	24965	-0.37	-0.22	-0.21
	0.05-0.25	37.60	218	49489	0.00		0.04
	0.25-0.65	37.96	271	51402	0.19		0.07

6.4.3 LSM

The correlation between the landslide occurrence and causative factors using *SI* is represented in Table 6-10. Two landslide susceptibility maps were generated: (i) using the six selected factors (*CF* > 0) and (ii) using the original 15 factors. The result is shown in Figure 6-13. Based on the natural breaks inherent in the data, the susceptible level is eventually divided into six classes; i.e., extremely low, low, moderate, high, very high and extremely high (Table 6-11). There are much more red color areas in Figure 6-13b, whereas there are more dark blue areas in Figure 6-13a. In the detailed maps for a small area in Figure 6-12, black lines denote main scarps and blue lines denote dissected crowns. Figure 6-14 shows that 90.18% of the total landslides occurred in the 69.66% of the area classified as high, very high and extremely high susceptibilities when the optimized six factors were used, while 73.41% of the landslides occurred in the 93.1% of the area classified as high, very high and extremely high susceptibilities when the original 15 factors were used.

According to Table 6-10, the slope angle class (15–20 ^{o}) with the highest *SI* value of 0.47 is most susceptible, having the highest percentage of landslide occurrence of 16.58%. The landslide occurrence gradually increases with increasing slope angle and then it drops after 35^{o}. This result is similar to that of CF modelling.

Landslides occurred more on northeast-, east-, southeast- and south-facing slopes. The highest percentage of landslides with the maximum *SI* value (0.1) is 14.9% along the south-facing slopes, followed by the 13.85% for the east-facing slopes. This also agrees with the results obtained from CF.

With an increase in drainage density the *SI* values are amplified, pointing to increased landslides here (*SI* > 2). The highest percentage of landslide occurrence in this class is 42.19%. This result is also in agreement with CF.

With respect to lithology, the results also display that the six lithology classes (similar to CF) have stronger relationships with landslide occurrence. The highest percentage of landslides among the lithology class (volcanic-dacite) is 50.93% with a maximum *SI* value of 0.27. The landslide frequency along the distance to lithologic boundary of dacite and dacite lava are greater than 50%. The distance to geological boundary indicates that classes >100 m have high probabilities of landslide occurrence (Table 6-10). The highest percentage (39.81%) of landslides is observed at the class >100 m. The distance to faults exhibits negative *SI* values for the classes, 0–100, 100–200, and 200–300 m and then the *SI* values become positive after 300 m.

Table 6-11 The boundaries classes for susceptibility maps.

Susceptible class	SI method		LR method	
	15 factors	6 factors	15 factors	6 factors
Extremely low	-12.31- -3.41	-3.83 - -2.07	0.00 - 0.13	0.03-0.207
Low	-3.41 - -2.17	-2.07- -1.20	0.13 - 0.32	0.20 - 0.38
Moderate	-2.17 - -0.94	-1.20- -0.577	0.32- 0.51	0.38 - 0.51
High	-0.94 - 0.23	-0.57- -0.04	0.51 - 0.67	0.51 - 0.64
Very high	0.23 - 1.33	-0.04 - 0.44	0.67 - 0.81	0.64 - 0.75
Extremly high	1.33- 4.26	0.44 - 1.85	0.81- 0.98	0.75 - 0.91

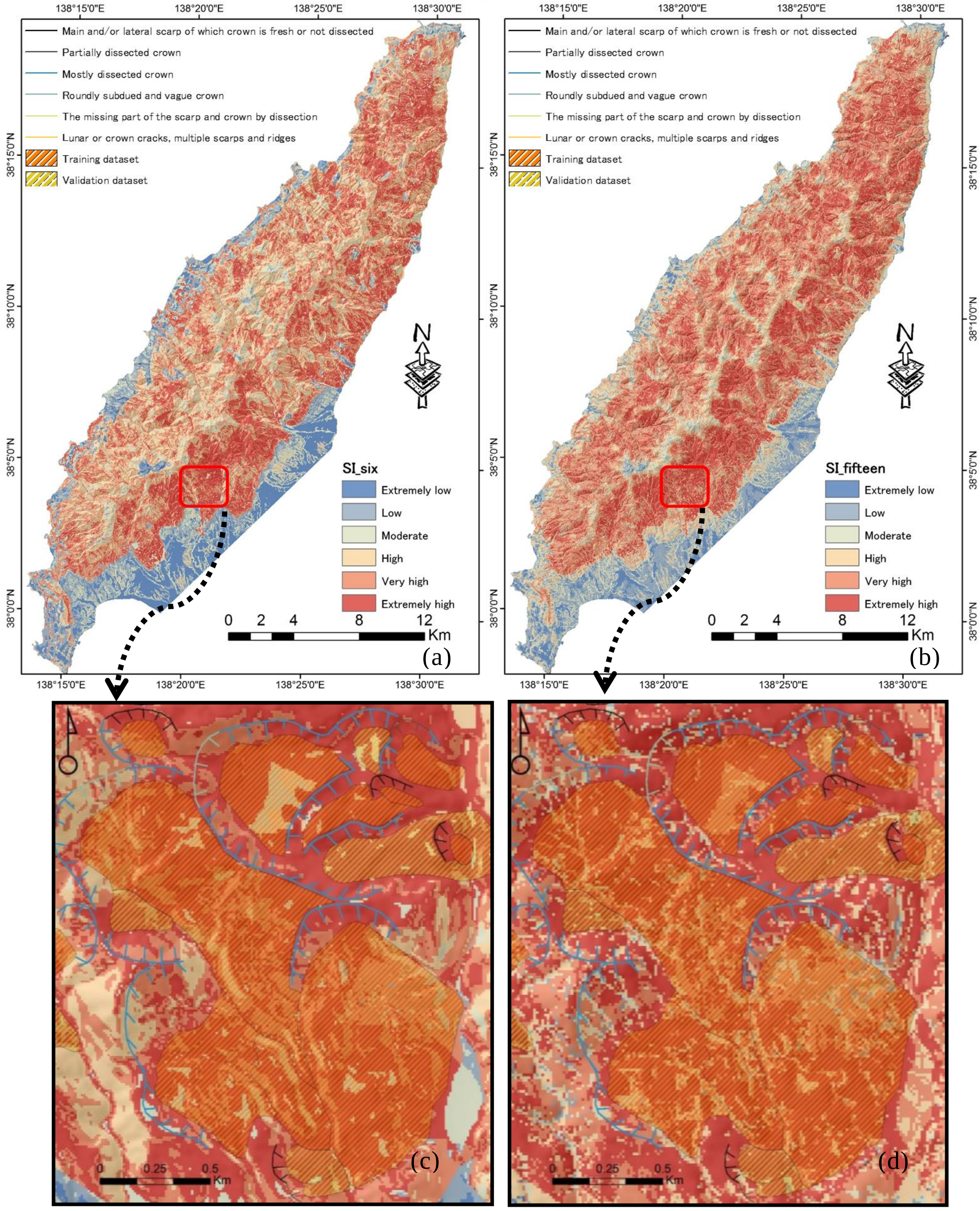

Figure 6-13 Landslide susceptibility maps generated by the SI method from a) selected six factors and b) original fifteen factors, with six susceptibility classes on the basis of natural break criterion. The lower maps are enlargements (c and d).

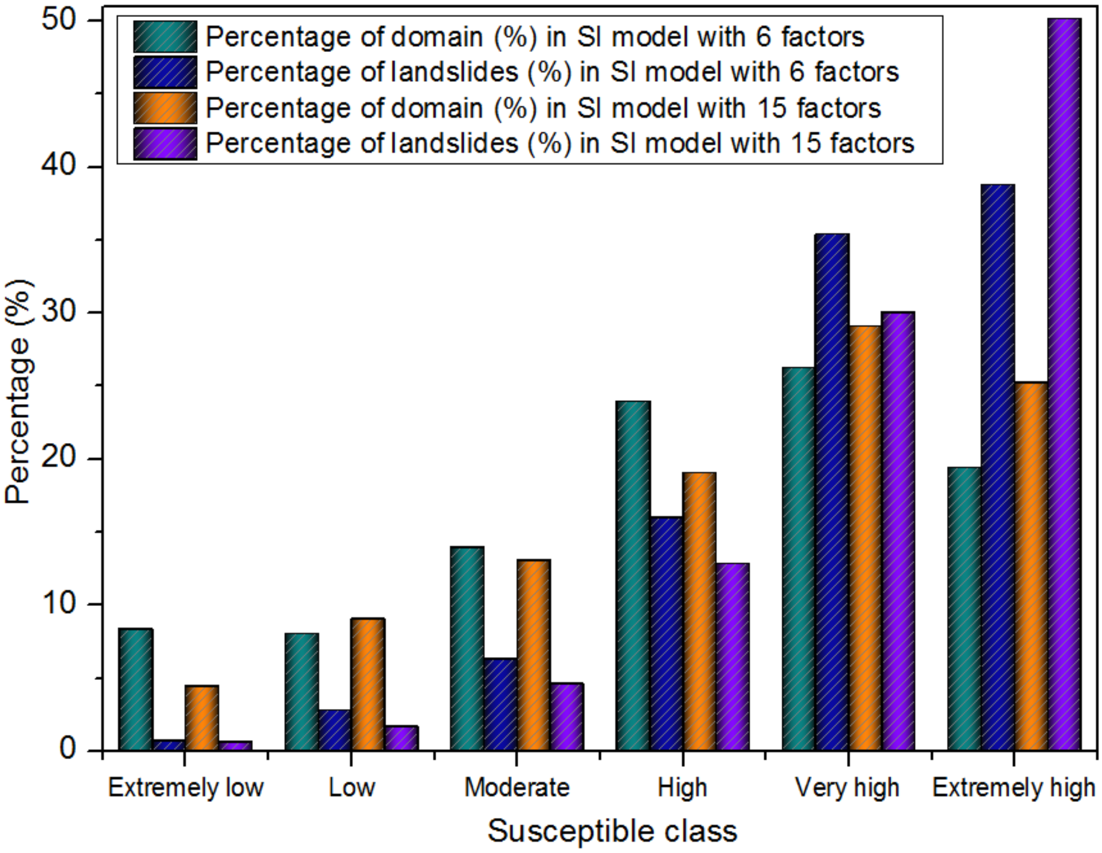

Figure 6-14 Frequency landslide susceptible classes obtained from the SI model.

In this study, the forward stepwise LR approach was used to incorporate predictor variables, using the SPSS 20 software. In the training dataset 578 landslides represented by points were assigned the value of 1. The same number of non-landslide points were randomly sampled from the landslide-free area and assigned the value of 0. The result of LR is shown in Table 6-12. All the causative factors have *p*-values less than 0.1, indicating a statistical correlation between factors and the susceptibility of landslides at the 90% confidence level (Tien Bui et al. 2011). The interpretation of the LR coefficient for each causative factor (Table 6-12) shows that elevation, slope angle, slope aspect, total curvature, *SPI*, drainage density, lithology, distance to drainage network, distance to geological boundary, and *NDVI* have positive values. Distance to drainage network has the highest value (1.7), followed by slope angle (1.2). The remaining factors have negative effects on landslide occurrence.

It is necessary to examine the effect of correlations between independent variables because LR is sensitive to their collinearity. *VIF* and tolerance (*TOL*) are widely used indexes of the degree of multi-collinearity. Table 6-12 shows that the maximum *VIF* and minimum *TOL* were 1.028 and 0.973, respectively. Therefore, there is no distinct multi-collinearity between these variables.

Lastly, the regression coefficients of the predictors and GIS were used to generate landslide susceptibility maps (Figure 6-15). The natural break classification was applied to divide classes (Table 6-11). Figure 6-16 shows that 91.39% of the total landslides took place in the 72.96% of the area classified as high, very high and extremely high when the optimized six factors were used, while 68.23% of the total landslides occurred in the 90.79% of the high, very high and extremely high areas if the original fifteen factors were used.

Table 6-12 Coefficients, statistics of the factors and the multi-collinearity diagnosis indexes for variables used for LR.

Causative factors	Coefficient	S.E.	Wald	*P*-value	Exp(B)	Collinearity Statistics	
						Tol	VIF
Elevation	0.96	0.13	51.83	0.00	2.60	1.00	1.00
Slope angle	1.21	0.19	39.36	0.00	3.35	0.98	1.02
Slope aspect	0.28	1.57	0.03	0.09	1.33	1.00	1.00
Total curvature	0.31	0.12	6.50	0.01	1.36	1.00	1.00
Profile curvature	-0.76	1.70	0.20	0.07	0.47	0.99	1.01
Plan curvature	-1.19	3.67	0.11	0.07	0.31	1.00	1.00
CTI	-0.23	0.99	0.06	0.08	0.79	0.99	1.01
SPI	0.70	0.49	2.03	0.02	2.01	1.00	1.01
Drainage density	0.12	0.43	0.08	0.08	1.13	1.00	1.00
Distance to drainage networks	1.71	0.33	26.88	0.00	5.51	1.00	1.00
Lithology	0.88	1.56	1.44	0.02	0.15	0.99	1.01
Density of geological boundaries	-0.05	0.54	0.01	0.09	0.96	1.00	1.00
Distance of geological boundaries	0.85	0.57	2.23	0.01	2.35	0.97	1.03
Distance to faults	-0.44	0.89	0.25	0.06	0.64	0.99	1.01
NDVI	0.28	0.84	0.11	0.07	1.33	0.99	1.01
Constant	0.79	0.12	42.69	0.00	2.21		

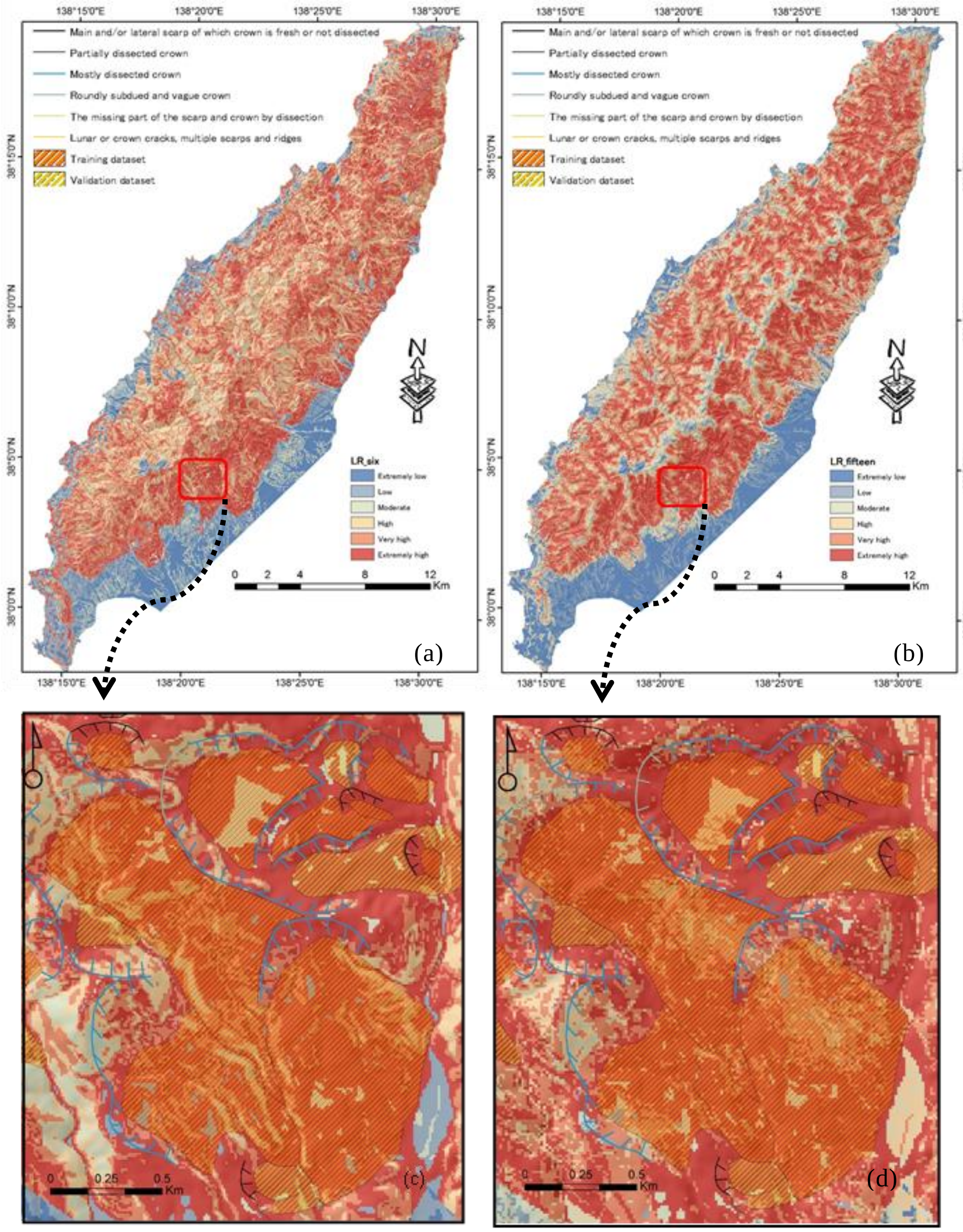

Figure 6-15 Landslide susceptibility maps generated by the LR method from a) selected six factors and b) original fifteen factors, with six susceptibility classes on the basis of natural break criterion. The lower maps are enlargements (c and d).

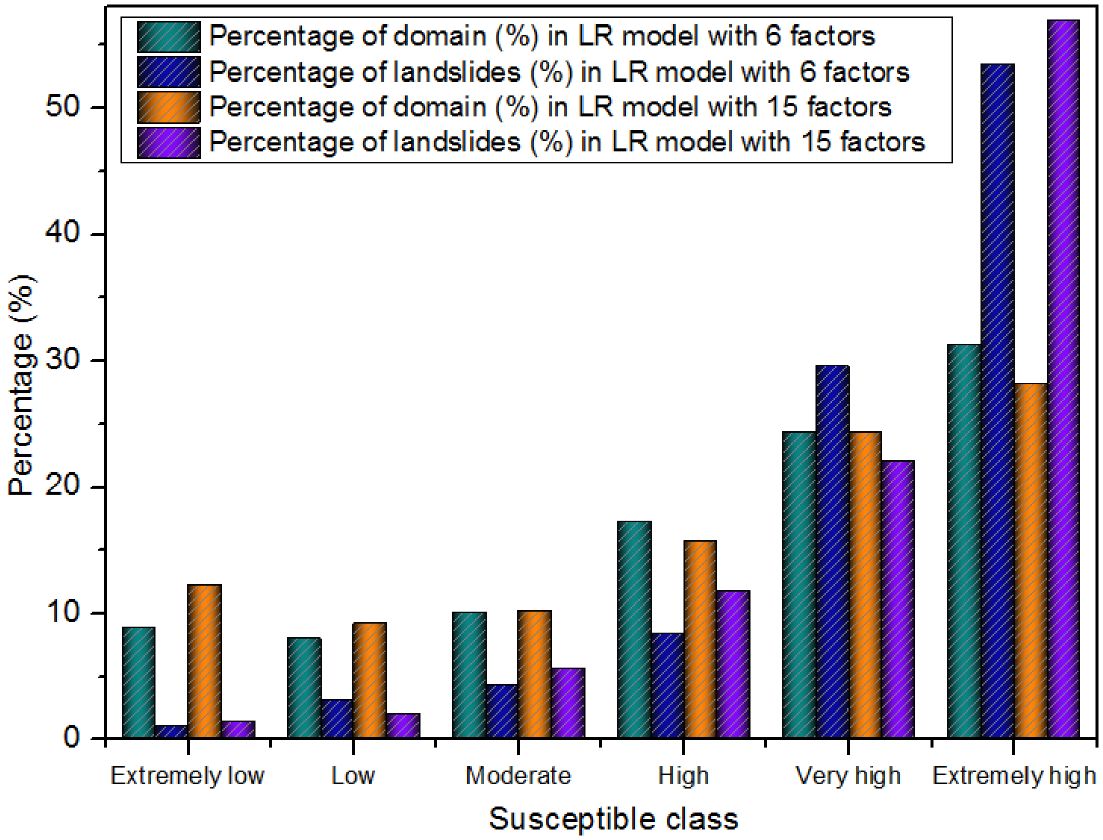

Figure 6-16 Comparison of landslide susceptible classes obtained from the LR model.

6.4.4 Accuracy assessment

LSM results can be validated using the known landslide locations. Accuracy assessment was performed by comparing the distribution of existing landslides that were not included in the data used for LSM. Both the training (70% of 825 polygons) and validation (the rest 30% of 825 polygons) datasets were selected to assess the models. The training data were used for computing the LSM success rate and the validation data for the prediction rate. To obtain both rates, the landslide susceptibility values of all cells were sorted in descending order. Then the ordered values were categorized into 100 classes with 1% cumulative intervals, for which the cumulative percentage of landslide occurrence in the classes was calculated to obtain *AUC*.

Figure 6-17 shows that for the *SI* method the *AUC* value of the success rate curve (80.1%)

using the six factors is higher than for the model using all the 15 factors (73.4%). For the prediction rate curve, the result is similar to that of the success rate curve. For the LR model, the *AUC* value of the success rate curve (81.7%) from the six factors is also higher than that from all the 15 factors (73.2%) as shown in Figure 6-18. The prediction rate shows a similar result to the success rate. Hence, using the six factors gives higher accuracy than using all the factors. Additionally, compared with the SI method, LR has a slightly higher accuracy for both success and predication rates.

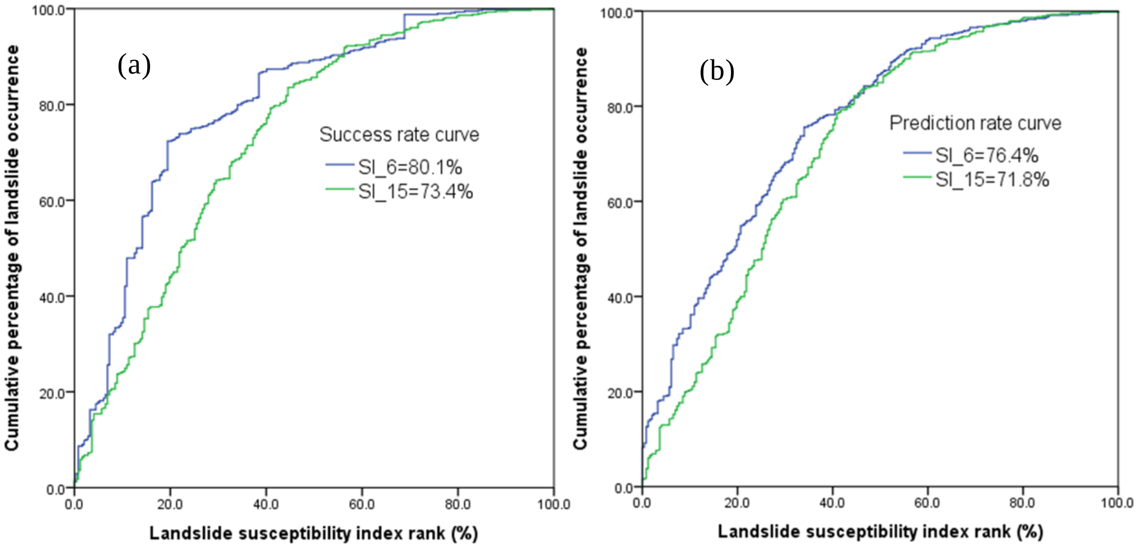

Figure 6-17 ROC curves representing (a) success rate, and (b) prediction rate for the SI method.

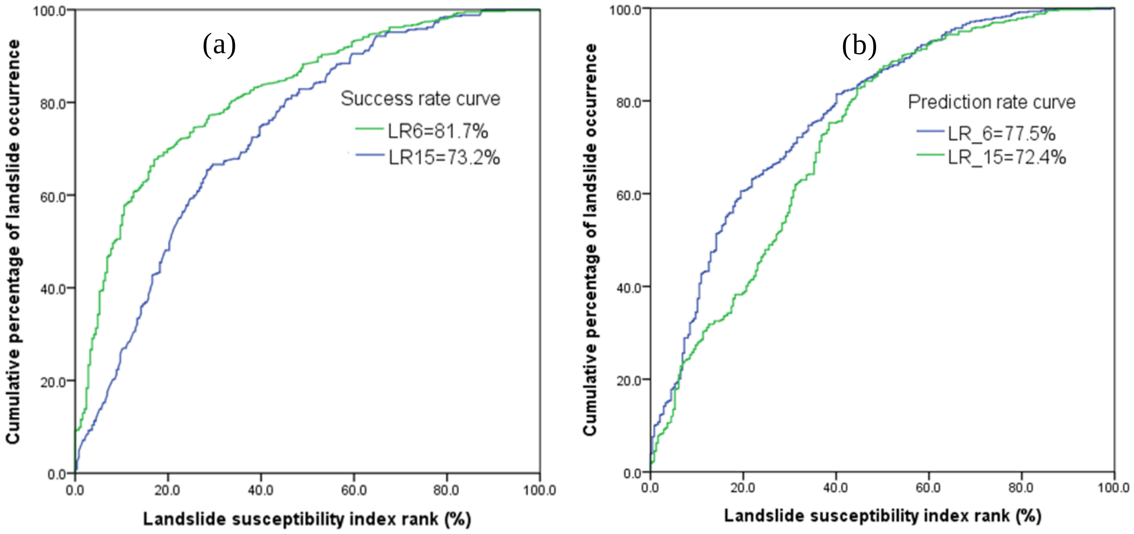

Figure 6-18 ROC curves representing (a) success rate, and (b) prediction rate for the LR model.

6.5. LSM for Dongjiang Reservoir, China

6.5.1 Feature selection

The results of the correlation analysis between the landslide occurrence and causative factors for the Dongjiang Reservoir area are shown in Table 6-13. The result of CF analysis shows that the *Z* value is positive for slope angle (0.25), curvature (0.82), plan curvature (0.21), drainage density (0.96), distance to drainage network (0.11), accumulative rainfall (0.97), and lithology (0.47) as shown in Figure 6-19. The *Z* value is negative for the other factors. Hence, these seven factors are selected for LSM.

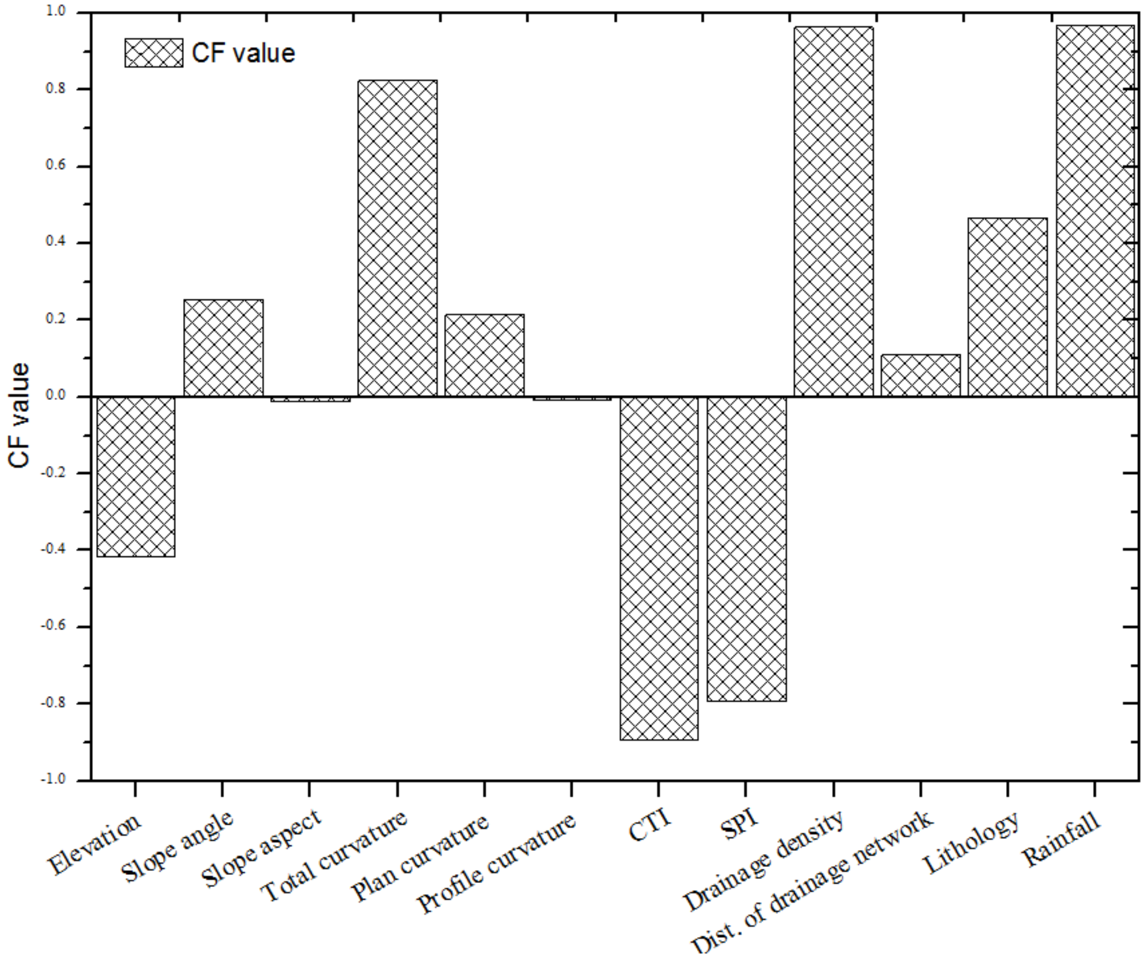

Figure 6-19 Calculation of CF values in the Dongjiang Reservoir

Table 6-13 Spatial relationship between the causative factors and landslide occurrence based on the CF and SI methods.

Factors	Class	No.of pixels in domain	Percentage of domain	No. of landslides	Percentage of landslides	*CF*	*Z*	*SI*
Elevation	<320	49499	16.63	381	21.57	0.56	-0.99	0.26
	320-400	86673	29.11	762	43.15	0.80		0.39
	400-500	70232	23.59	374	21.18	-0.22		-0.11
	500-600	35575	11.95	128	7.25	-0.61		-0.50
	600-700	23467	7.88	64	3.62	-0.74		-0.78
	700-900	25328	8.51	50	2.83	-0.83		-1.10
	>900	6934	2.33	7	0.40	-0.92		-1.77
Slope angle (o)	0-5	36715	12.33	97	5.49	-0.75	0.25	-0.81
	5-10	51448	17.28	259	14.67	-0.30		-0.16
	10-15	57080	19.17	391	22.14	0.33		0.14
	15-20	52738	17.71	361	20.44	0.33		0.14
	20-25	42919	14.42	297	16.82	0.35		0.15
	25-30	28884	9.70	166	9.40	-0.07		-0.03
	30-35	16179	5.43	121	6.85	0.51		0.23
	35-40	7440	2.50	50	2.83	0.29		0.12
	40-62	4305	1.45	24	1.36	-0.14		-0.06
Slope aspect	flat	10424	3.50	5	0.28	-0.97	-0.01	-2.52
	N	15273	5.13	51	2.89	-0.66		-0.57
	NE	32430	10.89	115	6.51	-0.62		-0.51
	E	38153	12.82	299	16.93	0.60		0.28
	SE	42716	14.35	399	22.59	0.90		0.45
	S	37354	12.55	368	20.84	0.98		0.51
	SW	36945	12.41	295	16.70	0.63		0.30
	W	34987	11.75	138	7.81	-0.55		-0.41
	NW	49426	16.60	96	5.44	0.59		-1.12
Curvature	<-3	5207	1.75	33	1.87	0.16	0.82	0.07
	-3--2	20378	6.85	138	7.81	0.30		0.13
	-2--1	43154	14.50	258	14.61	0.02		0.01
	-1-0	93076	31.26	494	27.97	-0.22		-0.11
	0-2	111216	37.36	660	37.37	0.00		0.00
	2-3	15509	5.21	112	6.34	0.44		0.20
	>3	9168	3.08	71	4.02	0.58		0.27
Plan curvature	Concave	56137	18.86	339	19.20	0.04	0.21	0.02
	Flat	167924	56.41	929	52.60	-0.15		-0.07
	Convex	73647	24.74	498	28.20	0.30		0.13

Table 6-13 Continued.

Factors	Class	No.of pixels in domain	Percentage of domain	No. of landslides	Percentage of landslides	*CF*	*Z*	*SI*
Profile curvature	-4--2	142755	47.95	825	46.72	-0.06	-0.01	-0.03
	-2-0	141351	47.48	847	47.96	0.02		0.01
	0-2	6247	2.10	41	2.32	0.24		0.10
	2-4	344	0.12	2	0.11	-0.05		-0.02
	>4	40	0.01	1	0.06	1.87		1.44
Drainage density	<0.8	36537	12.27	32	1.81	-0.93	0.96	-1.91
	.8-1	63095	21.19	218	12.34	-0.64		-0.54
	1-1.2	61055	20.51	608	34.43	0.99		0.52
	1.2-1.3	56106	18.85	411	23.27	0.47		0.21
	1.3-1.4	56897	19.11	284	16.08	-0.32		-0.17
	1.4-1.7	24018	8.07	213	12.06	0.81		0.40
Distance to drainage network	<130	84683	28.45	464	26.27	-0.17	0.11	-0.08
	130-280	78800	26.47	515	29.16	0.23		0.10
	280-450	65049	21.85	360	20.39	-0.15		-0.07
	450-650	49300	16.56	304	17.21	0.09		0.04
	>650	19876	6.68	123	6.96	0.10		0.04
SPI	<-6	102218	34.34	663	37.54	0.21	-0.79	0.09
	-6--2	27218	9.14	70	3.96	-0.76		-0.84
	-2-0	75938	25.51	514	29.11	0.30		0.13
	0-2	65551	22.02	377	21.35	-0.07		-0.03
	2-4	18343	6.16	103	5.83	-0.12		-0.05
	>4	8440	2.84	39	2.21	-0.41		-0.25
CTI	<-2	102436	34.41	663	37.54	0.21	-0.89	0.09
	-2-3	103447	34.75	648	36.69	0.13		0.05
	3-7	52876	17.76	324	18.35	0.08		0.03
	7-11	31389	10.54	116	6.57	-0.60		-0.47
	>11	7560	2.54	15	0.85	-0.83		-1.10
Accumulative rainfall	<300	40434	13.58	54	3.06	-0.89	0.97	-1.49
	300-320	48579	16.32	92	5.21	-0.84		-1.14
	320-345	73703	24.76	735	41.62	1.00		0.52
	345-360	65249	21.92	589	33.35	0.84		0.42
	360-375	54105	18.17	176	9.97	-0.67		-0.60
	375-390	15638	5.25	130	7.36	0.70		0.34
Lithology	Shaly limestone	97853	32.87	281	15.91	-0.72	0.47	-0.73
	Biotite adamellite	95797	32.18	878	49.72	0.87		0.44

Siltstone, shale	27045	9.08	197	11.16	0.46	0.21
Dolomites	12931	4.34	60	3.40	-0.41	-0.25
Sandstone, slate and siliceous rocks	64082	21.53	350	19.82	-0.17	-0.08

6.5.2 LSM

The correlations between the landslide occurrence and causative factors using *SI* is represented in Table 6-13. Two landslide susceptibility maps were generated: (i) using the six selected factors (*CF* > 0) and (ii) using the original factors (Figure 6-20). Based on the natural breaks, the susceptibility level was divided into six classes; i.e., extremely low, low, moderate, high, very high and extremely high. There are much more red color areas in Figure 6-20b, whereas there are more dark blue areas in Figure 6-20a. Figure 6-21 shows that 90.84% of the total landslides occurred in the 52.56% of the area classified as high, very high and extremely high susceptibilities when the original factors were used, while 51.73% of the total landslides occurred in the 92.03% of the area classified as high, very high and extremely high susceptibilities if the optimized seven factors were used (Figure 6-22 and Table 6-15).

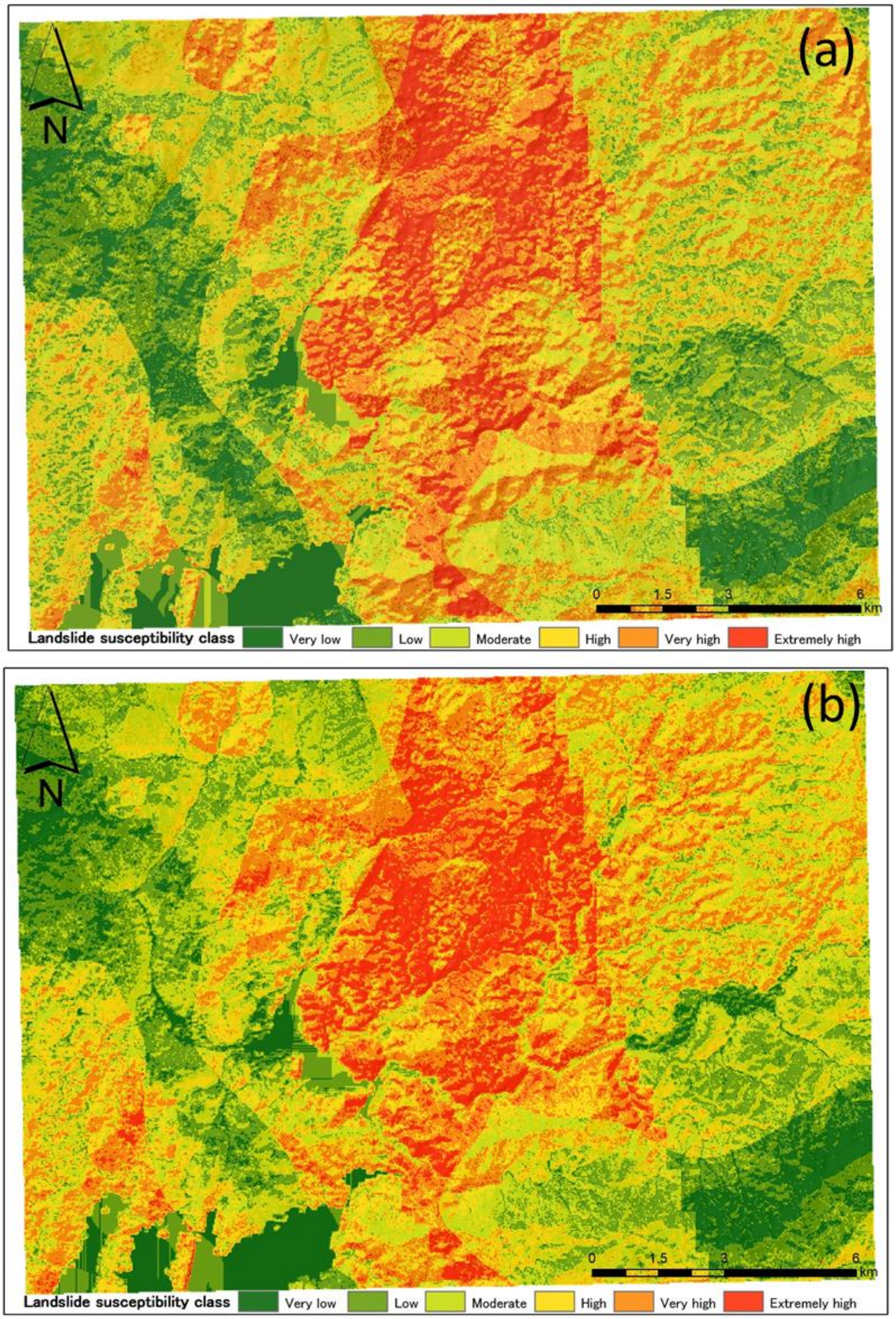

Figure 6-20 LSM maps produced by the SI method: a) selected seven factors, and b) original 12 factors. Maps show the spatial probability of landslide occurrence in six classes.

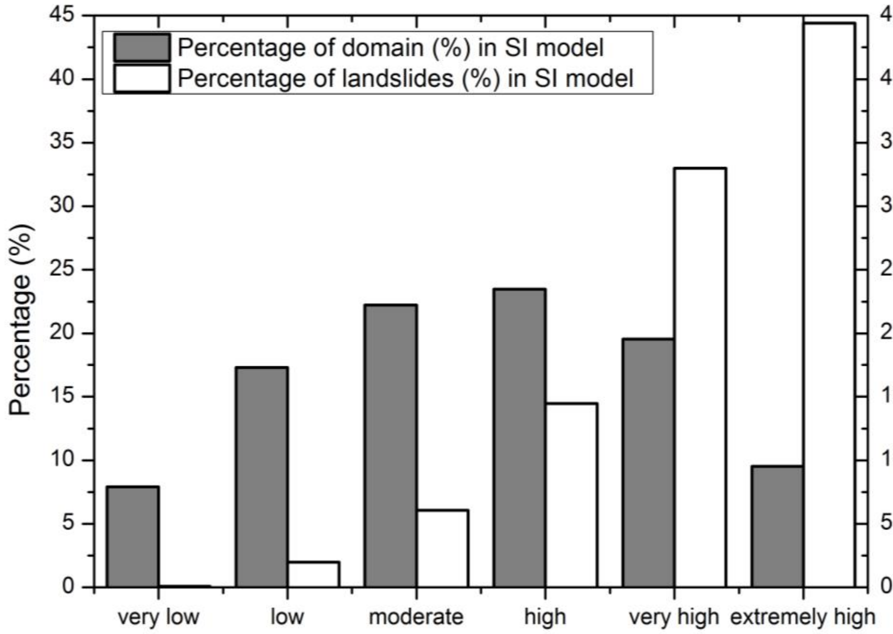

Figure 6-21 Susceptibility class distribution within the study area and the occurrence of landslides according to the classification scheme for LSM using the SI method with the original 12 factors.

Table 6-14 Result of statistical analysis concerning landslide susceptibility from the SI method with the original 12 factors.

Class	Area of each class	Percentage of domain （%）	No. of landslides	Percentage of landslides (%)
Very low	23530	7.90	2	0.09
Low	51529	17.31	44	1.99
Moderate	66159	22.22	134	6.07
High	69919	23.49	319	14.45
Very high	58197	19.55	728	32.99
Extremely high	28374	9.53	980	44.40

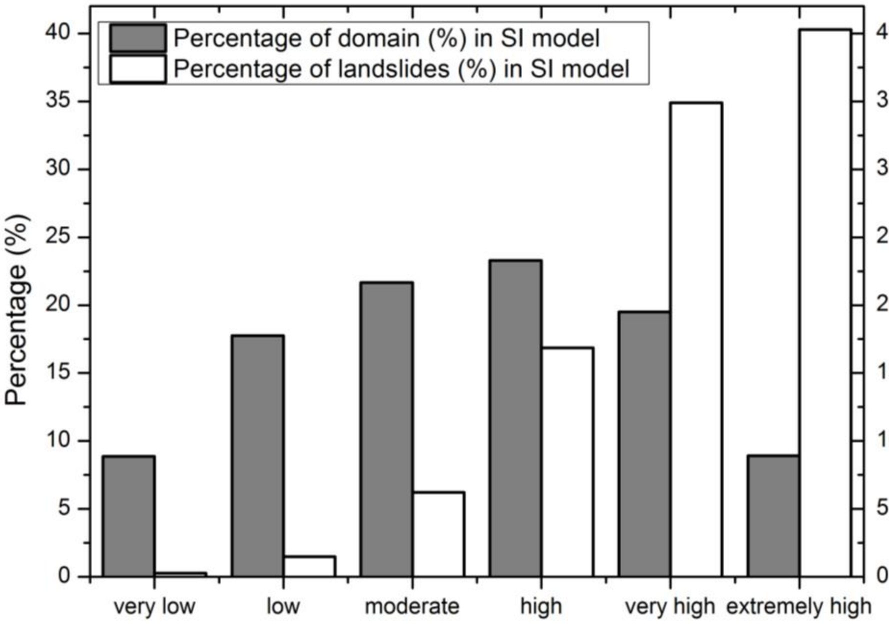

Figure 6-22 Susceptibility class distribution within the study area and the occurrence of landslides according to the classification scheme for LSM using the SI method with the selected seven factors.

Table 6-15 Result of statistics analysis concerning landslide susceptibility from the SI method with the selected seven factors.

Class	Area of each class	Percentage of domain （%）	No. of landslides	Percentage of landslides (%)
Very low	26388	8.86	6	0.27
Low	52840	17.75	33	1.50
Moderate	64488	21.66	137	6.21
High	69376	23.30	372	16.86
Very high	58052	19.50	770	34.89
Extremely high	26564	8.92	889	40.28

The forward stepwise LR approach was used to incorporate the predictor variables using the SPSS 20 software. The training dataset (1776 of total landslides) represented by points were assigned the value of 1. The same number of non-landslide points were randomly sampled from the landslide-free area and assigned the value of 0.

The result based on all original factors is shown in Table 6-16. According to the table of this logistic regression, it shows that all the causative factors have a P-value less than 0.05, indicating a statistical correlation between factors and the susceptibility of landslides at the 90% confidence level (Tien Bui et al. 2011). Based on equation, the occurrence of landslide probability (P) can be computed as mentioned before.

Lastly, the regression coefficients of the predictors, GIS and the natural break criterion were used to generate the landslide susceptibility maps (Figure 6-23). In the maps there are places where differences are subtle but also areas with obvious dissimilarities. There are more red colors in the map from all factors, which segregate at the very high and extremely high ends of the color ramp than the seven-factor counterpart. The map from the seven factors is less heterogeneous. Figure 6-24 and Table 6-17 show that 95.51% of the total landslides occurred in the 66.73% of the area classified as high, very high and extremely high susceptibilities if the all the original factors were used, while if the optimal seven factors were used 96.1% of the total landslides occurred in the 64.09% of the area classified as high, very high and extremely high susceptibilities (Figure 6-25 and Table 6-18).

Table 6-16 Coefficients, statistics of the factors with all factors used in the LR equation.

Factors	*B*	*S.E.*	Wald	df	Sig.	Exp(*B*)	95% C.I.for EXP(*B*)	
							Lower	Upper
Elevation	-0.50	0.04	187.79	1	0.00	0.61	0.56	0.65
Slope	0.07	0.02	9.50	1	0.00	1.08	1.03	1.13
Aspect	-0.06	0.02	10.75	1	0.00	0.95	0.91	0.98
Plan curvature	0.13	0.08	2.52	1	0.01	1.14	0.97	1.34
Profile curvature	-0.25	0.07	12.86	1	0.00	1.28	1.12	1.47
Curvature	0.08	0.06	1.88	1	0.02	1.08	0.97	1.20
Drainage density	0.03	0.03	0.92	1	0.03	1.03	0.97	1.09
Dist. drainage network	0.21	0.03	42.40	1	0.00	1.24	1.16	1.32
SPI	0.22	0.05	21.43	1	0.00	1.25	1.14	1.37
CTI	-0.45	0.06	55.03	1	0.00	0.64	0.57	0.72
Rainfall	0.01	0.03	0.17	1	0.05	1.01	0.95	1.08
Lithology	0.01	0.02	0.06	1	0.01	1.01	0.97	1.05
Constant	-0.18	0.15	1.48	1	0.02	0.84		

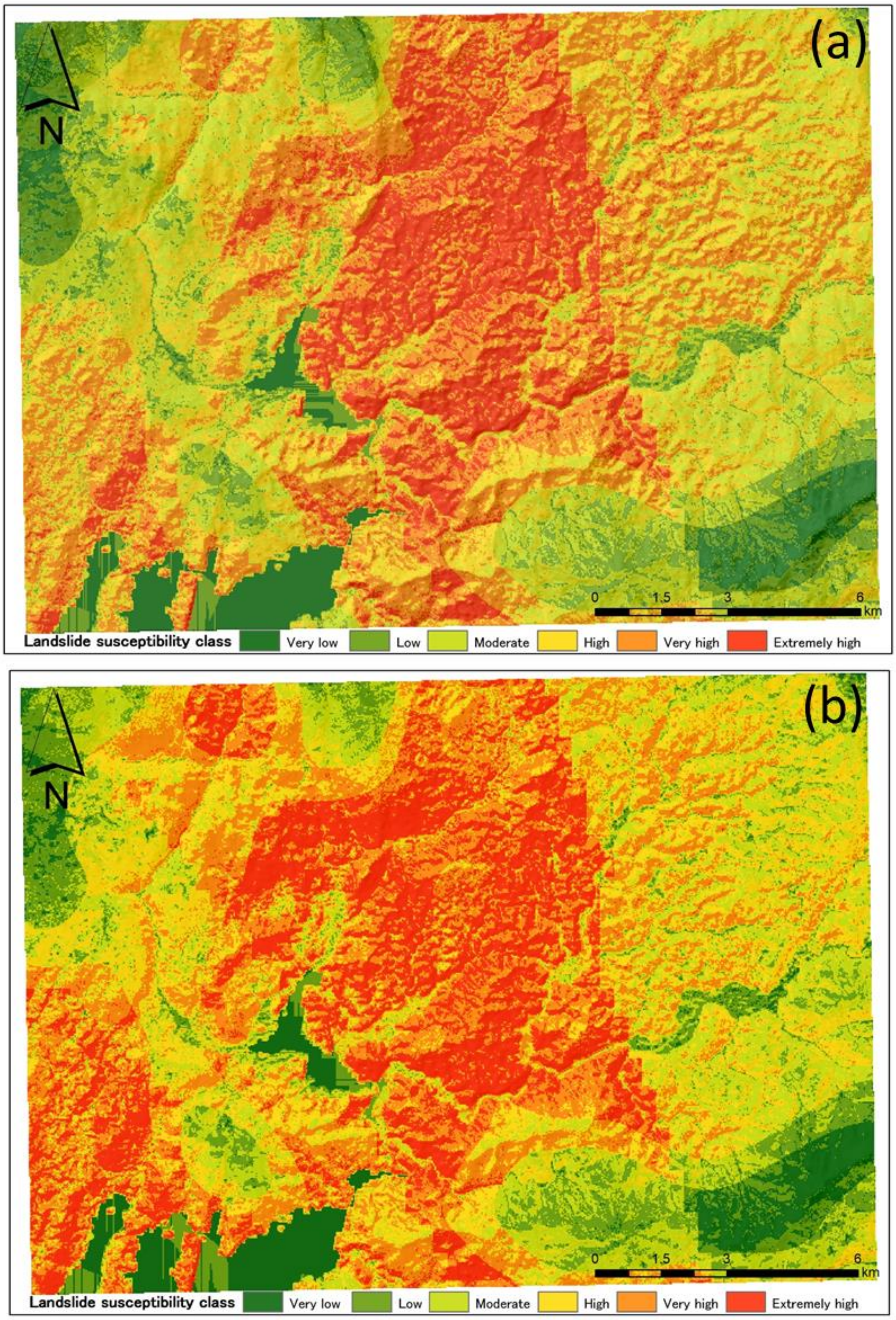

Figure 6-23 LSM maps produced by the LR method: a) selected seven factors, and b) original 12 factors. Maps show the spatial probability of landslide occurrence in six classes.

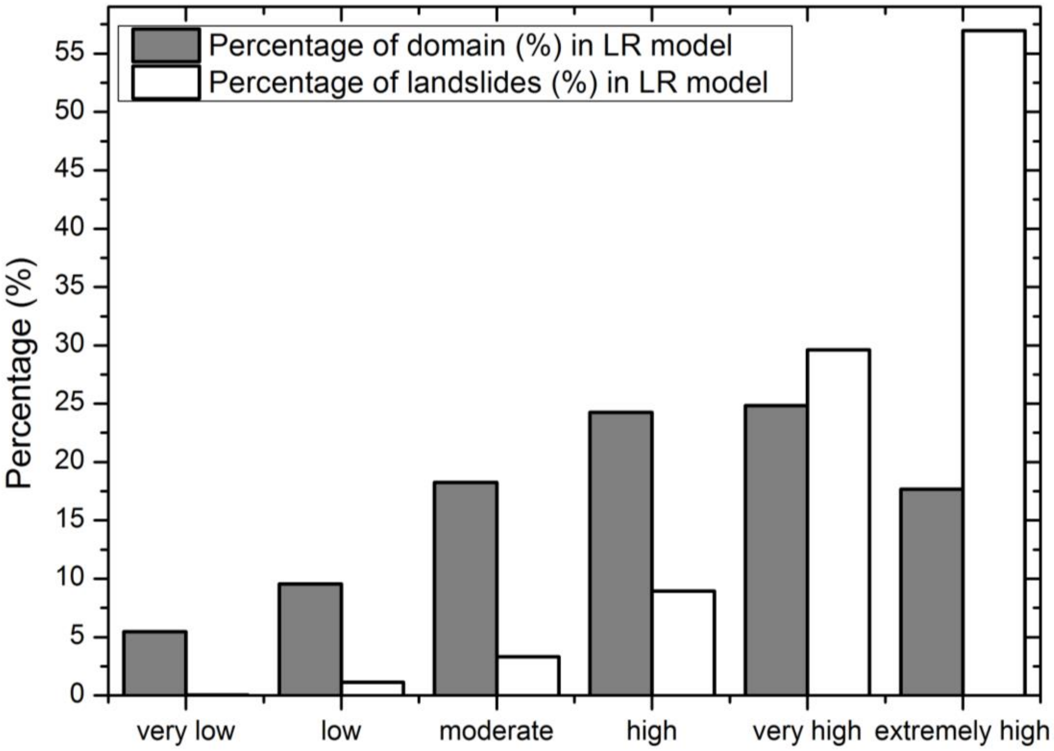

Figure 6-24 Susceptibility class distribution within the study area and the occurrence of landslides according to the classification scheme for LSM using the LR method with the original 12 factors.

Table 6-17 Result of statistical analysis concerning landslide susceptibility from the LR method with the original 12 factors.

Class	Area of each class	Percentage of domain（%）	No. of landslides	Percentage of landslides (%)
Very low	16245	5.46	1	0.05
Low	28468	9.56	25	1.13
Moderate	54336	18.25	73	3.31
High	72201	24.25	197	8.93
Very high	73890	24.82	654	29.63
Extremely high	52568	17.66	1257	56.96

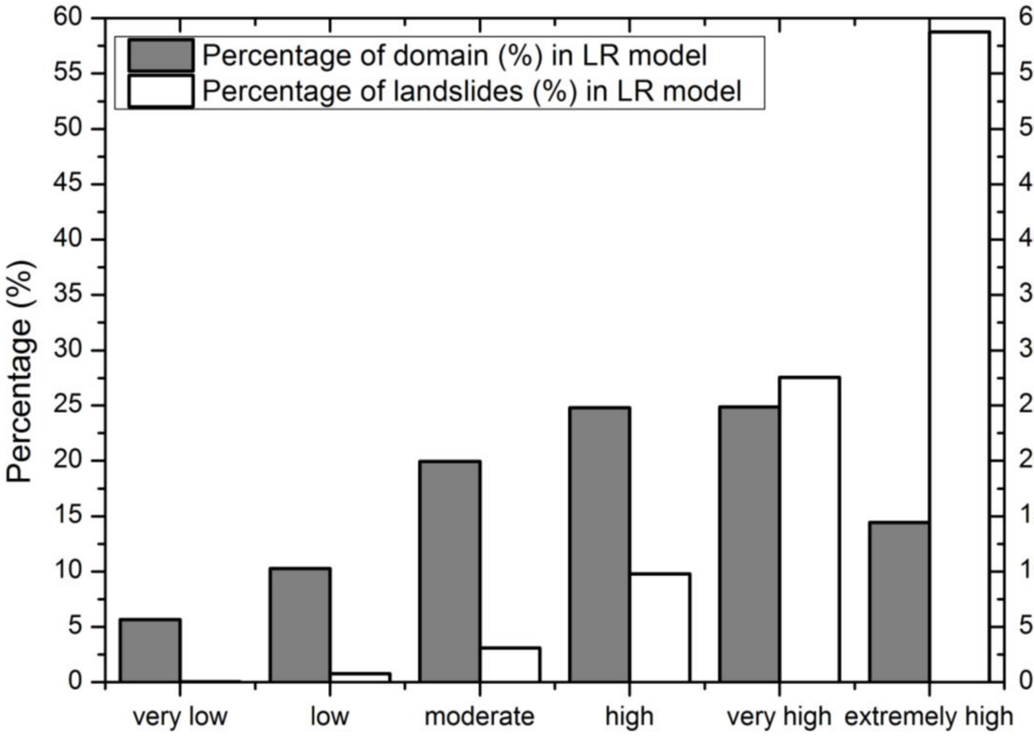

Figure 6-25 Susceptibility class distribution within the study area and the occurrence of landslides according to the classification scheme for LSM using the LR method with selected seven factors.

Table 6-18 Result of statistical analysis concerning landslide susceptibility from the LR method with selected seven factors.

Class	Area of each class	Percentage of domain（%）	No. of landslides	Percentage of landslides (%)
Very low	16889	5.67	1	0.05
Low	30600	10.28	17	0.77
Moderate	59412	19.96	68	3.08
High	73816	24.79	216	9.79
Very high	74014	24.86	608	27.55
Extremely high	42977	14.44	1297	58.77

6.5.3 Accuracy assessment

For the verification, the total landslides were divided into two groups, training data and validation data and an *ROC* plot of sensitivity (true positive rate) and 1-specificity (false positive rate) was made. For the SI method, the *AUC* value (0.837) is higher when optimal seven factors were used than 0.794 from all the original factors (Figure 6-26). For the LR model, the *AUC* value of the prediction rate curve (84.8%) form the seven factors is higher than that from all factors (80.8%) as shown in Figure 6-27. Consequently, using the seven factors give a higher accuracy than using all the original factors. In addition, LR has a slightly higher accuracy than SI.

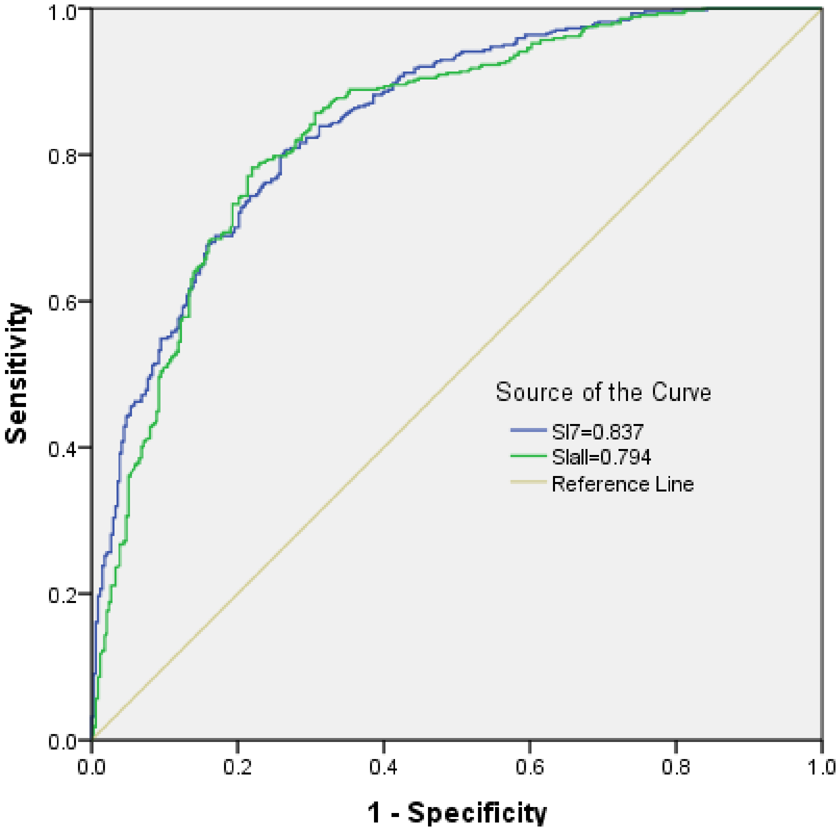

Figure 6-26 ROC curves for landslide susceptibility maps produced using *SI* with the selected seven and original 12 factors for Dongjiang Reservoir.

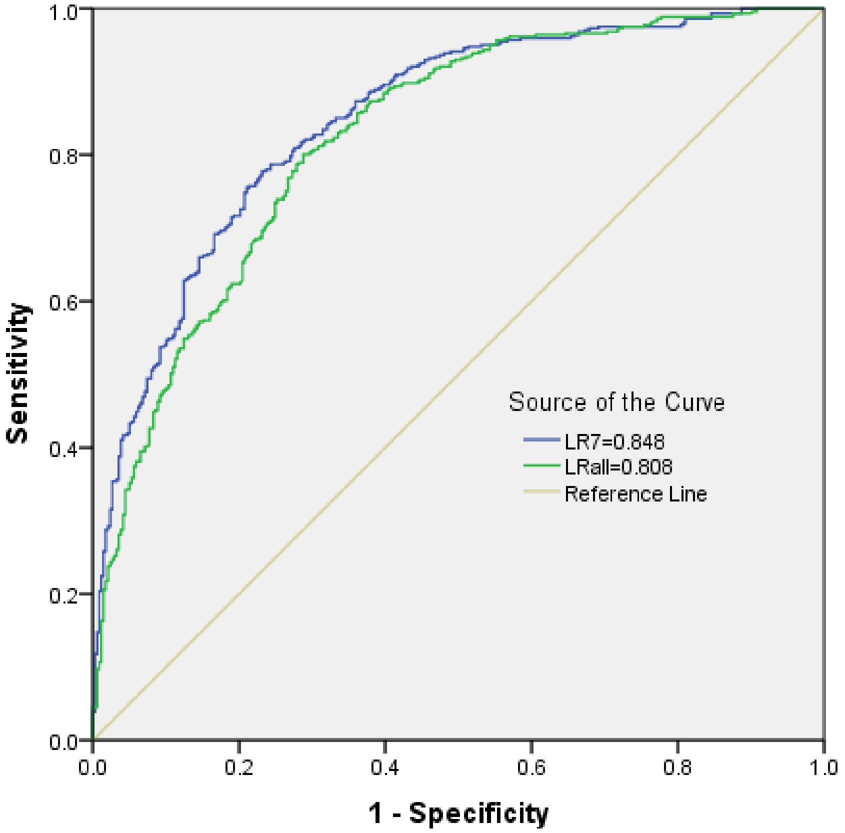

Figure 6-27 ROC curves for landslide susceptibility maps produced using *LR* with the selected seven and original 12 factors for Dongjiang Reservoir.

CHAPTER 7 DISCUSSION

7.1. Completeness of landslide inventory maps

In mountainous terrains, field investigation for landslide inventory mapping is challenging and tedious (Martha et al. 2010). Although direct visual interpretation of remote sensing data is a reliable method, this task is also laborious and time-consuming. New perspectives on detecting landslides more automatically have been facilitated with the increasing availability of high-resolution remote sensing images and improved computer technology in relation to image analysis. In this context, automated detection techniques seem to be suitable for efficiently establishing spatial inventories or databases of landslides that are useful for qualitative/quantitative hazard assessment. Various methods for automated landslide mapping have been proposed; however, few methods have successfully demonstrated ideal results. Nichol and Wong (2005) attempted using SPOT and IKONOS images to map landslides through pixel-based methods. Martha et al. (2010) recently applied OOIA to map landslides semiautomatically through contextual analysis. They empirically used expert knowledge to select features. Although Chang et al. (2012) identified shallow landslides by using high-resolution remote sensing observation data, they did not use a DEM and *NDVI*. For OOIA, applying expert knowledge to characterize landslides plays a key role in semiautomatic landslide detection. However, selecting features related to a landslide occurrence is highly subjective because of the high dependence on personal knowledge. Hence, an appropriate feature selection of landslide types is required. In this study, the information of objects was created through the multi-segmentation of fine resolution QuickBird (0.61 m) multispectral data and 5-m resolution elevation data, and major landslides were successfully characterized as old and young landslides.

In this thesis multi-segmentation was performed to extract the landslide 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. (2014) built a complicated rule set to detect and delineate landslides by using OOIA, but it 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 4-2). 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 to a bright appearance of landslide areas in the satellite image, as shown in Figure 7-1. To reduce such false positive classification in the future, we must integrate topographical GLCM that can substantially reduce the misclassification of objects (Blaschke et al. 2014). 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.

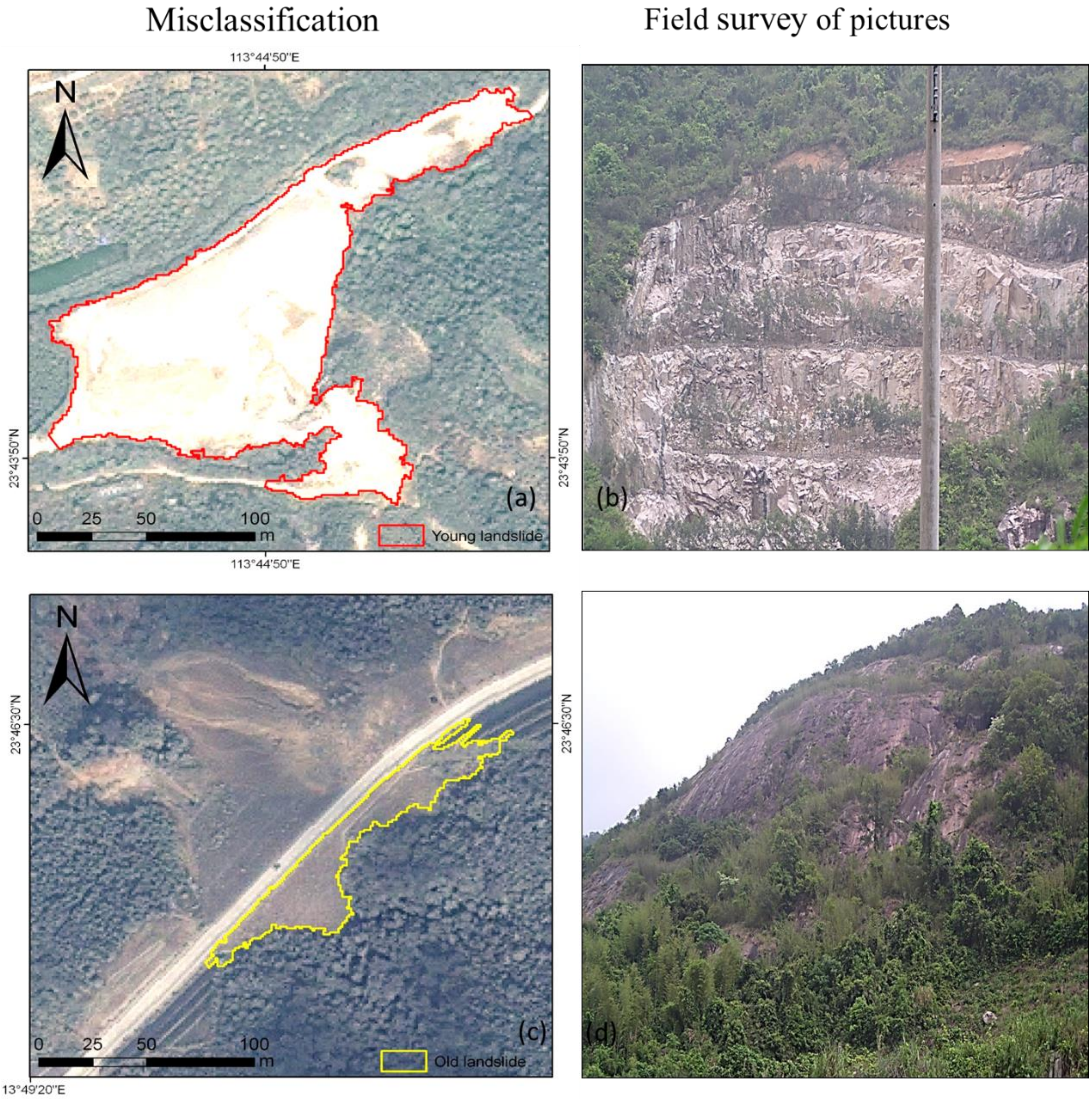

Figure 7-1 Examples of misclassification: (a) bare quarry misclassified as a young landslide, (b) related field photo; (c) rock outcrop with sparse forest misclassified as an old landslide, and (d) related field photo. The tones of misclassification were similar, with a bright appearance of the landslide area in the satellite image.

7.2. Importance of landslide classification in susceptibility mapping.

Active geological processes like landslides play an important role in reshaping topography (van Westen et al. 1997; Farrokhzad et al. 2011; Tarolli 2014; Li et al. 2014). They also cause serious natural disasters; about one quarter of the natural disasters in the world seems to be directly or indirectly related to landslides (Sassa et al. 2006; Miyagi et al. 2011; Huang et al. 2012). The risks posed by landslides differ according to their types (shallow or deep-seated) and therefore “when, where and what type” of landslides are important aspects of susceptibility mapping. The problem is especially critical in developing countries where warning and protection measures are difficult to implement due to the limitation of economic conditions (Kojima et al. 2000). The risk imposed as well as the extent of the damage caused by a shallow landslide is usually local whereas deep-seated landslides can have regional implications due their large sheer size and volumes.

We found that the variations in topographic and geographic factors used for evaluating landslide susceptibility are equally useful in predicting and differentiating shallow and deep-seated landslides. This showed that the forces behind the occurrence of different landslide types are intrinsically different and that the parametric differences of the landslide causative factors can classify them into the two types. In a similar previous study, Kojima et al. (2000) used different landslide types as independent training datasets for prediction maps according to landslide types; however, this thesis has successfully combined these two analytical aspects (prediction and differentiation). This will provide an important frontier to discuss the geomorphological evolution of hillslopes. Moreover, the prediction of landslide types will be a useful guidance for geotechnical engineers and the local government managing and mitigating local hazards.

7.3. Significance of causative-factor optimization.

A prior knowledge on appropriate causative factors related to landslide events are required to map landslide susceptibility (Guzzetti et al. 1999). Several studies in the past have shown that a manual selection of the causative factors by a subject specialist was considered the best approach, but it is rather subjective. Indeed, so far there is no general criteria or guidelines available on how to identify and select a set of landslide causative factors. Due to this fact, numerous scholars have used varied numbers of different causative factors to produce landslide hazard maps. Sometimes even 20 to 60 factors have been used for building discriminant susceptibility models (Guzzetti et al. 1999). Nevertheless, most frequently 10–15 factors were used based on abundance and accessibility (Lee et al. 2008). Hence, it seems to be possible to narrow down the factors based on the knowledge of triggering mechanism involved. For instance, for earthquake induced landslides, the triggering factors associated are no way related to precipitation and their varieties, but are linked to ground acceleration and its intensity. In such a case, it is a common understanding that one can easily omit those unnecessary factors in the analysis. However, when the triggering mechanism is not single or unknown, where the landslide inventory database were created from multiple imageries in different periods of time, the screening out process requires statistical or computational models. Lee et al. (2008), on computing the standardized difference of causative factors, screened six factors out of 14 for landslide susceptibility mapping in some parts of Taiwan. Although this method includes less computation, it requires to categorize the data into landslide and non-landslide groups which is further tedious. Similar statistical equations based on correlation or association indexes limit the predictive performances on multivariate models. On the other hand, Costanzo et al. (2012) identified the factors based on the ranks associated with each factor's expected contribution to the predictive skill of a multivariable model. Approaches adopting discriminant analysis and LR on the forward selection of variables, however, fail when most of the variables are statistically significant. For the same reason, this study did not consider the stepwise LR model because we found most of the variables are significant in the statistical tests ($p < 0.1$). As indicated in the results, none of the factors were screened with the stepwise LR model. Furthermore, stepwise LR model in landslide susceptibility assessment requires both landslide and non-landslide pixels in the calculation. The proposed model using *CF* eliminated these limitations because it used only landslide pixels in the computation, and hence is very fast. Prior definition of hazard classes is not required in the *CF* approach and it also supplies advantage of rendering the definition of susceptible classes transparent. Moreover, the proposed model is a relatively straightforward method that allows the causative factors to be ranked according to their certainty values in the range between -1 and 1. It is assumed that a positive *CF* values have a strong influence on the landslide occurrence, and vice versa. As shown in the result, and the criterion discussed in Section 6.3, six to seven causative factors were finally identified and they were ranked based on their *CF* values. We believe, *CF*-based factor screening process for the identification of the most determinant factors is an important step in the landslide hazard mapping.

The results of our analyses demonstrated that a larger number of causative factors are not necessarily to obtain a better landslide predictability map. This is probably either because of the data redundancy or spatial self-correlation with the study area. For instance, in Sado Island, one of the causative factors, *NDVI*, has no significant effect on the landslide occurrence in this study, as most of the landslides were large. Relatively short roots of the vegetation cover do not considerably influence large landslides and should be obvious to our understandings. In addition, geology and faults may have a positive influence on triggering deep-seated landslides. As demonstrated in the previous study (Ayalew et al. 2005b; Dou et al. 2015g; Dou et al. 2015f) , landslide activity is mostly concentrated in the lithologies dominated by volcanic dacites, and volcanic andesites, followed by volcanic dacite and sandstone. Volcanic dacite and andesite are characterized by a high silica and alumina content and low in potash. They generally have relatively low shear strength and are strongly fractured, resulting in most concentrated landsliding in these rocks. Furthermore, slopes consisting of these lithologies are relatively steeper and more susceptible to failure. Some authors (Tien Bui et al. 2011; Devkota et al. 2013) invoked faults as the triggering cause of many deep-seated landslides. Ayalew et al. (2005) reported the presence of active faults in Sado Island that could potentially trigger landslides. This is in agreement with our results as confirmed from the *CF* analysis.

Although the method proposed in this study has not been tested at other sites, there are indications, which suggests its applicability to other landslide prone regions. Firstly, notwithstanding the fact that CF methods have seldom been used in identifying causative factors in landslide susceptibility mapping, they are used worldwide for managing uncertainty in rule-based systems. Because of their favorability functions to handle different data layers and the heterogeneity and uncertainty of the data, CF models are largely appreciated in slope stability studies (Binaghi et al. 1998; Tien Bui et al. 2011; Pourghasemi et al. 2013a; Devkota et al. 2013).

7.4. Comparative analysis of landslide susceptibility models

Predication of the precise locations of the instabilities for landslide susceptibility assessment is rather difficult owing to the uncertainty of the spatial and temporal distribution of rainfall. Susceptibility assessments may also be influenced by other important factors such as geology, slope aspect, and slope angle. These issues are commonly addressed by GIS-based landslide susceptibility studies. Different researchers utilize various landslide conditioning factors for LSM. Though the selection of factors is a fundamental step for landslide susceptibility evaluation, universal standard or rule to select the conditioning factors is absent (Ayalew and Yamagishi 2005).Selecting factors is a fundamental step for landslide susceptibility evaluation and influences the result of LSM. To address this problem, we propose the CF method to select important factors. Using this method we selected the conditioning factors highly related to landslide occurrence. The resultant improvement in the values of *AUC* validate our approach. The use of the optimized factors led to a higher accuracy than when all possible factors were simultaneously used. Spatial auto-correlation and data redundancy among the conditioning factors before optimization is the possible cause for this observation.

Additionally, among the factors for the three study areas in China and Japan as shown in Figure 7-2, we found that slope angle and lithology are the most important. This finding agrees with the other studies around the world (Can et al. 2005; Solaimani et al. 2013; Zare et al. 2013b; Xu et al. 2013; Fuchs et al. 2014). For instance, the slope angle is regarded as a very important factor in landslide research (Keefer 1994; Lu and Rosenbaum 2003; Fourniadis et al. 2007; Vita et al. 2012; Zare et al. 2013b), because it affects slope stability against gravity and hillslope hydrology such as sub-surface flow velocity. Lithology also plays an important role in the occurrence of landslides in relation to rock permeability and strengths (Yamagishi et al. 2004; Ayalew and Yamagishi 2005; García-Rodríguez and Malpica 2010; Pourghasemi et al. 2013b; Dou et al. 2015g). Thus, these landslide factors may be common to various areas in the world. We believe that our research findings are universal and provides a method to select and qualify the landslide-conditioning factors.

Landslides in Dongjiang reservoir watershed in China and the Chuetsu area in Japan differ in their triggering mechanisms. Landslides in Dongjiang were rainfall-induced while the same in Chuetsu were earthquake-triggered. In the Chinese study area, intense rainfall caused slope failures in areas of severely weathered-granite resulting in numerous shallow landslides. The permeability and drainage characteristics of the area also affected the distance travelled by the of landslide debries. Total curvature and distance to drainage network were found to be important in the Dongjian area along with the other common factors. Total curvature represents the morphological measurement of the topography (Lee and Sambath 2006). A more upwardly concave or convex slope holds more water and keeps it longer (Lee et al. 2004a) and these hydrological controls of topography are more expressed in mountainous areas and lower in the flat areas. Likewise, factors supplementing ground hydrology, distance to drainage network and rainfall, were also found to be important in the Dongjiang reservoir watershed area.

Devastative landslides as a result of intense-rainfall are common in many places around the world. For example, on 20th August, 2015, extreme rainfall and the resultant serious debris-flow caused 44 injuries, 74 deaths and severely damaged 296 houses in Hiroshima City of Japan (Chigira 2001; Wang et al. 2015). On 4th July, 2013, high intensity rainfall induced by a torrential rainstorm in Sichuan Province, China, triggered a debris flow that caused 19 deaths, and additionally, transported a large amounts of sediments to the Zhuma

River and the Nanya River, generating three debris dams that blocked the two rivers temporarily (Ni et al. 2014). Similar recent hydro-geological disasters make mitigation efforts an urgent necessity. Our study of rainfall-induced landslides in the Dongjiang reservoir watershed may be applicable to many similar cases.

An earthquake was the primary trigger for the landslides in the Chuetsu area. This study found that alongside the factors common to the other study areas, density of geological boundaries was found to be an important factor for the area. Geological boundaries depict geological heterogeneity of an area and its increasing density signifies a decreasing slope cohesiveness. The higher density of geological boundaries means lower stability due to the weak zones that can lead to an increase in landslide occurrences (Kawabata and Bandibas 2009). The earthquake led to numerous slip along such weak zones. Plan curvature was also found to be an important factor in the study of landslides at Chuetsu. It delineates the morphology of the topography and governs the divergence or convergence of surface flow (Parsons 1979; Ohlmacher 2007; Conforti et al. 2010; Peckham 2011; Conforti et al. 2014). However, in the case of Chuetsu, not flow but high soil moisture along concave topography may have affected landslide occurrence. Ohlmacher (2007) demonstrated that any study of landslide susceptibility needs to account for the complex relationship between plan curvature and landslide susceptibility.

Drainage density represents fluvial erosion by streams but can be related to landslides. For example, drainage density and erosion rates in steep Japanese mountains are negatively correlated due to active landslides (Oguchi 1997). Several scholars have therefore studied the inter-relationship of landslides and geomorphological characteristics of drainage networks (Benda and Dunne 1997; Hovius et al. 1998; Nath 2004; Lee and Talib 2005; Ng 2006; Schwab et al. 2008; Devkota et al. 2013; Santangelo et al. 2013; Conforti et al. 2014). Drainage density is an important factor both in the Chuetsu area and the Dongjiang reservoir watershed, perhaps because these two study areas have major rivers. In the Dongjiang reservoir watershed, the Dongjiang river flows into the reservoir, while in the Chuetsu area, rivers such as the Shinano and the Uono cross over this area, hence, the drainage density in these areas changed largely place by place (Yoong and Okada 2005; Ayalew and Yamagishi 2005; Wang and Zhao 2006). A lot of landslides have been observed along the eastern side of the Shinano River near its confluence with the Asahi River (Kieffer et al. 2006). However, in Sado Island, there is no big river flowing across the study area and thus this causative factor is not important.

Details about the triggering mechanism of landslides in Sado Island are unavailable but rainfall and heavy snowmelt are reported to be the primary reasons (Ayalew et al. 2005a; Ayalew et al. 2005b; Yamagishi 2008). Besides the factors common to the three study areas, aspect was found to be an important factor for Sado island. Slope-aspect is related to the meteorological parameters like precipitation asymmetry, exposure to sunshine, and wind velocity (Ercanoglu et al. 2008; Aksoy and Ercanoglu 2012) and hence an important causative factor in LSM (Carrara 1983; Van Den Eeckhaut et al. 2009b; Bijukchhen et al. 2013). Most landslides in Sado occurred on the southern and southwest slopes where the snow is easy to melt. The periodic loading/melting of snow results in variation of soil water content and over-burden stress. These factors over time gradually decrease the stability of slope forming materials (Dou et al. 2015f). In the Chuetsu areas, landslides mostly occurred on the hanging-wall side of active faults, and thus slope aspect became important (Kieffer et al. 2006; Has et al. 2010). These situations do not occur in the Chinese study area at last typically.

Neotectonic movements like earthquakes often create “broken zones” along the huge faults segments in superficial deposits on the higher ground consisting mostly of colluvium. Such weak zones are favorable for landslides (Liu et al. 2004). The major structural discontinuities and shear zones generated by the faults are primarily causative factors in the analysis of landslide susceptibility (Sarkar and Kanungo 2004; Collins et al. 2012; Tien Bui et al. 2012a; Zare et al. 2013b). The distance to faults was found to be a major causative factors in Sado Island (Dou et al. 2015g) because of active faulting there. From these three case studies in China and Japan we found that the triggering mechanisms dictates the causative factors relevant for that kind of landslide susceptibility mapping.

In this study, several traditional models, such as statistical index (SI), logistic regression (LR), frequency ratio (FR), and weight of evidence (WOE) are compared with the help of respective AUC values. We found that logistic regression has a better performance than the others as shown in Figure 7-3. This conclusion is also in a good agreement with the other researchers around the world (Dai et al. 2004; Gorsevski and Gessler 2006; Chen and Wang 2007; Devkota et al. 2013) .

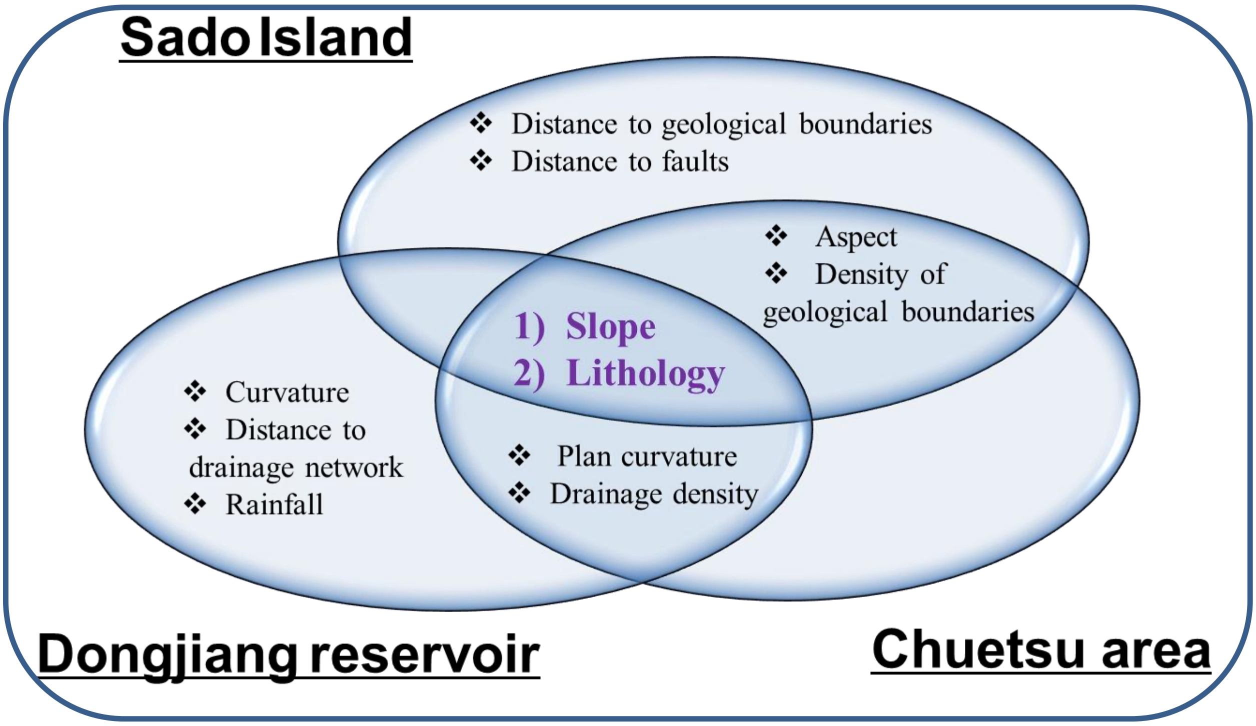

Figure 7-2 Important collective factors for the Chuetsu area, Sado Island and the Dongiang reservoir watershed.

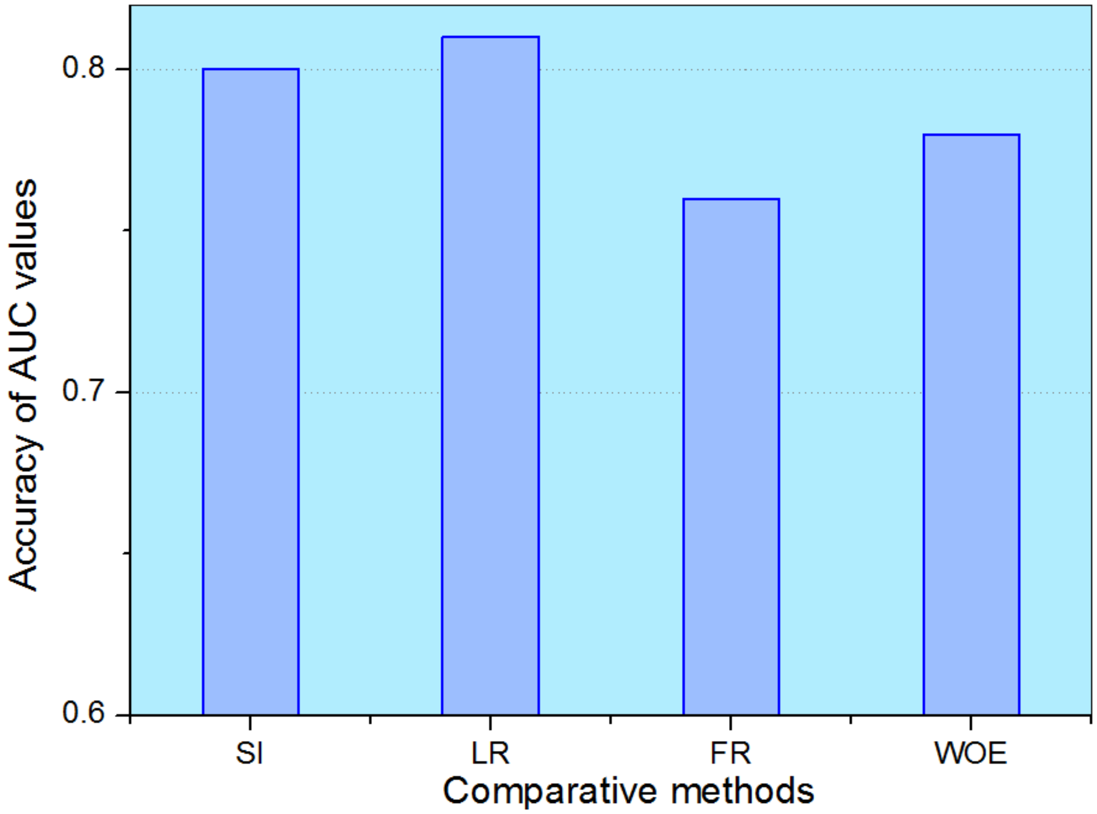

Figure 7-3 Comparative different models for LSM by AUC values.

CHAPTER 8 CONCLUSIONS

This study developed an integrated approach for detecting and characterizing landslide features to automatically construct a landslide inventory in a district in the PRD, China. The proposed model incorporates each method's advantage and eludes some of the problems such as bottlenecks of knowledge-based selection in creating an EKS. The established inventory is the foundation of forecasting the spatial and temporal distributions of future landslides. The predictive mapping of landslide susceptibility can be conducted based on independent landslides with a satisfactory consistency. This valuable knowledge is suitable for disaster management to identify landslide-prone areas, mitigation after the occurrence of landslide disasters, and prevent future unnecessary economic losses.

Additionally, this study differentiated the landslide types (i.e., shallow and deep-seated landslides) using an SVM model with the assistance of a 2-m high resolution Lidar DEM. The outcomes with high accuracy suggest that our assumptions employed are valid. The existing landslides matched the predictions in most cases. SVMs also performed better than ANN (BP) in terms of model stability and accuracy with the relatively small training samples. Among the four SVM kernels, RBF was selected after a comparative test. Besides, reduction in the size of the training dataset from 50% to 30% of the total dataset did not significantly affect the accuracy of the SVMs model, confirming that SVMs work even with a smaller training dataset. However, we found that a higher resolution DEM is necessary for studying the details of shallow landslides. Active geological processes like landslides play an important role in reshaping topography. Therefore, differentiating the types of landslides is important for discussing the geomorphological evolution of hillslopes, and also for supporting the local government managing and mitigating local hazards.

Further studies using not only a finer DEM, but also other detailed information such as the peak ground acceleration (*PGA*) and volume of landslides are necessary.

This study also demonstrates the usefulness of the CF model in identifying the best fitted causative factors for landslide susceptibility mapping. Based on the CF model, six or seven influencing factors with the high correlations to landslide occurrence were selected from a set of original factors. The LSM maps were then produced by applying both the SI and LR methods for the CF-identified causative factors and the original set of factors. Both the success rate and prediction rate indicated for both the SI and LR methods that the six or seven factors achieve better results than that of all factors. In addition, we noticed that the maps prepared from using six or seven causative factors have much more homogeneous classes than the original factors. We found that they all the three case studies include some common collective causative factors, such as slope angle, and lithology. These common factors may be applied to the other similar study areas. The proposed method provides a useful way to select the controlling factors of landslides in particular where data redundancy or scarcity is critical. The findings acknowledge that in the mountainous regions suffering from data scarcity, it is possible to select key factors related to landslide occurrence based on the CF models in a GIS platform. Moreover, in this research, LR has slightly outperformed the others among the traditional methods, which agrees with results from some other researchers in the world.

We assume that the results of our studies provide helpful information for disaster management, urban planning, risk mitigation and related decision making in landslide-prone areas. For example, in the study areas, the resultant landslide susceptibility maps can be conducive to select appropriate locations for urban development to increase economic benefits and decrease future damages and loss of lives.

In the future, to further understand the geomorphic process and landscape evolution related to landslides in steep mountainous regions, auxiliary research is required concerning geomorphic, topographic, anthropogenic, hydrologic, and geologic settings of landslide sources. Additionally, to quantitatively asses the influence of environmental changes on slope stability, physical-based or process-based models such as, TRIGRS, SHALSTAB, SINMAP and SHETEAN models (Santini et al. 2009; Chien et al. 2015) need to be used to simulate landslides. Moreover, the relatively new models, such as the random forest (RF), deep learning (DP) algorithms will also develop to generate the landslide susceptibility maps and enable hazard warnings more precisely.

ACKNOWLEGEMENTS

This dissertation would not have been possible without the help of so many people in countless ways. First and foremost, I would like to express my deepest gratitude to my supervisor Prof. Takashi Oguchi for his kind consideration and constructive comments over of last four years. He sets a good example for us on how to be a scientist, not to be a technician. Without his guidance and persistent help this dissertation would not have been possible.

In addition, I would like to express my sincere gratitude to the committee members of this thesis: Prof. Toshihiko Sugai, Prof. Kaoru Saito, Assoc. Prof. Juichiro Ashi, and Assoc. Prof. Yuichi S. Hayakawa for their valuable comments and suggestions that have helped in improving this manuscript.

Assoc. Prof. Yuichi S. Hayakawa also provided a chance of field survey. I also thank to my lab members for their mental and research support, especially Dr. Yunus Ali. P., Mr. Uttam Paudel and Mr. Chen Chiwen. Herein, I extend the friendship.

During the years, I have opportunities to collaborate with the other excellent researchers. I acknowledge Prof. Hiromitsu Yamagishi for providing the aerial photographs and landslide data. I also acknowledge my sincere gratitude towards Mr. Shoichiro Uchiyama at NIED for providing the Chuetsu landslide data. I express my gratefulness to the former supervisor, Prof. Xia Li for his instructions and suggestions on a paper. Additionally, I am indebted to Dr. Yueren Xu from the China Earthquake Administration and Drs. Yi Zou and Zhongfan Zhu from the Beijing Normal University for their generous data collection and assistance. I thank my former classmate, Dr. Yi Zou at University of Chinese Academy of Sciences for providing data for the Chinese study area. Additionally, I am grateful for Drs. Kuan-Tsung Chang and Jin-King Liu offering some data of Taiwan for future collaboration .

Moreover, I appreciate the Japanese Government (Monbukagakusho-MEXT) scholarship provide by the Ministry of Education, Culture, Sports, Science, and Technology, Japan for supporting my study and daily life.

Last but not the least, I highly express my special sincere cheers to family and friends for their constant understanding and continuous encourage, especially for my parents, wife Zhang Xiaorui and my son that they understand, support and tolerate me to spend almost of my time on the research work in the University of Tokyo, particularly when I felt hopeless and depressed of my future. Their love is an inspiration for my research. I dedicate this work to them.

Life is not easy for any of us, herein I highly express my heartfelt thanks to the people mentioned above, and many more, have together contributed to what has been for me an enormously valuable experience, and hence, there is a kind of sadness in leaving. Nevertheless, I will take many joyful memories of the University of Tokyo with me.

REFERENCES

Aamodt A, Plaza E (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun 7:39–59.

Addis B, Cassioli A, Locatelli M, Schoen F (2011) A global optimization method for the design of space trajectories. Comput Optim Appl 48:635–652. doi: 10.1007/s10589-009-9261-6

Ahmed B (2014) Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. Landslides. doi: 10.1007/s10346-014-0521-x

Ahn H, Kim K, Han I (2006) Hybrid genetic algorithms and case-based reasoning systems for customer classification. Expert Syst 23:127–144.

Aksoy B, Ercanoglu M (2012) Landslide identification and classification by object-based image analysis and fuzzy logic: An example from the Azdavay region (Kastamonu, Turkey). Comput Geosci 38:87–98. doi: 10.1016/j.cageo.2011.05.010

Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. Bull Eng Geol Environ 58:21–44. doi: 10.1007/s100640050066

Anbazhagan S, Ramesh V (2014) Landslide hazard zonation mapping in ghat road section of Kolli hills, India. J Mt Sci 11:1308–1325. doi: 10.1007/s11629-012-2618-9

Anders NS, Seijmonsbergen AC, Bouten W (2011) Segmentation optimization and stratified object-based analysis for semi-automated geomorphological mapping. Remote Sens Environ 115:2976–2985. doi: 10.1016/j.rse.2011.05.007

Ardizzone F, Cardinali M, Galli M, et al. (2007) Identification and mapping of recent rainfall-induced landslides using elevation data collected by airborne Lidar. 637–650.

Arora MK, Das Gupta a. S, Gupta RP (2004) An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. Int J Remote Sens 25:559–572. doi: 10.1080/0143116031000156819

Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65:15–31. doi: 10.1016/j.geomorph.2004.06.010

Ayalew L, Yamagishi H, Marui H, Kanno T (2005a) Landslides in Sado Island of Japan Part II GIS-based susceptibility mapping with comparisons of results from two methods and verifications. Eng Geol 81:432–445.

Ayalew L, Yamagishi H, Marui H, Kanno T (2005b) Landslides in Sado Island of Japan: Part I. Case studies, monitoring techniques and environmental considerations. Eng Geol 81:419–431. doi: 10.1016/j.enggeo.2005.08.005

Baatz M, Benz U, Dehghani S, Heynen M (2004) eCognition User Guide 4 Definiens Imagine GmbH, Munich, Germany.

Bahsan E, Liao HJ, Ching J, Lee SW (2014) Statistics for the calculated safety factors of undrained failure slopes. Eng Geol 172:85–94. doi: 10.1016/j.enggeo.2014.01.005

Bai S, Lü G, Wang J, et al. (2010) GIS-based rare events logistic regression for landslide-susceptibility mapping of Lianyungang, China. Environ Earth Sci 62:139–149. doi: 10.1007/s12665-010-0509-3

Benda L, Dunne T (1997) Stochastic forcing of sediment supply to channel networks from landsliding and debris flow. Water Resour Res 33:2849–2863. doi: 10.1029/97WR02388

Benz UC, Hofmann P, Willhauck G, et al. (2004) Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J Photogramm Remote Sens 58:239–258. doi: 10.1016/j.isprsjprs.2003.10.002

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci J 24:43–69.

Bijukchhen SM, Kayastha P, Dhital MR (2013) A comparative evaluation of heuristic and bivariate statistical modelling for landslide susceptibility mappings in Ghurmi-Dhad Khola, east Nepal. Arab J Geosci 6:2727–2743. doi: 10.1007/s12517-012-0569-7

Binaghi E, Luzi L, Madella P, et al. (1998) Slope Instability Zonation: a Comparison Between Certainty Factor and Fuzzy Dempster–Shafer Approaches. Nat Hazards 17:77–97. doi: 10.1023/A:1008001724538

Blaschke T (2010) Object based image analysis for remote sensing. ISPRS J Photogramm Remote Sens 65:2–16. doi: 10.1016/j.isprsjprs.2009.06.004

Blaschke T, Feizizadeh B, Daniel H (2014) 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 7:1–12. doi: 10.1109/JSTARS.2014.2350036

Brabb EE (1984) Innovative approaches to landslide hazard mapping. Proc 4th Int Symp Landslides 1:307–324.

Burges C (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 2:121–167.

Can T, Nefeslioglu H a., Gokceoglu C, et al. (2005) Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. Geomorphology 72:250–271. doi: 10.1016/j.geomorph.2005.05.011

Caniani D, Pascale S, Sdao F, Sole A (2007) Neural networks and landslide susceptibility: a case study of the urban area of Potenza. Nat Hazards 45:55–72. doi: 10.1007/s11069-007-9169-3

Carrara A (1983) Multivariate models for landslide hazard evaluation. J Int Assoc Math Geol 15:403–426. doi: 10.1007/BF01031290

Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat Hazards Earth Syst Sci 13:2815–2831. doi: 10.5194/nhess-13-2815-2013

Chang C, Lin C (2013) LIBSVM : A Library for Support Vector Machines. 1–39.

Chang C-C, Lin C-J (2011) LIBSVM. ACM Trans Intell Syst Technol 2:1–27. doi: 10.1145/1961189.1961199

Chang K, Chiang S, Chen Y, Mondini AC (2014) Modeling the spatial occurrence of shallow landslides triggered by typhoons. Geomorphology 208:137–148. doi: 10.1016/j.geomorph.2013.11.020

Chang K-T, Liu J, Wang C (2012) An Object-oriented Analysis for Characterizing the Rainfall-induced Shallow Landslide. J Mar Sci Technol 20:647–656. doi: 10.6119/JMST-012-0430-2

Chang T-C, Chao R-J (2006) Application of back-propagation networks in debris flow prediction. Eng Geol 85:270–280. doi: 10.1016/j.enggeo.2006.02.007

Chen WT, Li XJ, Wang YX, Liu SW (2013) Landslide susceptibility mapping using LiDAR and DMC data: a case study in the Three Gorges area, China. Environ Earth Sci 70:673–685. doi: 10.1007/s12665-012-2151-8

Chen Z, Wang J (2007) Landslide hazard mapping using logistic regression model in Mackenzie Valley, Canada. Nat Hazards 42:75–89. doi: 10.1007/s11069-006-9061-6

Cheng C, Lee C, Lin P, et al. (2010) Probabilistic Earthquake Hazard in Metropolitan Taipei and Its Surrounding Regions. 21:429–446. doi: 10.3319/TAO.2009.11.11.01(TH)1.

Chi MM, Feng R, Bruzzone L (2008) Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. Adv Sp Res 41:1793–1799. doi: 10.1016/j.asr.2008.02.012

Chien L-K, Hsu C-F, Yin L-C (2015) Warning Model for Shallow Landslides Induced by Extreme Rainfall. Water 7:4362–4384. doi: 10.3390/w7084362

Chigira M (2001) Micro-sheeting of granite and its relationship with landsliding specifically after the heavy rainstorm in June 1999, Hiroshima Prefecture, Japan. Eng Geol 59:219–231.

Chigira M, Yagi H (2005) Geological and geomorphological characteristics of landslides triggered by the 2004 Mid Niigta prefecture earthquake in Japan. Eng Geol 82:202–221.

Chung C-J, Fabbri AG (1993) The representation of geoscience information for data integration. Nonrenewable Resour 2:122–139. doi: 10.1007/BF02272809

Collins BD, Kayen R, Tanaka Y (2012) Spatial distribution of landslides triggered from the 2007 Niigata Chuetsu–Oki Japan Earthquake. Eng Geol 127:14–26.

Conforti M, Aucelli PPC, Robustelli G, Scarciglia F (2010) Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). Nat Hazards 56:881–898. doi: 10.1007/s11069-010-9598-2

Conforti M, Pascale S, Robustelli G, Sdao F (2014) Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). Catena 113:236–250. doi: 10.1016/j.catena.2013.08.006

Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37:35–46. doi: 10.1016/0034-4257(91)90048-B

Conoscenti C, Ciaccio M, Caraballo-arias NA, et al. (2014) Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines : A case of the Belice River basin (western Sicily , Italy). Geomorphology. doi: 10.1016/j.geomorph.2014.09.020

Conoscenti C, Ciaccio M, Caraballo-Arias NA, et al. (2015) Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the Belice River basin (western Sicily, Italy). Geomorphology 242:49–64. doi: 10.1016/j.geomorph.2014.09.020

Costanzo D, Rotigliano E, Irigaray C, et al. (2012) Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: application to the river Beiro basin (Spain). Nat Hazards Earth Syst Sci 12:327–340. doi: 10.5194/nhess-12-327-2012

Cross M (1998) Landslide susceptibility mapping using the Matrix Assessment Approach: a Derbyshire case study. Geol. Soc. London, Eng. Geol. Spec. Publ. Geological Society, pp 247–261

Crowther PS, Cox RJ (2006) Accuracy of Neural Network Classifiers as a Property of the Size of the Data Set. In: Gabrys B, Howlett R, Jain L (eds) Knowledge-Based Intell. Inf.

Eng. Syst. Springer Berlin Heidelberg, pp 1143–1149

Dai FC, Lee CF (2003) A spatiotemporal probabilistic modelling of storm-induced shallow landsliding using aerial photographs and logistic regression. Earth Surf Process Landforms 28:527–545. doi: Doi 10.1002/Esp.456

Dai FC, Lee CF, Tham LG, et al. (2004) Logistic regression modelling of storm-induced shallow landsliding in time and space on natural terrain of Lantau Island, Hong Kong. Bull Eng Geol Environ 63:315–327. doi: 10.1007/s10064-004-0245-6

Dai FC, Xu C, Yao X, et al. (2011) Spatial distribution of landslides triggered by the 2008 Ms 8.0 Wenchuan earthquake, China. J Asian Earth Sci 40:883–895. doi: 10.1016/j.jseaes.2010.04.010

Danneels G, Pirard E, Havenith H-B (2007) Automatic landslide detection from remote sensing images using supervised classification methods. 2007 IEEE Int Geosci Remote Sens Symp 3014–3017. doi: 10.1109/IGARSS.2007.4423479

De Waele J, Gutiérrez F, Parise M, Plan L (2011) Geomorphology and natural hazards in karst areas: A review. Geomorphology 134:1–8. doi: 10.1016/j.geomorph.2011.08.001

Deng J-L (1982) Control problems of grey systems. Syst Control Lett 1:288–294. doi: 10.1016/S0167-6911(82)80025-X

Devkota KC, Regmi AD, Pourghasemi HR, et al. (2013) Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. Nat Hazards 65:135–165. doi: 10.1007/s11069-012-0347-6

Dou J, Chang K, Chen S, et al. (2015a) Automatic Case-Based Reasoning Approach for Landslide Detection: Integration of Object-Oriented Image Analysis and a Genetic

Algorithm. Remote Sens 4318–4342. doi: 10.3390/rs70404318

Dou J, Chang K-T, Chen S, et al. (2015b) Automatic Case-Based Reasoning Approach for Landslide Detection: Integration of Object-Oriented Image Analysis and a Genetic Algorithm. Remote Sens 7:4318–4342. doi: 10.3390/rs70404318

Dou J, Li X, Yunus AP, et al. (2015c) Automatic detection of sinkhole collapses at finer resolutions using a multi-component remote sensing approach. Nat Hazards 26:1–24. doi: 10.1007/s11069-015-1756-0

Dou J, Li X, Yunus AP, Paudel U (2015d) Automatic detection of sinkhole collapses at finer resolutions using a multi-component remote sensing approach. Nat Hazards. doi: 10.1007/s11069-015-1756-0

Dou J, Oguchi T, Hayakawa YS, et al. (2014a) Susceptibility Mapping Using a Certainty Factor Model and Its Validation in the Chuetsu Area, Central Japan. Landslide Sci a Safer Geoenvironment 2:483–489. doi: 10.1007/978-3-319-05050-8_65

Dou J, Oguchi T, Hayakawa YSS, et al. (2014b) Susceptibility Mapping Using a Certainty Factor Model and Its Validation in the Chuetsu Area, Central Japan. Landslide Sci a Safer Geoenvironment 2:483–489. doi: 10.1007/978-3-319-05050-8_65

Dou J, Paudel U, Oguchi T, et al. (2015e) Differentiation of shallow and deep-seated landslides using support vector machines: a case study of the Chuetsu area, Japan. Terr Atmos Ocean Sci 26:227–239. doi: 10.3319/TAO.2014.12.02.07(EOSI)

Dou J, Qian J, Zhang H, et al. (2009a) Landslides detection: a case study in Conghua city of Pearl River delta. Second Int Conf Earth Obs Glob Chang 74711K–11. doi: 10.1117/12.836328

Dou J, Tien Bui D, P. Yunus A, et al. (2015f) Optimization of Causative Factors for Landslide

Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan. PLoS One 10:e0133262. doi: 10.1371/journal.pone.0133262

Dou J, Yamagishi H, Pourghasemi HR, et al. (2015g) An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Nat Hazards 1–28. doi: 10.1007/s11069-015-1799-2

Dou J, Zheng X, Qian J, et al. (2009b) Intelligence Based Automatic Detection and Classification of Ground Collapses Using Object-Based Image Analysis Method: A Case Study in Paitan of Pearl River delta. 7146:714623–714623–12. doi: 10.1117/12.813168

Dou J, Zheng X, Qian JJ, et al. (2010) Object-based and case-based reasoning method for ground collapses detection. J Image Graph 15:900–910.

Dragićević S, Lai T, Balram S (2015) GIS-based multicriteria evaluation with multiscale analysis to characterize urban landslide susceptibility in data-scarce environments. Habitat Int 45:114–125. doi: 10.1016/j.habitatint.2014.06.031

Drăguţ L, Tiede D, Levick SR (2010) ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. Int J Geogr Inf Sci 24:859–871. doi: 10.1080/13658810903174803

Dramis F, Sorriso-Valvo M (1994) Deep-seated gravitational slope deformations, related landslides and tectonics. Eng Geol 38:231–243. doi: 10.1016/0013-7952(94)90040-X

Du Y, Wu D, Liang F, Li C (2013) 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. GIScience Remote Sens 50:574–589. doi: 10.1080/15481603.2013.842292

Ercanoglu M, Kasmer O, Temiz N (2008) Adaptation and comparison of expert opinion to analytical hierarchy process for landslide susceptibility mapping. Bull Eng Geol Environ 67:565–578. doi: 10.1007/s10064-008-0170-1

ESRI (2011) ArcGIS Desktop.

Falaschi F, Giacomelli F, Federici PR, et al. (2009) Logistic regression versus artificial neural networks: landslide susceptibility evaluation in a sample area of the Serchio River valley, Italy. Nat Hazards 50:551–569.

Farrokhzad F, Barari A, Ibsen L, Choobbasti A (2011) Predicting subsurface soil layering and landslide risk with Artificial Neural Networks: a case study from Iran. Geol Carpathica 62:477–485. doi: 10.2478/v10096-011-0034-7

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874. doi: 10.1016/j.patrec.2005.10.010

Felicísimo ÁM, Cuartero A, Remondo J, Quirós E (2012) Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10:175–189. doi: 10.1007/s10346-012-0320-1

Fire and Disaster Management Agency of Japan (2004) Disaster information on the mid Niigata Prefecture earthquake in 2004 (61st report on December 16).

Florinsky I (2012) Digital Terrain Analysis in Soil Science and Geology. doi: 10.1016/B978-0-12-385036-2.00029-8

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

Foody GM, Mathur A (2004) A relative evaluation of multiclass image classification by support vector machines. Geosci Remote Sensing, IEEE Trans 42:1335–1343.

Fourniadis IG, Liu JG, Mason PJ (2007) Landslide hazard assessment in the Three Gorges area, China, using ASTER imagery: Wushan–Badong. Geomorphology 84:126–144. doi: 10.1016/j.geomorph.2006.07.020

Fuchs M, Torizin J, Kühn F (2014) The effect of DEM resolution on the computation of the factor of safety using an infinite slope model. Geomorphology. doi: 10.1016/j.geomorph.2014.07.015

García-Rodríguez MJ, Malpica J a. (2010) Assessment of earthquake-triggered landslide susceptibility in El Salvador based on an Artificial Neural Network model. Nat Hazards Earth Syst Sci 10:1307–1315. doi: 10.5194/nhess-10-1307-2010

García-Rodríguez MJ, Malpica JA, Benito B, Díaz M (2008) Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. Geomorphology 95:172–191. doi: 10.1016/j.geomorph.2007.06.001

Geographical Survey Institute (2007) 2m grid elevation data. DVD-ROM.

Gessler PE, Moore ID, McKenzie NJ, Ryan PJ (1995) Soil-landscape modelling and spatial prediction of soil attributes. Int J Geogr Inf Syst 9:421–432.

Glenn NF, Streutker DR, Chadwick DJ, et al. (2006) Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity. Geomorphology 73:131–148. doi: 10.1016/j.geomorph.2005.07.006

Goldberg DE (1989) Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, p 343

Gorsevski P, Gessler P (2006) Spatial prediction of landslide hazard using logistic regression and ROC analysis. Trans. …

Guzzetti F (1999) Landslide Cartography , Hazard Assessment and Risk Evaluation : Overview , Limits and Prospective. 1–12.

Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31:181–216. doi: 10.1016/S0169-555x(99)00078-1

Guzzetti F, Malamud BD, Turcotte DL, Reichenbach P (2002) Power-law correlations of landslide areas in central Italy. Earth Planet Sci Lett 195:169–183. doi: 10.1016/S0012-821X(01)00589-1

Guzzetti F, Mondini AC, Cardinali M, et al. (2012) Landslide inventory maps: New tools for an old problem. Earth-Science Rev 112:42–66. doi: 10.1016/j.earscirev.2012.02.001

Guzzetti F, Reichenbach P, Ardizzone F, et al. (2006) Estimating the quality of landslide susceptibility models. Geomorphology 81:166–184. doi: 10.1016/j.geomorph.2006.04.007

Guzzetti F, Reichenbach P, Cardinali M, et al. (2005) Probabilistic landslide hazard assessment at the basin scale. Geomorphology 72:272–299. doi: 10.1016/j.geomorph.2005.06.002

Harp EL, Keefer DK, Sato HP, Yagi H (2011) Landslide inventories: The essential part of seismic landslide hazard analyses. Eng Geol 122:9–21. doi: 10.1016/j.enggeo.2010.06.013

Has B, Ishii Y, Maruyama K, et al. (2010) Relation between distance from earthquake source fault and scale of landslide triggered by recent two strong earthquakes in the Niigata Prefecture, Japan. Director 760:412–419.

Hoopes JC, McBride JH, Christiansen EH, et al. (2014) Characterizing a Landslide Hazard along the Wasatch Mountain Front (Utah). Environ Eng Geosci 20:1–24. doi: 10.2113/gseegeosci.20.1.1

Hovius N, Stark CP, Tutton MA, Abbott LD (1998) Landslide-driven drainage network evolution in a pre-steady-state mountain belt: Finisterre Mountains, Papua New Guinea. Geology 26:1071–1074.

Huang C, Davis L, Townshend J (2002a) An assessment of support vector machines for land cover classification. Int J Remoste Sesning 23:725–749.

Huang C, Davis L, Townshend J (2002b) An assessment of support vector machines for land cover classification. Int J Remoste Sesning 23:725–749.

Huang R, Pei X, Fan X, et al. (2012) The characteristics and failure mechanism of the largest landslide triggered by the Wenchuan earthquake, May 12, 2008, China. Landslides 9:131–142. doi: 10.1007/s10346-011-0276-6

IHRR (2010) How many people die from landslides? http://ihrrblog.org/2010/11/25/how-many-people-die-from-landslides/ (Date Accessed: November 15, 2014).

Iovine G, D'Ambrosio D, Di Gregorio S (2005) Applying genetic algorithms for calibrating a hexagonal cellular automata model for the simulation of debris flows characterised by strong inertial effects. Geomorphology 66:287–303. doi: 10.1016/j.geomorph.2004.09.017

Irigaray C, Fernández T, El Hamdouni R, Chacón J (2006) Evaluation and validation of landslide-susceptibility maps obtained by a GIS matrix method: examples from the Betic Cordillera (southern Spain). Nat Hazards 41:61–79. doi: 10.1007/s11069-006-9027-8

Jamsawang P, Voottipruex P, Boathong P, et al. (2015) Three-dimensional numerical investigation on lateral movement and factor of safety of slopes stabilized with deep cement mixing column rows. Eng Geol. doi: 10.1016/j.enggeo.2015.01.017

Japan Meteorological Agency (2004) Press release materials (28th report on Novemner 10).

Jebur MN, Pradhan B, Tehrany MS (2014) Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. Remote Sens Environ 152:150–165. doi: 10.1016/j.rse.2014.05.013

Jonassen DH, Hernandez-Serrano J (2002) Case-based reasoning and instructional design: Using stories to support problem solving. Educ Technol Res Dev 50:65–77. doi: 10.1007/BF02504994

Kavzoglu T, Sahin EK, Colkesen I (2013a) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. Landslides 11:425–439. doi: 10.1007/s10346-013-0391-7

Kavzoglu T, Sahin EK, Colkesen I (2013b) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. Landslides 11:425–439. doi: 10.1007/s10346-013-0391-7

Kawabata D, Bandibas J (2009) Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). Geomorphology 113:97–109. doi: 10.1016/j.geomorph.2009.06.006

Keefer DK (2000) Statistical analysis of an earthquake-induced landslide distribution — the 1989 Loma Prieta, California event. Eng Geol 58:231–249. doi: 10.1016/S0013-7952(00)00037-5

Keefer DK (1994) The importance of earthquake-induced landslides to long-term slope erosion and slope-failure hazards in seismically active regions. Geomorphology 10:265–284. doi: 10.1016/0169-555X(94)90021-3

Kieffer DS, Jibson R, Rathje EM, Kelson K (2006) Landslides triggered by the 2004 Niigata Ken Chuetsu, Japan, earthquake. Earthq Spectra 22:S47–S73. doi: 10.1193/1.2173021

Kojima H, Chung C-JF, van Westen CJ (2000) Strategy on the landslide type analysis based on the expert knowledge and the quantitative prediction model. Int. Arch. Photogramm. Remote Sens. pp 701–708

Komac M (2012) Regional landslide susceptibility model using the Monte Carlo approach–the case of Slovenia. Geol Q 56:41–54.

Korup O (2005a) Large landslides and their effect on sediment flux in South Westland, New Zealand. Earth Surf Process Landforms 30:305–323. doi: 10.1002/esp.1143

Korup O (2006) Effects of large deep-seated landslides on hillslope morphology, western Southern Alps, New Zealand. J Geophys Res 111:F01018. doi: 10.1029/2004JF000242

Korup O (2005b) Large landslides and their effect on sediment flux in South Westland, New Zealand. Earth Surf Process Landforms 30:305–323. doi: 10.1002/esp.1143

Kudo M, Sklansky J (2000) Comparison of algorithms that select features for pattern classifiers. Pattern Recognit 33:25–41.

Larsen IJ, Montgomery DR, Korup O (2010) Landslide erosion controlled by hillslope material. Nat Geosci 3:247–251. doi: 10.1038/ngeo776

Lee C, Tsai B (2008) Mapping VS30 in Taiwan. Terr Atmos Ocean Sci 19:671–682.

Lee CT, Huang CC, Lee JF, et al. (2008) Statistical approach to earthquake-induced landslide susceptibility. Eng Geol 100:43–58. doi: 10.1016/j.enggeo.2008.03.004

Lee S, Choi J, Min K (2004a) Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. Int J Remote Sens 25:2037–2052. doi: 10.1080/01431160310001618734

Lee S, Choi J, Woo I (2004b) The effect of spatial resolution on the accuracy of landslide susceptibility mapping: a case study in Boun, Korea. Geosci J 8:51–60. doi: 10.1007/BF02910278

Lee S, Pradhan B (2006) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides 4:33–41. doi: 10.1007/s10346-006-0047-y

Lee S, Sambath T (2006) Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. Environ Geol 50:847–855. doi: 10.1007/s00254-006-0256-7

Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. Environ Geol 47:982–990. doi: DOI 10.1007/s00254-005-1228-z

Li X, Yeh AG (2004) Multitemporal SAR images for monitoring cultivation systems using case-based reasoning. Photogramm Eng Remote Sens 90:524–534. doi: 10.1016/j.rse.2004.01.018

Li Y, Chen G, Han Z, et al. (2014) A hybrid automatic thresholding approach using panchromatic imagery for rapid mapping of landslides. GIScience Remote Sens 1–21. doi: 10.1080/15481603.2014.972867

Liu J., Mason P., Clerici N, et al. (2004) Landslide hazard assessment in the Three Gorges area of the Yangtze river using ASTER imagery: Zigui–Badong. Geomorphology 61:171–187. doi: 10.1016/j.geomorph.2003.12.004

Lu P, Rosenbaum MS (2003) Artificial Neural Networks and Grey Systems for the Prediction of

Slope Stability. Nat Hazards 30:383–398. doi: 10.1023/B:NHAZ.0000007168.00673.27

Lu P, Stumpf A, Kerle N, Casagli N (2011) Object-Oriented Change Detection for Landslide Rapid Mapping. IEEE Geosci Remote Sens Lett 8:701–705. doi: 10.1109/LGRS.2010.2101045

Lucas PJF (2001) Certainty-factor-like structures in Bayesian belief networks. Knowledge-Based Syst 14:327–335. doi: 10.1007/3-540-46238-4-3

Magliulo P, Di Lisio A, Russo F, Zelano A (2008) Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: a case study in southern Italy. Nat Hazards 47:411–435. doi: 10.1007/s11069-008-9230-x

Mantero P, Moser G, Serpico SB (2005) Partially supervised classification of remote sensing images through SVM-based probability density estimation. Geosci Remote Sensing, IEEE Trans 43:559–570.

Marjanovic M, Bajat B, Kovacevic M (2009) Landslide Susceptibility Assessment with Machine Learning Algorithms. 2009 Int Conf Intell Netw Collab Syst 273–278. doi: 10.1109/INCOS.2009.25

Martha TR, Kerle N, Jetten V, et al. (2010) Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. Geomorphology 116:24–36. doi: 10.1016/j.geomorph.2009.10.004

Martínez-Álvarez F, Reyes J, Morales-Esteban a., Rubio-Escudero C (2013) Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. Knowledge-Based Syst 50:198–210. doi: 10.1016/j.knosys.2013.06.011

Matlab (2010) Version 7.11(R2010b). MATLAB:2010

May C (2007) Sediment and Wood Routing in Steep Headwater Streams : An Overview of Geomorphic Processes and their Topographic Signatures. For Sci 53:119–130.

McKean J, Roering J (2004a) Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. Geomorphology 57:331–351. doi: 10.1016/S0169-555X(03)00164-8

McKean J, Roering J (2004b) Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. Geomorphology 57:331–351. doi: 10.1016/S0169-555X(03)00164-8

Meinhardt M, Fink M, Tünschel H (2015) Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: Comparison of a new method to calculate weighting factors by means of bivariate statistics. Geomorphology 234:80–97. doi: 10.1016/j.geomorph.2014.12.042

Minor M, Bergmann R, Görg S (2014) Case-based adaptation of workflows. Inf Syst 40:142–152. doi: 10.1016/j.is.2012.11.011

Miyagi T, Higaki D, Yagi H, et al. (2011) Reconnaissance report on landslide disasters in northeast Japan following the M 9 Tōhoku earthquake. Landslides 8:339–342. doi: 10.1007/s10346-011-0281-9

Mondini a. CC, Guzzetti F, Reichenbach P, et al. (2011) Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. Remote Sens Environ 115:1743–1757. doi: 10.1016/j.rse.2011.03.006

Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrol Process 5:3–30. doi:

10.1002/hyp.3360050103

Myint SW, Gober P, Brazel A, et al. (2011) Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens Environ 115:1145–1161. doi: 10.1016/j.rse.2010.12.017

Nath SK (2005) An initial model of seismic microzonation of Sikkim Himalaya through thematic mapping and GIS integration of geological and strong motion features. J Asian Earth Sci 25:329–343. doi: 10.1016/j.jseaes.2004.03.002

Nath SK (2004) Seismic hazard mapping and microzonation in the Sikkim himalaya through GIS integration of site effects and strong ground motion attributes. Nat Hazards 31:319–342.

Neuhäuser B, Damm B, Terhorst B (2011) GIS-based assessment of landslide susceptibility on the base of the Weights-of-Evidence model. Landslides 9:511–528. doi: 10.1007/s10346-011-0305-5

Ng KY (2006) Landslide locations and drainage network development: A case study of Hong Kong. Geomorphology 76:229–239. doi: 10.1016/j.geomorph.2005.10.008

Ni H, Zheng W, Song Z, Xu W (2014) Catastrophic debris flows triggered by a 4 July 2013 rainfall in Shimian, SW China: faormation mechanism, disaster characteristics and the lessons learned. Landslides 11:909–921. doi: 10.1007/s10346-014-0514-9

Nichol J, Wong MS (2005) Satellite remote sensing for detailed landslide inventories using change detection and image fusion. Int J Remote Sens 26:1913–1926. doi: 10.1080/01431160512331314047

Oguchi T (1996) Factors affecting the magnitude of post-glacial hillslope incision in Japanese mountains. Catena 26 :171–186.

Oguchi T (1997) Drainage Density and Relative Relief in Humid Steep Mountains with Frequent Slope Failure. Earth Surf Process Landforms 22:107–120.

Ohlmacher GC (2007) Plan curvature and landslide probability in regions dominated by earth flows and earth slides. Eng Geol 91:117–134.

Parsons AJ (1979) Short communication plan form and profile form of hillsopes. Earth Surf Process S 4:395–402.

Peckham S (2011) Profile, plan and streamline curvature: a simple derivation and applications. Geomorphometry 27–30.

Peng L, Niu R, Huang B, et al. (2014) Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. Geomorphology 204:287–301. doi: 10.1016/j.geomorph.2013.08.013

Pettorelli N, Vik JO, Mysterud A, et al. (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol Evol 20:503–10. doi: 10.1016/j.tree.2005.05.011

Pontius RG, Millones M (2011) Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int J Remote Sens 32:4407–4429. doi: 10.1080/01431161.2011.552923

Pourghasemi H, Pradhan B, Gokceoglu C, Moezzi KD (2013a) A comparative assessment of prediction capabilities of Dempster-Shafer and Weights-of-evidence models in landslide susceptibility mapping using GIS. Geomatics Nat Hazards Risk 4:93–118. doi: 10.1080/19475705.2012.662915

Pourghasemi HR, Moradi HR, Fatemi Aghda SM, et al. (2013b) GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). Arab J Geosci. doi: 10.1007/s12517-012-0825-x

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci 51:350–365. doi: 10.1016/j.cageo.2012.08.023

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. Environ Model Softw 25:747–759. doi: 10.1016/j.envsoft.2009.10.016

Pradhan B, Lee S, Buchroithner MF (2010a) A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. Comput Environ Urban Syst 34:216–235. doi: 10.1016/j.compenvurbsys.2009.12.004

Pradhan B, Youssef A, Varathrajoo R (2010b) Approaches for delineating landslide hazard areas using different training sites in an advanced artificial neural network model. Geo-spatial Inf Sci 13:93–102. doi: 10.1007/s11806-010-0236-7

Prasad R, Pandey a., Singh KP, et al. (2012) Retrieval of spinach crop parameters by microwave remote sensing with back propagation artificial neural networks: A comparison of different transfer functions. Adv Sp Res 50:363–370. doi: 10.1016/j.asr.2012.04.010

Pulko B, Majes B, Mikoš M (2012) Reinforced concrete shafts for the structural mitigation of large deep-seated landslides: an experience from the Macesnik and the Slano blato landslides (Slovenia). Landslides 11:81–91. doi: 10.1007/s10346-012-0372-2

Qian J, Liu R, Chen S, Dou J (2007) Man-made Ground Collapse Detection Using High

Resolution Aerial Image and Object-Based Classification : Example of Pearl River Delta. Proc. Second Int. ISCRAM Work. Inf. Syst. Cris. Response Manag. pp 1–6

Radoux J, Bogaert P (2014) Accounting for the area of polygon sampling units for the prediction of primary accuracy assessment indices. Remote Sens Environ 142:9–19. doi: 10.1016/j.rse.2013.10.030

Roering JJ, Kirchner JW, Dietrich WE (2005) Characterizing structural and lithologic controls on deep-seated landsliding: Implications for topographic relief and landscape evolution in the Oregon Coast Range, USA. Geol Soc Am Bull 117:654. doi: 10.1130/B25567.1

Roering JJ, Schmidt KM, Stock JD, et al. (2003) Shallow landsliding, root reinforcement, and the spatial distribution of trees in the Oregon Coast Range. Can Geotech J 40:237–253. doi: 10.1139/T02-113

Rumelhart D, Hinton G, Williams R (1986) Learning representations by back-propagating errors. Nature 323:533–536.

Santangelo M, Gioia D, Cardinali M, et al. (2013) Interplay between mass movement and fluvial network organization: An example from southern Apennines, Italy. Geomorphology 188:54–67. doi: 10.1016/j.geomorph.2012.12.008

Santini M, Grimaldi S, Nardi F, et al. (2009) Pre-processing algorithms and landslide modelling on remotely sensed DEMs. Geomorphology 113:110–125. doi: 10.1016/j.geomorph.2009.03.023

Sarkar S, Kanungo DP (2004) An Integrated Approach for Landslide Susceptibility Mapping Using Remote Sensing and GIS. Photogramm Eng Remote Sens 70:617–625. doi: 10.14358/PERS.70.5.617

Sassa K, Fukuoka H, Wang F, et al. (2006) Landslide Disasters Triggered by the 2004 Mid-

Niigata Prefecture Earthquake in Japan.

Schmidt KM, Roering JJ, Stock JD, et al. (2001) The variability of root cohesion as an influence on shallow landslide susceptibility in the Oregon Coast Range. Can Geotech J 38:995–1024. doi: 10.1139/cgj-38-5-995

Schulz WH (2007) Landslide susceptibility revealed by LIDAR imagery and historical records, Seattle, Washington. Eng Geol 89:67–87. doi: 10.1016/j.enggeo.2006.09.019

Schwab M, Rieke-Zapp D, Schneider H, et al. (2008) Landsliding and sediment flux in the Central Swiss Alps: A photogrammetric study of the Schimbrig landslide, Entlebuch. Geomorphology 97:392–406. doi: 10.1016/j.geomorph.2007.08.019

Seppelt R, Voinov A, Lange S, Bankamp D (2012) Application of support vector machines in landslide susceptibility assessment for the Hoa Binh province (Vietnam) with kernel functions analysis. iemss.org

Shahabi H, Ahmad BB, Khezri S (2013) Evaluation and comparison of bivariate and multivariate statistical methods for landslide susceptibility mapping (case study: Zab basin). Arab J Geosci 6:3885–3907. doi: 10.1007/s12517-012-0650-2

Shahabi H, Khezri S, Ahmad B Bin, Hashim M (2014) Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models. Catena 115:55–70. doi: 10.1016/j.catena.2013.11.014

Solaimani K, Mousavi SZ, Kavian A (2013) Landslide susceptibility mapping based on frequency ratio and logistic regression models. Arab J Geosci 6:2557–2569. doi: 10.1007/s12517-012-0526-5

Süzen ML, Doyuran V (2004) Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. Eng Geol 71:303–321. doi: 10.1016/S0013-7952(03)00143-1

Swets JA (1988) Measuring the Accuracy of Diagnostic Systems. Science 240:1285–1293.

Takeuchi K, Yanagisawa Y (2004) 1:50,000 Digital Geological Map of the Uonuma Tegion, Niigata Prefecture (Ver. 1), GSJ Open-file Report, 478, Geological Survey of Japan.

Tang AM, Quek C, Ng GS (2005) GA-TSKfnn: Parameters tuning of fuzzy neural network using genetic algorithms. Expert Syst Appl 29:769–781. doi: 10.1016/j.eswa.2005.06.001

Tarolli P (2014) High-resolution topography for understanding Earth surface processes: opportunities and challenges. Geomorphology. doi: 10.1016/j.geomorph.2014.03.008

Tien Bui D, Lofman O, Revhaug I, Dick O (2011) Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. Nat Hazards 59:1413–1444. doi: 10.1007/s11069-011-9844-2

Tien Bui D, Pradhan B, Lofman O, et al. (2012a) Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. Catena 96:28–40. doi: 10.1016/j.catena.2012.04.001

Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012b) Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naïve Bayes Models. Math Probl Eng 2012:1–26. doi: 10.1155/2012/974638

Tofani V, Del Ventisette C, Moretti S, Casagli N (2014) Integration of Remote Sensing Techniques for Intensity Zonation within a Landslide Area: A Case Study in the Northern Apennines, Italy. Remote Sens 6:907–924. doi: 10.3390/rs6020907

Tsai T-L, Chen H-F (2009) Effects of degree of saturation on shallow landslides triggered by rainfall. Environ Earth Sci 59:1285–1295. doi: 10.1007/s12665-009-0116-3

Turner AK, Schuster R (1996) Landslides: Investigation and Mitigation. Transp. Res. Board, Washington, D. C.

Van Den Eeckhaut M, Marre A, Poesen J (2010) Comparison of two landslide susceptibility assessments in the Champagne-Ardenne region (France). Geomorphology 115:141–155. doi: 10.1016/j.geomorph.2009.09.042

Van Den Eeckhaut M, Moeyersons J, Nyssen J, et al. (2009a) Spatial patterns of old, deep-seated landslides: A case-study in the northern Ethiopian highlands. Geomorphology 105:239–252. doi: 10.1016/j.geomorph.2008.09.027

Van Den Eeckhaut M, Reichenbach P, Guzzetti F, et al. (2009b) Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium. Nat Hazards Earth Syst Sci 9:507–521.

van Westen CJ, Rengers N, Terlien MTJ, Soeters R (1997) Prediction of the occurrence of slope instability phenomenal through GIS-based hazard zonation. Geol Rundschau 86:404–414. doi: 10.1007/s005310050149

Vancoillie F, Verbeke L, Dewulf R (2007) Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. Remote Sens Environ 110:476–487. doi: 10.1016/j.rse.2007.03.020

Vapnik VN (1998) Statistical Learning Theory (Adaptive and Learning Systems for Signal Processing, Communications and Control Series). Wiley-Interscience

Vita P, Napolitano E, Godt JW, Baum RL (2012) Deterministic estimation of hydrological thresholds for shallow landslide initiation and slope stability models: case study from the Somma-Vesuvius area of southern Italy. Landslides 10:713–728. doi:

10.1007/s10346-012-0348-2

Wang F, Wu Y-H, Yang H, et al. (2015) Preliminary investigation of the 20 August 2014 debris flows triggered by a severe rainstorm in Hiroshima City, Japan. Geoenvironmental Disasters 2:17. doi: 10.1186/s40677-015-0025-6

Wang H, Wang G, Wang F, et al. (2008) Probabilistic modeling of seismically triggered landslides using Monte Carlo simulations. Landslides 5:387–395. doi: 10.1007/s10346-008-0131-6

Wang L-J, Sawada K, Moriguchi S (2013) 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 7:073561. doi: 10.1117/1.JRS.7.073561

Wang Z, Zhao D (2006) Seismic images of the source area of the 2004 Mid-Niigata prefecture earthquake in Northeast Japan. Earth Planet. Sci. Lett.

Wieczorek GF, Jäger S (1996) Triggering mechanisms and depositional rates of postglacial slope-movement processes in the Yosemite Valley, California. Geomorphology 15:17–31.

Wu Q, Zhang H, Chen F, Dou J (2008) A web-based spatial decision support system for spatial planning and governance in the Guangdong Province. Geoinformatics 2008 Jt Conf GIS Built Environ Adv Spat Data Model Anal 7144:71442G–12. doi: 10.1117/12.812837

Xu C, Xu X, Dai F, et al. (2013) Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping related to the May 12, 2008 Wenchuan earthquake of China. Nat Hazards 68:883–900. doi: 10.1007/s11069-013-0661-7

Yalcin a., Reis S, Aydinoglu a. C, Yomralioglu T (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. CATENA 85:274–287. doi: 10.1016/j.catena.2011.01.014

Yamagishi H (2008) GIS mapping of landscape and disasters of Sado Island, Japan. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. Vol. XXXVII. Part B7. Beijing 2008. pp 1429–1432

Yamagishi H, Marui H, Ayalew L, et al. (2004) Estimation of the sequence and size of the Tozawagawa landslide, Niigata, Japan, using aerial photographs. Landslides 1:299–303. doi: 10.1007/s10346-004-0032-2

Yang M (2007) A genetic algorithm (GA) based automated classifier for remote sensing imagery. Can J Remote Sens 33:203–213.

Yang W, Wang M, Shi P (2013) Using MODIS NDVI Time Series to Identify Geographic Patterns of Landslides in Vegetated Regions. IEEE Geosci Remote Sens Lett 10:707–710. doi: 10.1109/LGRS.2012.2219576

Yesilnacar E, Topal T (2005) Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Eng Geol 79:251–266. doi: 10.1016/j.enggeo.2005.02.002

Yilmaz C, Topal T, Süzen ML (2011) GIS-based landslide susceptibility mapping using bivariate statistical analysis in Devrek (Zonguldak-Turkey). Environ Earth Sci 65:2161–2178. doi: 10.1007/s12665-011-1196-4

Yoong K, Okada A (2005) Surface deformations associated with the October 2004 Mid-Niigata earthquake: Description and discussion. Earth, Planets, Sp 57:1093–1102.

Zare M, Pourghasemi HR, Vafakhah M, Pradhan B (2013a) Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arab J Geosci 6:2873–2888. doi: 10.1007/s12517-012-0610-x

Zare M, Pourghasemi HR, Vafakhah M, Pradhan B (2013b) Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: A comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arab J Geosci 6:2873–2888. doi: 10.1007/s12517-012-0610-x

Zêzere J, Trigo R, Trigo I (2005) Shallow and deep landslides induced by rainfall in the Lisbon region (Portugal): assessment of relationships with the North Atlantic Oscillation. Nat Hazards Earth Syst Sci 5:331–344.

Zhu J, Qian J, Zhang Y, Dou J (2010) Research progress of nature reserve using remote sensing. J Anhui Agric Sci 38:10828–10831.

Zinko U, Seibert J, Dynesius M, Nilsson C (2005) Plant Species Numbers Predicted by a Topography-based Groundwater Flow Index. Ecosystems 8:430–441. doi: 10.1007/s10021-003-0125-0

APPENDICES

Matlab matching coding for detecting landslides in the CBR:

```
%extract the variable

[datalength,m_datawidth]=size(data);

[m_casedatalength,m_casedatawidth]=size(case_data)

m_fid = fopen('d: ¥test.txt', 'w');

 %matches the entire unknown patch to match the attribute.

for j=1:datalength

    m_data_id=data(j,1);

    m_data=data(j,2:m_datawidth);

     for i=1:m_casedatalength

        m_casedata_id=case_data(i,1);

        m_casedata=case_data(i,2:m_datawidth);

        m_minus=m_casedata-m_data;

        m_dist(i,1)=m_casedata_id;

        m_dist(i,2)=sqrt(sum(m_minus.^2));

    end

     [m_matchcase,m_matchcase_index] = sortrows(m_dist,2);

%      m_matchcasetype(j,1)={m_data_id};

_matchcasetype(j,2)={case_textdata(m_matchcase_index(1)+1,6)};

% %      fwrite(m_fid,cell2mat(m_matchcasetype(j,1)),'%d');

%      fwrite(m_fid,cell2mat(m_matchcasetype(j,2)),'%s');

     m_matchcasetype=case_textdata(m_matchcase_index(1)+1,6);
```

```
                fprintf(m_fid,'%d,',m_data_id);
                fprintf(m_fid,'%s ¥n',cell2mat(m_matchcasetype));
                  %fprintf(m_fid,'%d,%8.2d,%s,%s/n',m_data_id,m_casedata_id,cell2mat(m_mat
chcasetype),cell2mat(m_matchcasetype));
        end
          %save('d: ¥test.txt','m_matchcasetype','-ASCII')
        %save('d: ¥test.txt','m_matchcasetype')
        fclose(m_fid);
        fprintf(' ¥n');
        fprintf('finished');
        fprintf(' ¥n');
          [datalength,m_datawidth]=size(data);
        [m_casedatalength,m_casedatawidth]=size(case_data)
          m_data_id=data(1,1);
        m_data=data(1,2:m_datawidth);
          for i=1:m_casedatalength
                m_casedata_id=case_data(i,1);
                m_casedata=case_data(i,2:m_datawidth);
                m_minus=m_casedata-m_data;
                m_dist(i,1)=m_casedata_id;
                m_dist(i,2)=sqrt(sum(m_minus.^2));
        end
          [m_matchcase,m_matchcase_index] = sortrows(m_dist,2);
        m_matchcasetype=case_textdata(m_matchcase_index(1)+1,6);
```