Terr. Atmos. Ocean. Sci., Vol. 26, No. 2, Part II, 227-239, April 2015 doi: 10.3319/TAO.2014.12.02.07(EOSI)

Shallow and Deep-Seated Landslide Differentiation Using Support Vector Machines: A Case Study of the Chuetsu Area, Japan

Jie Dou[1, *], Uttam Paudel[1], Takashi Oguchi[2], Shoichiro Uchiyama[3], and Yuichi S. Hayakawa[2]

[1] *Department of Natural Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo , Chiba, Japan*
[2] *Center for Spatial Information Science, The University of Tokyo, Chiba, Japan*
[3] *National Research Institute for Earth Science and Disaster Prevention, Ibaraki, Japan*

Received 30 December 2013, revised 3 August 2014, accepted 2 December 2014

ABSTRACT

Landslides are one of the most destructive geological disasters affecting Japan every year, resulting in huge losses in life and property. Numerous susceptibility studies have been conducted to minimize the risk of landslides; however, most of these studies do not differentiate landslide types. This study 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. This study utilizes a 2-m DEM derived from airborne Light detection and ranging (Lidar), geological information and support vector machines (SVMs) to study the 1225 landslides triggered by the M 6.8 Chuetsu earthquake in Japan and the successive aftershocks. 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. 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. We also found that the overall accuracy of SVMs 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.

Key words: Airborne Lidar DEM, Shallow and deep-seated landslide, Differentiation, SVMs

Citation: Dou, J., U. Paudel, T. Oguchi, S. Uchiyama, and Y. S. Hayakawa, 2015: Shallow and deep-seated landslide differentiation 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)

1. INTRODUCTION

Landslides are induced by earthquakes, rainfall, snow melt and human intervention, 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 type of natural disasters (Guzzetti et al. 1999; García-Rodríguez et al. 2008). As indicated by Turner and Schuster (1996) this trend will continue and become clearer under the influence of urbanization, economic development, deforestation and increased regional precipitation in landslide-prone areas due to changing climate.

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 landslide (Lee et al. 2008; Cheng et al. 2010; Chen et al. 2013; Chang et al. 2014). Few studies have differentiated the probabilities for 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 hillslope erosion and sediment discharge for the protection of human settlements and infrastructures (Dramis and Sorriso-Valvo 1994; Korup 2005, 2006; Larsen et al. 2010; Lin et al. 2013). Some studies focused on the factors controlling the occurrence of deep-seated landslides. Roering et al. (2005) applied an algorithm developed from the

* *Corresponding author*
E-mail: doujie@csis.u-tokyo.ac.jp

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 (Guzzetti et al. 1999, 2006; Chang et al. 2014; Dou et al. 2014; Hoopes et al. 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 maximum likelihood classification (MLC), decision tree classification (DTC) and artificial neural network (ANN) classification with larger training data. SVMs have been widely applied in various fields including remote sensing, computer science, pattern recognition and economics (Marjanović et al. 2009; Bui et al. 2012a).

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 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; Kasai et al. 2009; Van Den Eeckhaut et al. 2009; Pulko et al. 2014).

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 were made on this event (Chigira and Yagi 2006; 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.

2. MATERIALS AND METHODS

2.1 Study Area

The study area is located in a mountainous region (Chuetsu) of Niigata Prefecture, Japan (Fig. 1). 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 - 734 m with a mean of 206 m. Approximately 2000 mm of precipitation, a few typhoons and heavy snow occur annually in this region. The area contains sedimentary and metamorphic rocks from the Paleogene to Quaternary periods (Takeuchi and Yanagisawa 2004). Land-use 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 magnitude 6.8

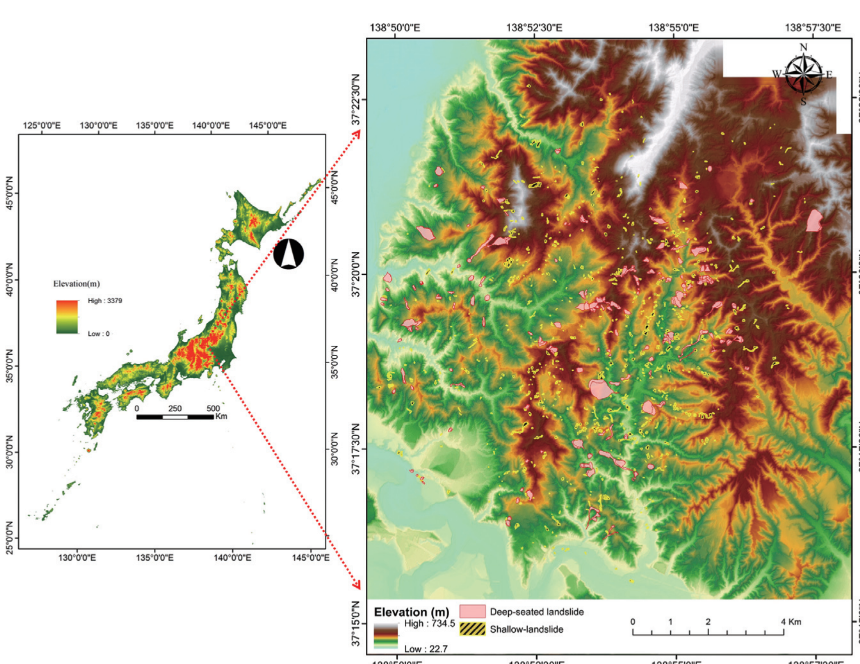

Fig. 1. Location of the study area and distribution of landslides (data from NIED).

on 23 October 2004, with the depth of the hypocenter being 13 km (Japan Meteorological Agency 2004; Fig. 2). The ground shaking caused by the earthquake and numerous aftershocks (four ≥ M 6 and ten M 5-6) triggered numerous shallow and deep-seated landslides (Fig. 3). The Fire and Disaster Management Agency of Japan (2004) reported that 40 people died and 4496 were injured during this event. The property and infrastructure damage from landslides alone was initially estimated at US $8 billion, making it one of the costliest landslide events in history (Kieffer et al. 2006). It is therefore necessary to assess this event in great detail.

2.2 Data

Previous landslides are key to predicting the distribution of future landslides (Guzzetti et al. 1999). We used a landslide inventory prepared by the National Research Institute for Earth Science and Disaster Prevention (NIED), Japan, based on the aerial photograph interpretation (Fig. 1). The inventory contains 895 shallow landslides and 330 deep-seated landslides with an average area of 187 and 9600 m^2 respectively (Fig. 4). Their minimum areas are 42 and 271 m^2, and maximum areas are 28178 and 205461 m^2, respectively. The inventory contains landslides represented by polygons; however, for this study, the landslide polygons were changed into points at the centroid of each landslide polygon, using ArcGIS 10.1, a GIS package from ESRI.

Landslides are classified as deep and shallow in relation to the material and movement mechanism (Dai et al. 2011) (Fig. 5). NIED used the sliding plane depth to differentiate the two types: depth < 10 m = shallow and >10 m = deep-seated. This landslide differentiation scheme is also used by Roering et al. (2003). Shallow landslides with surface soil mantle movement 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 to cause large scale debris flows and landslide dams, with more disastrous consequences.

The thematic data used are summarized in Table 1. An airborne Lidar DEM with a spatial resolution of 2 m was provided by the Geospatial Information Authority of Japan in 2005 (Geographical Survey Institute 2007). The geological information used is based on the geological maps provided by the Geological Survey of Japan (Takeuchi and Yanagisawa 2004). All conditioning factors used are continuous except the categorical lithology data.

Landslide causative factor selection 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. We therefore considered landslide susceptibility studies by Guzzetti et al. (1999, 2006), Lee and Sambath (2006), Caniani et al. (2008), Lee and Tsai, (2008), Lee et al. (2008), and Dou et al. (2009) for selecting the factors summarized in Table 2. The factors include several DEM derivatives: elevation, slope angle, aspect, curvature, distance from drainage network, the compound topographic index (CTI), and the stream power index (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 Fig. 6.

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. 2014). Curvature controls hydraulic flow in relation to convergence and divergence and hence landslide occurrence (Dai et al. 2011). Shallow and deep-seated landslide differentiation may depend on the lithology (Wieczorek and Jäger 1996) that affects the thickness of weak beds. Groundwater and soil moisture conditions in relation to topography affect landslides (Zinko et al. 2005). Several topographic indices have been proposed to describe these conditions. CTI and SPI, developed by Beven and Kirkby (1979) and Gessler et al. (1995), respectively, are used in this study:

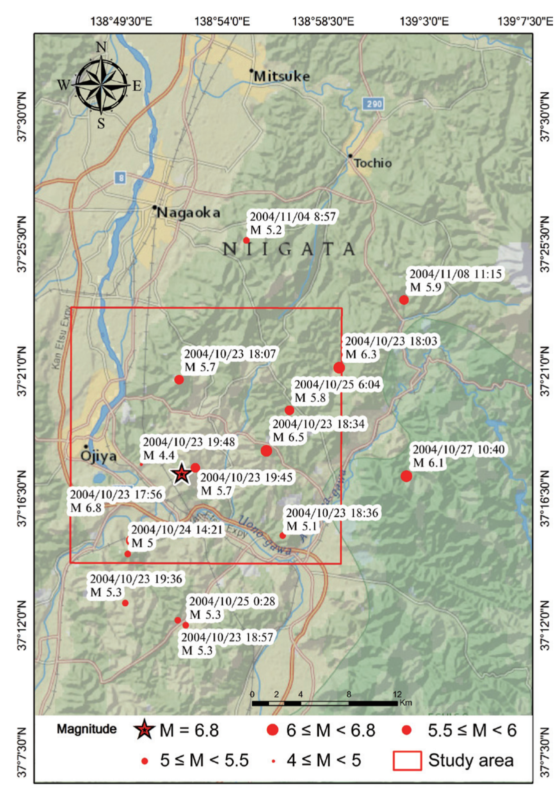

Fig. 2. Epicentral distribution of major earthquakes in and around mid-Niigata prefecture region. Map: Courtesy of ESRI.

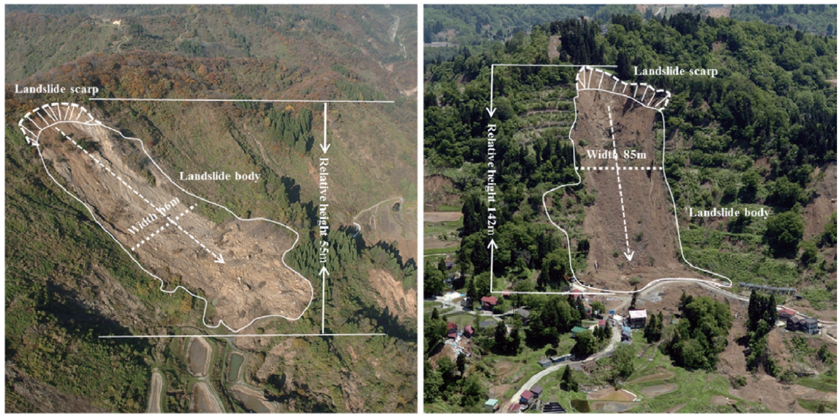

Fig. 3. Figure of shallow (left) in the southwest of Nigorizawa Nagaoka city; and deep-seated (right) landslides in the western entrance of Haguro tunnel, Nagaoka city. Photographs: Courtesy of NIED, Japan.

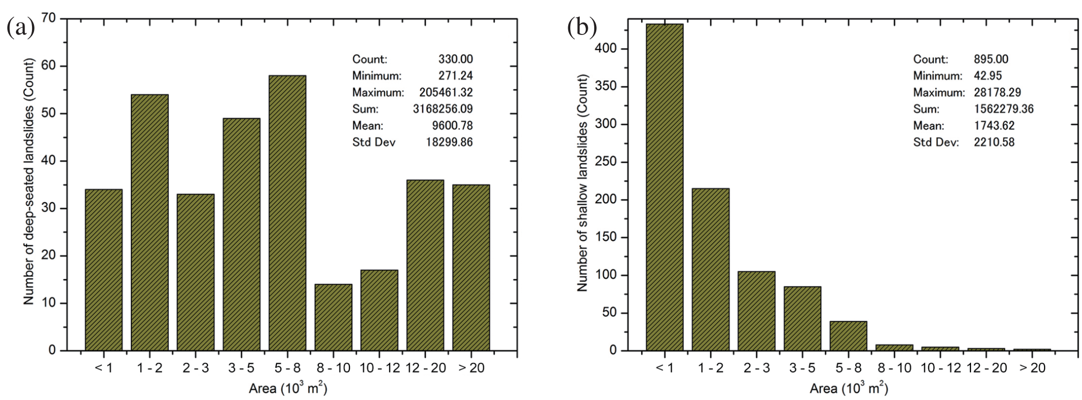

Fig. 4. Histograms showing characteristics of landslide types: (a) area of deep-seated landslide, (b) area of shallow landslide.

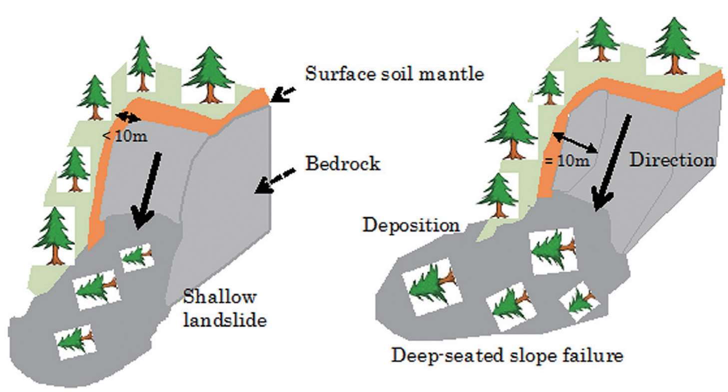

Fig. 5. Sketch of shallow landslides (left) and deep-seated landslides (right).

Table 1. 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	1:50000	Continuous	NIED
Geological map	Lithology	Polygon coverage	1:50000	Categorical	GSJ
	Geological boundary	Line coverage		Continuous	
Topographic map	DEM	ARC/INFO Grid	2 × 2 m	Continuous	GSI

Table 2. 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 elevation	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/tanβ) 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.

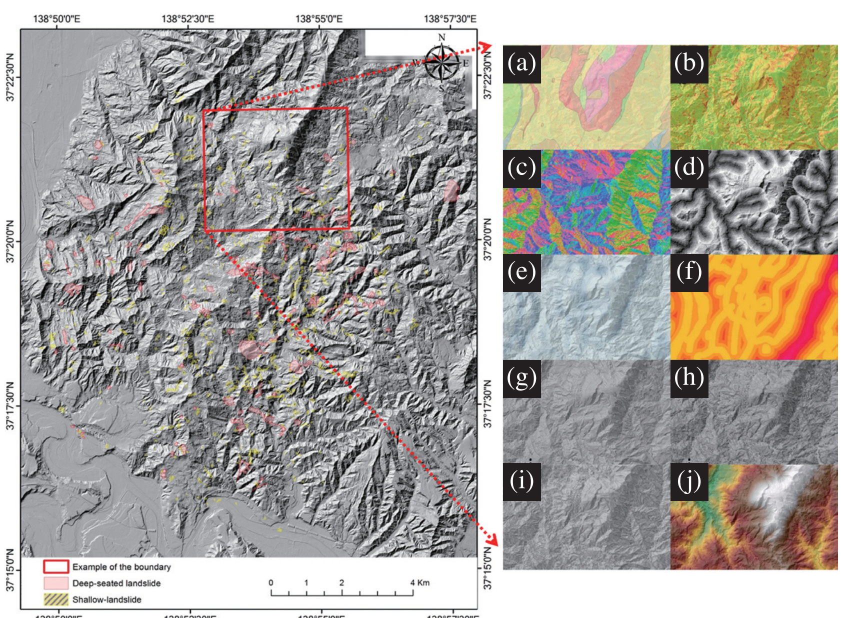

Fig. 6. Factors maps: (a) lithology, (b) slope angle, (c) aspect, (d) distance from drainage network, (e) density of geologic boundaries, (f) distance from the nearest geologic boundary, (g) curvature, (h) compound topographic index (CTI), (i) stream power index (SPI), and (j) elevation.

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

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

where As is the specific catchment area per unit channel width orthogonal to the flow direction ($m^2\ m^{-1}$) and β is the slope angle expressed in degrees. CTI is strongly correlated with soil moisture, while SPI taking into account both slope and flow accumulation is correlated with erosion potential. These indices also provide information on soil depth and soil constituents (Moore et al. 1991; Florinsky 2011) suggesting that they can be invaluable factors in predicting the landslide types (Wieczorek and Jäger 1996). Geological boundary density may reflect the tectonic activity that results in slope instability (Dou et al. 2014).

2.3 SVMs Model

SVMs provide supervised learning models with associated algorithms based on the concept of optimal separating hyper plane 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 7 illustrates this concept: ovals and squares are two kinds of samples and the separating hyper plane (H) is one of possible planes which can separate the two classes; and the distance between the two dotted lines in Fig. 7 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).

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 hyper plane by differentiating the two classes by solving the following optimization function (Vapnik 1998):

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

Subjected to the constraints of the following equation:

$$y_i [w^T \phi(x_i) + b] \geq 1 - \xi_i, \xi_i \geq 0 \tag{4}$$

where w is a coefficient vector, b is the offset of the hyper plane 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:

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

Normally four basic kernel functions, linear (LF), polynomial (PF), radial basis (RBF), and sigmoid (SF) functions, are used in the SVMs. Table 3 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.

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:

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

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 4.

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. 2010; Bui et al. 2012b). Each time a random set of landslides was selected for

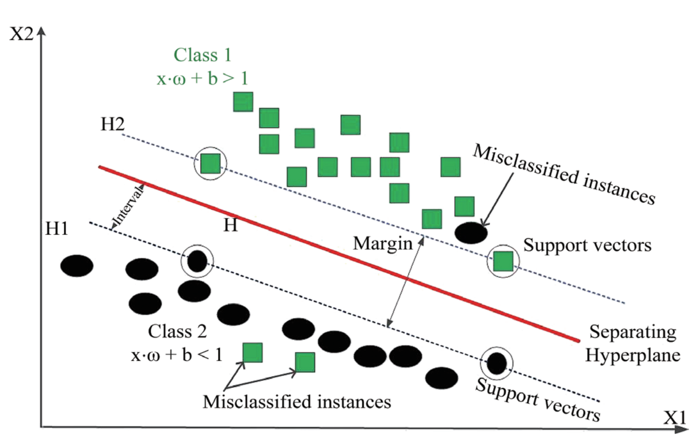

Fig. 7. Illustration of the optimal separating hyperplane.

Table 3. 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)	$k(x_i, x_j) = (\gamma x_i^T x_j + \Upsilon)^a, \Upsilon > 0$	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%

Table 4. 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	0.1862	0.1675	0.7516	0.2313	0.2222	0.2714	0	0	1
non-landslide	0.1970	0.2952	0.4871	0.2545	0.7823	0.0428	0	0	1
non-landslide	0.2824	0.0905	0.2792	0.2536	0.2838	0.3347	0	0	1
deep	0.0228	0.3355	0.0473	0.2516	0.4106	0.0518	1	0	0
deep	0.0562	0.5063	0.8478	0.2468	0.4706	0.1312	1	0	0
deep	0.0700	0.4798	0.8495	0.2536	0.4619	0.0631	1	0	0
shallow	0.4310	0.0095	0.7516	0.2554	0.1152	0.0147	0	1	0
shallow	0.4080	0.0667	0.9786	0.2554	0.1241	0.0147	0	1	0
shallow	0.4497	0.4067	0.1854	0.2528	0.1521	0.0147	0	1	0
deep	0.2114	0.5017	0.9613	0.2572	0.1826	0.4758	1	0	0
deep	0.2163	0.1865	0.8253	0.2506	0.2033	0.3733	1	0	0
shallow	0.4794	0.5104	0.2249	0.2462	0.2325	0.0147	0	1	0
shallow	0.4176	0.1233	0.7273	0.2542	0.1172	0.0147	0	1	0
shallow	0.4310	0.0095	0.7516	0.2554	0.1152	0.0147	0	1	0
shallow	0.4080	0.0667	0.9786	0.2554	0.1241	0.0147	0	1	0
shallow	0.4497	0.4067	0.1854	0.2528	0.1521	0.0147	0	1	0
shallow	0.4286	0.3098	0.2846	0.2510	0.2112	0.0147	0	1	0
shallow	0.4451	0.2773	0.7070	0.2688	0.3700	0.0147	0	1	0
shallow	0.4306	0.0538	0.3574	0.2596	0.3545	0.0147	0	1	0
non-landslide	0.3299	0.1215	0.8615	0.2076	0.2331	0.5122	0	0	1
non-landslide	0.3373	0.0903	0.2788	0.1975	0.3628	0.3347	0	0	1
deep	0.0189	0.3004	0.9941	0.2423	0.4322	0.0374	1	0	0
deep	0.1811	0.3794	0.8790	0.2564	0.2769	0.3214	1	0	0
deep	0.1590	0.3130	0.8507	0.2606	0.2774	0.2903	1	0	0
deep	0.0635	0.3947	0.9786	0.2657	0.4352	0.1495	1	0	0
deep	0.0549	0.4451	0.8359	0.2555	0.4472	0.1057	1	0	0

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.

3. RESULTS AND DISCUSSION

SVM model performance is directly related to kernel function and parameter selection. Each kernel was trained and tested; the prediction results for the two landslide types from the 10 factors for each kernel (Table 3) show that the PF sample training accuracy is the highest (95.26%) followed by that of RBF (94.38%). RBF outperformed for the testing samples (84.31%) and was hence selected as the SVMs kernel for this study. This classifier selection may be an SVM limitation because only one constraint is active at a time (Burges 1998; Kavzoglu et al. 2014).

The average back propagation (BP) technique accuracy using ANN applied to the same dataset from an unpublished work from the same authors, showing 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). 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 Fig. 8. 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 2005; May 2007). Although we used the 2-m DEM, using higher resolution topographical information seems to be necessary to study shallow landslides in detail.

For the reduced size of the training dataset (30, 40, and 50%), the model performed equally well (Table 6 and Fig. 9). 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) , Huang et al. (2002), Chi et al. (2008), and Kavzoglu et al. (2014) in terms of the stability of SVMs even with fewer training datasets, compared to other models like ANN.

Figure 10 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, based on 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. 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).

4. CONCLUSION

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

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

Iteration number	Accuracy (%) of SVMs		Accuracy (%) of BP	
	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

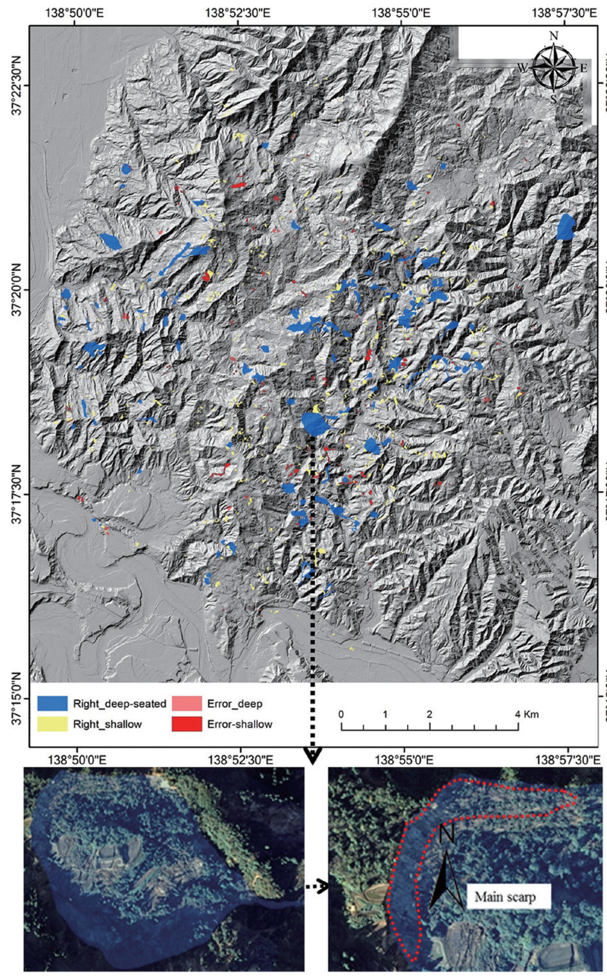

Fig. 8. Map prepared using the confusion matrix obtained from classification of landslide types using SVM with 50% trainings sample (upper); representational aerial photographs (lower). Photographs: Courtesy of GSI, Japan.

Table 6. Accuracy of the SVMs model using 30, 40, and 50% of the data to train the model.

	Prediction accuracy (%) with the testing samples		
Iteration number	**Model trained with 50% data**	**Model trained with 40% data**	**Model trained with 30% data**
1	75.87	78.6	74.7
2	79.38	75.1	80.8
3	73.74	73.5	73.74
4	74.22	74.1	77.4
5	76.87	72.3	79.5
6	77.91	80.0	72.3
7	77.39	76.8	72.6
8	78.09	75.1	72.9
9	80.04	74.8	72.3
10	84.31	72.1	74.8
Min	73.74	72.1	72.3
Max	84.31	80.0	80.8
Standard deviation	2.89	2.43	2.93
Average	77.78	75.24	75.1

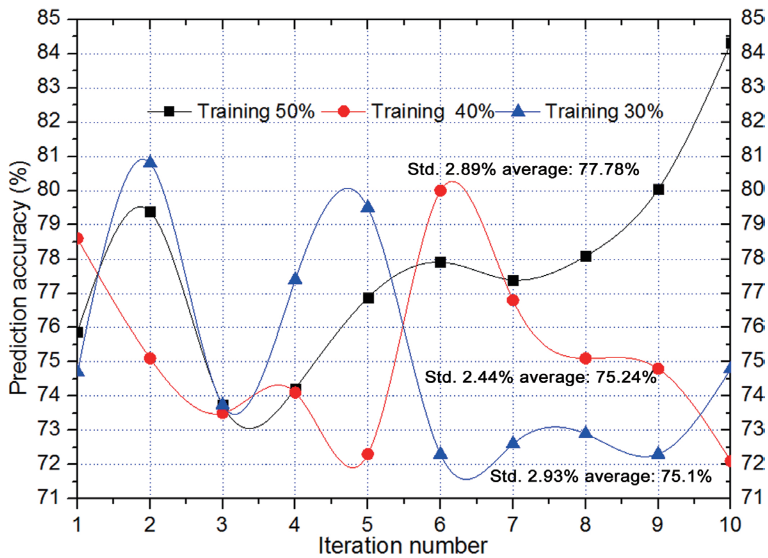

Fig. 9. Prediction accuracy (testing samples) of the SVMs model using 30, 40, and 50% of the data to train the model.

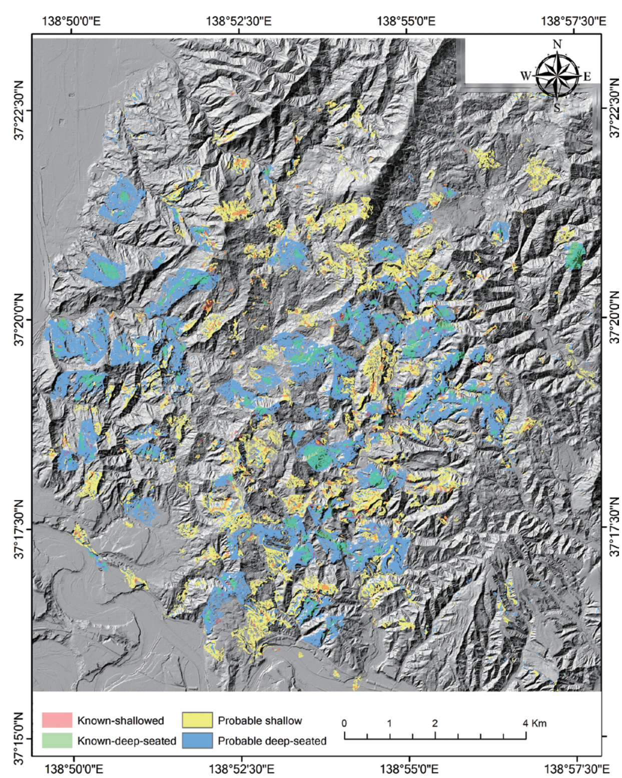

Fig. 10. Map showing the probable occurrences of shallow and deep-seated landslides in the whole study area.

(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 - 30% of the total dataset did not significantly affect the SVMs model accuracy 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.

Acknowledgements The authors would like to express their sincere gratitude to the NIED for allowing us to use the landslide data. We deeply thank Prof. Chyi-Tyi Lee and an anonymous reviewer who provided their constructive comments on a previous draft of the manuscript.

REFERENCES

Ardizzone, F., M. Cardinali, M. Galli, F. Guzzetti, and P. Reichenbach, 2007: Identification and mapping of recent rainfall-induced landslides using elevation data collected by airborne Lidar. *Nat. Hazards Earth Syst. Sci.*, **7**, 637-650, doi: 10.5194/nhess-7-637-2007. [Link]

Beven, K. J. and M. J. Kirkby, 1979: A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.*, **24**, 43-69, doi: 10.1080/02626667909491834. [Link]

Bui, D. T., B. Pradhan, O. Lofman, and I. Revhaug, 2012a: 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. [Link]

Bui, D. T., B. Pradhan, O. Lofmana, I. Revhauga, and O. B. Dick, 2012b: Application of support vector machines in landslide susceptibility assessment for the Hoa Binh province (Vietnam) with kernel functions analysis. In: Seppelt, R., A. A. Voinov, S. Lange, and D. Bankamp (Eds.), International Environmental Modelling and Software Society (iEMSs), 2012 International Congress on Environmental Modelling and Software, Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting, Leipzig, Germany, 382-389.

Burges, C. J. C., 1998: A tutorial on Support Vector Machines for pattern recognition. *Data Min. Knowl. Disc.*, **2**, 121-167, doi: 10.1023/A:1009715923555. [Link]

Caniani, D., S. Pascale, F. Sdao, and A. Sole, 2008: 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. [Link]

Chang, C. C. and C. J. Lin, 2011: LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Tech.*, **2**, 1-27, doi: 10.1145/1961189.1961199. [Link]

Chang, K., S. Chiang, Y. Chen, and A. C. Mondini, 2014: Modeling the spatial occurrence of shallow landslides triggered by typhoons. *Geomorphology*, **208**, 137-148, doi: 10.1016/j.geomorph.2013.11.020. [Link]

Chang, T. C. and R. J. Chao, 2006: Application of back-propagation networks in debris flow prediction. *Eng. Geol.*, **85**, 270-280, doi: 10.1016/j.enggeo.2006.02.007. [Link]

Chen, W., X. Li, Y. Wang, and S. Liu, 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. [Link]

Cheng, C. T., C. T. Lee, P. S. Lin, B. S. Lin, Y. B. Tsai, and S. J. Chiou, 2010: Probabilistic earthquake hazard in metropolitan Taipei and its surrounding regions. *Terr. Atmos. Ocean. Sci.*, **21**, 429-446, doi: 10.3319/TAO.2009.11.11.01(TH). [Link]

Chi, M., R. Feng, and L. Bruzzone, 2008: Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. *Adv. Space Res.*, **41**, 1793-1799, doi: 10.1016/j.asr.2008.02.012. [Link]

Chigira, M. and H. Yagi, 2006: Geological and geomorphological characteristics of landslides triggered by the 2004 Mid Niigta prefecture earthquake in Japan. *Eng. Geol.*, **82**, 202-221, doi: 10.1016/j.enggeo.2005.10.006. [Link]

Crowther, P. S. and R. J. Cox, 2006: Accuracy of neural network classifiers as a property of the size of the data set. In: Gabrys, B., R. J. Howlett, and L. C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, Vol. 4253, Springer Berlin Heidelberg, 1143-1149, doi: 10.1007/11893011_144. [Link]

Dai, F. C., C. Xu, X. Yao, L. Xu, X. B. Tu, and Q. M. Gong, 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. [Link]

Dou, J., J. Qian, H. Zhang, S. Chen, X. Zheng, J. Zhu, Z. Xie, and Y. Zou, 2009: Landslides detection: A case study in Conghua city of Pearl River delta. In: Zhang, X., J. Li, G. Liu, and X. Yang (Eds.), Proc. SPIE 7471, Second International Conference on Earth Observation for Global Changes, 74711K, doi: 10.1117/12.836328. [Link]

Dou, J., T. Oguchi, Y. S. Hayakawa, S. Uchiyama, H. Saito, and U. Paudel, 2014: GIS-based landslide susceptibility mapping using a certainty factor model and its validation in the Chuetsu area, Central Japan. In: Sassa, K., P. Canuti, and Y. Yin (Eds.), Landslide Science for a Safer Geoenvironment, Springer International Publishing, Cham, 419-424, doi: 10.1007/978-3-319-05050-8_65. [Link]

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

Fire and Disaster Management Agency of Japan, 2004: Disaster information on the mid Niigata Prefecture earthquake in 2004. Available at http://j.mp/1waDD7B. (in Japanese)

Florinsky, I. V., 2011: Digital Terrain Analysis in Soil Science and Geology, Academic Press, 432 pp.

Foody, G. M. and A. Mathur, 2004: A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans. Geosci. Remote Sensing*, **42**, 1335-1343, doi: 10.1109/TGRS.2004.827257. [Link]

Foody, G. M. and A. Mathur, 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, doi: 10.1016/j.rse.2006.04.001. [Link]

García-Rodríguez, M. J., J. A. Malpica, B. Benito, and M. Díaz, 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. [Link]

Geographical Survey Institute, 2007: 2m grid elevation data, DVD-ROM.

Gessler, P. E., I. D. Moore, N. J. McKenzie, and P. J. Ryan, 1995: Soil-landscape modelling and spatial prediction of soil attributes. *Int. J. Geogr. Inform. Syst.*, **9**, 421-432, doi: 10.1080/02693799508902047. [Link]

Glenn, N. F., D. R. Streutker, D. J. Chadwick, G. D. Thackray, and S. J. Dorsch, 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. [Link]

Guzzetti, F., A. Carrara, M. Cardinali, and P. Reichenbach, 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. [Link]

Guzzetti, F., P. Reichenbach, F. Ardizzone, M. Cardinali, and M. Galli, 2006: Estimating the quality of landslide susceptibility models. *Geomorphology*, **81**, 166-184, doi: 10.1016/j.geomorph.2006.04.007. [Link]

Harp, E. L., D. K. Keefer, H. P. Sato, and H. Yagi, 2011: Landslide inventories: The essential part of seismic landslide hazard analyses. *Eng. Geol.*, **122**, 9-21, doi: 10.1016/j.enggeo.2010.06.013. [Link]

Hoopes, J. C., J. H. McBride, E. H. Christiansen, B. J. Kowallis, T. J. Thompson, D. G. Tingey, A. O. Okojie-Ayoro, and J. R. Yaede, 2014: Characterizing a landslide hazard along the Wasatch Mountain Front (Utah). *Environ. Eng. Geosci.*, **20**, 1-24, doi: 10.2113/gseegeosci.20.1.1. [Link]

Huang, C., L. S. Davis, and J. R. G. Townshend, 2002: An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.*, **23**, 725-749, doi: 10.1080/01431160110040323. [Link]

Japan Meteorological Agency, 2004: Niigata Chuetsu Earthquake. Available at http://j.mp/1AL3N5X. (in Japanese)

Kasai, M., M. Ikeda, T. Asahina, and K. Fujisawa, 2009: LiDAR-derived DEM evaluation of deep-seated landslides in a steep and rocky region of Japan. *Geomorphology*, **113**, 57-69, doi: 10.1016/j.geomorph.2009.06.004. [Link]

Kavzoglu, T., E. K. Sahin, and I. Colkesen, 2014: 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. [Link]

Kawabata, D. and J. Bandibas, 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. [Link]

Keefer, D. K., 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. [Link]

Kieffer, D. S., R. Jibson, E. M. Rathje, and K. Kelson, 2006: Landslides triggered by the 2004 Niigata Ken Chuetsu, Japan, earthquake. *Earthq. Spectra*, **22**, 47-73, doi: 10.1193/1.2173021. [Link]

Korup, O., 2005: Large landslides and their effect on sediment flux in South Westland, New Zealand. *Earth Surf. Process. Landf.*, **30**, 305-323, doi: 10.1002/esp.1143. [Link]

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. [Link]

Larsen, I. J., D. R. Montgomery, and O. Korup, 2010: Landslide erosion controlled by hillslope material. *Nat. Geosci.*, **3**, 247-251, doi: 10.1038/ngeo776. [Link]

Lee, C. T. and B. R. Tsai, 2008: Mapping Vs30 in Taiwan. *Terr. Atmos. Ocean. Sci.*, **19**, 671-682, doi: 10.3319/TAO.2008.19.6.671(PT). [Link]

Lee, C. T., C. C. Huang, J. F. Lee, K. L. Pan, M. L. Lin, and J. J. Dong, 2008: Statistical approach to earthquake-induced landslide susceptibility. *Eng. Geol.*, **100**, 43-58, doi: 10.1016/j.enggeo.2008.03.004. [Link]

Lee, S. and T. Sambath, 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. [Link]

Lin, C. W., C. M. Tseng, Y. H. Tseng, L. Y. Fei, Y. C. Hsieh, and P. Tarolli, 2013: Recognition of large scale deep-seated landslides in forest areas of Taiwan using high resolution topography. *J. Asian Earth Sci.*, **62**, 389-400, doi: 10.1016/j.jseaes.2012.10.022. [Link]

Mantero, P., G. Moser, and S. B. Serpico, 2005: Partially

Supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Trans. Geosci. Remote Sensing*, **43**, 559-570, doi: 10.1109/TGRS.2004.842022. [Link]

Marjanović, M., B. Bajat, and M. Kovačević, 2009: Landslide susceptibility assessment with machine learning algorithms. INCOS '09, International Conference on Intelligent Networking and Collaborative Systems, IEEE, 273-278, doi: 10.1109/INCOS.2009.25. [Link]

May, C., 2007: Sediment and wood routing in steep headwater streams: An overview of geomorphic processes and their topographic signatures. *Forest Sci.*, **53**, 119-130.

McKean, J. and J. Roering, 2004: 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. [Link]

Moore, I. D., R. B. Grayson, and A. R. Ladson, 1991: Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.*, **5**, 3-30, doi: 10.1002/hyp.3360050103. [Link]

Oguchi, T., 1996: Factors affecting the magnitude of post-glacial hillslope incision in Japanese mountains. *CATENA*, **26**, 171-186, doi: 10.1016/0341-8162(95)00039-9. [Link]

Pradhan, B., A. Youssef, and R. Varathrajoo, 2010: Approaches for delineating landslide hazard areas using different training sites in an advanced artificial neural network model. *Geo-spatial Inform. Sci.*, **13**, 93-102, doi: 10.1007/s11806-010-0236-7. [Link]

Pulko, B., B. Majes, and M. Mikoš, 2014: 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. [Link]

Roering, J. J., K. M. Schmidt, J. D. Stock, W. E. Dietrich, and D. R. Montgomery, 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. [Link]

Roering, J. J., J. W. Kirchner, and W. E. Dietrich, 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**, 654-668, doi: 10.1130/B25567.1. [Link]

Schmidt, K. M., J. J. Roering, J. D. Stock, W. E. Dietrich, D. R. Montgomery, and T. Schaub, 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/t01-031. [Link]

Schulz, W. H., 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. [Link]

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

Turner, A. K. and R. L. Schuster, 1996: Landslides: Investigation and Mitigation, National Academy Press, Washington, DC, 673 pp.

Van Den Eeckhaut, M., J. Moeyersons, J. Nyssen, A. Abraha, J. Poesen, M. Haile, and J. Deckers, 2009: 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. [Link]

Vapnik, V. N., 1998: Statistical Learning Theory, Wiley-Interscience, 768 pp.

Wieczorek, G. F. and S. Jäger, 1996: Triggering mechanisms and depositional rates of postglacial slope-movement processes in the Yosemite Valley, California. *Geomorphology*, **15**, 17-31, doi: 10.1016/0169-555X(95)00112-I. [Link]

Zêzere, J. L., R. M. Trigo, and I. F. Trigo, 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, doi: 10.5194/nhess-5-331-2005. [Link]

Zinko, U., J. Seibert, M. Dynesius, and C. Nilsson, 2005: Plant species numbers predicted by a topography-based groundwater flow index. *Ecosystems*, **8**, 430-441, doi: 10.1007/PL00021513. [Link]