

Article

The Analysis on Similarity of Spectrum Analysis of Landslide and Bareland through Hyper-Spectrum Image Bands

Shiuan Wan ¹, Tsu Chiang Lei ^{2,*}, Hong Lin Ma ¹ and Ru Wen Cheng ¹

¹ Department of Information Technology, Ling Tung University, Taichung 40851, Taiwan; shiuan123@teamail.ltu.edu.tw (S.W.); df916118@gmail.com (H.L.M.); 1989r67@gmail.com (R.W.C.)

² Department of Urban Planning and Spatial Information, Feng Chia University, Taichung 40724, Taiwan

* Correspondence: tclei@fcu.edu.tw; Tel.: +886-0424517250

Received: 24 September 2019; Accepted: 12 November 2019; Published: 17 November 2019



Abstract: Landslides of Taiwan occur frequently in high mountain areas. Soil disturbance causes by the earthquake and heavy rainfall of the typhoon seasons often produced the earth and rock to landslide in the upper reaches of the catchment area. Therefore, the landslide near the hillside has an influence on the catchment area. The hyperspectral images are effectively used to monitor the landslide area with the spectral analysis. However, it is rarely studied how to interpret it in the image of the landslide. If there are no elevation data on the slope disaster, it is quite difficult to identify the landslide zone and the bareland area. More specifically, this study used a series of spectrum analysis to identify the difference between them. Therefore, this study conducted a spectrum analysis for the classification of the landslide, bareland, and vegetation area in the mountain area of NanXi District, Tainan City. On the other hand, this study used the following parallel study on Support Vector Machine (SVM) for error matrix and thematic map for comparison. The study simultaneously compared the differences between them. The spectral similarity analysis reaches 85% for testing data, and the SVM approach has 98.3%.

Keywords: landslide; image classification; spectrum similarity analysis

1. Introduction

Landslides cause a great loss of human lives and properties. Landslides are frequent phenomena in Taiwan in which a more effective solution to estimate landslide area is desired through considering the remote sensing data [1–3]. Conventionally, monitoring of landslides for their locations and distributions are generally used in situ or field geotechnical techniques through aerial photos by human-power or unmanned aerial devices [4–6]. In the past, the investigation of landslide areas requires much manpower, material resources, and funding, and is very time-consuming. Various modeling approaches have been taken in the form of multivariate statistical analyses or Data Mining techniques of landslide characteristics corresponding to past landslide records. Many researchers studied the landslide through various evaluation/estimation through a Geographic Information System [7,8] with different techniques. The usage of aerial images in large-scale land cover surveys is of great help to the problem [9–11]. Nowadays, spatial information technology is the most proper solution for spatial analysis, which is to effectively and accurately judge the landslide through remotely-sensed images [12]. Hyperspectral image data have been developed for more than 20 years. Hyperspectral image data combine the spectrum shape and image data. In general, the wavelengths of spectrum are divided into visible, near-infrared, and part-short-wave infrared—three different parts. Those instruments recorded the spectral reflection information of the material to obtain complete geospatial information quickly and

extensively [13]. Due to the high spectral resolution of hyperspectral images, it can provide rich material details in landslide analysis [14,15].

Landslides cause lots of human life and economic losses every year. With the progressing techniques of spatial data survey in geosciences, large amounts of data for observing the change in the landslide area can easily be collected. Accordingly, the advancement and development of science and technology have enabled remotely to obtain large-scale and high-resolution quantitative information in a short period of time. To find the most valuable knowledge from the target, statistical classification and data mining techniques are usually used to predict the results of the analysis [2,4,8]. The aim of this research is to produce landslide susceptibility mapping by remote sensing data processing and GIS spatial analysis. To identify the unknown species, the spectral reflection diagram of the ground object could be used [16,17]. This action is like to discover the identification code of the ground object which can help us identify different features of land cover. Hyperspectral Imaging has a large number of bands and is almost continuous, which displays a relatively narrow on the spectral range of each band is relatively narrow. The amount of data obtained is huge and it can completely show slight differences in the spectrum of different features.

Due to the lack of accurate DEM (Digital Elevation Modeling) map/data in this study, only the hyperspectral with multi-band data is used to identify landslide and bareland based on a series of spectral intensities of the band reflection (see Figure 1). Therefore, the study aims to answer the question on whether the hyperspectral data can substitute for DEM data or not. On the other hand, landslide and bareland differentiation have drawn more attention to scientists and researchers. Landslide and bareland both have the same ingredient of soil but are usually at different locations on the hill. If the spectrum similarity analysis can be done to determine these two different categories, it could reduce a great amount of time in generating the DEM data/Map.

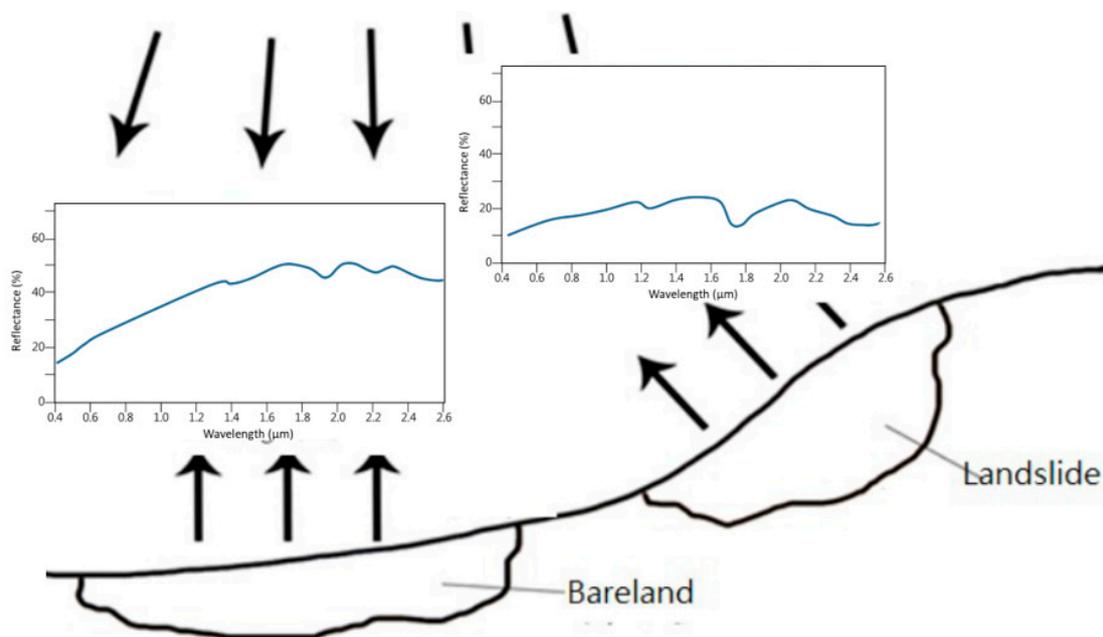


Figure 1. Theory for different reflection of target category (bareland vs. landslide).

In parallel studies, this study intends to use data mining methods: Support Vector Machine (SVM). A total of 72 spectral data of hyper-spectrum remote sensing images are distinguished from the traditional high-resolution data of traditional R, G, B and IR images which can clearly resolve the topography of the surface. If each category is carefully determined, it will be beneficial to compare them by similarity analysis. Between the classification of landslide and bareland, various machine learning classifiers may have different characteristics and solutions. The classification of the image can

be conducted either in stage or in combination with each other. Therefore, the study intends to adopt the following two approaches: (a) Spectrum analysis and (b) Support Vector Machine (SVM). These two approaches are used to compare the outcomes in advantages and disadvantages, respectively.

2. Data Collection for Study Plan and Area

The study area is located in Zhuzizao Mountain, Nanxi District, Tainan City, Taiwan. It is located at the northeastern end of Tainan City, north of Dongshan District and Taipu Township of Chiayi, adjacent to Nanhua District in the east, Liujia District and Dazhong District in the west and the south of Yujing District. Nanxi District is located at the tail edge of the Alishan Mountain. The central part is the Dawu Ridge Basin. The hyperspectral image telemetry can reach a large area of the empirical area. To achieve the control and prediction of the collapse disaster, this study used the image data from the Chung-Hsing measurement Company in 2016. They purchased the UAV (Unmanned Aerial Vehicle), which is used to capture the hyperspectral image of Zhuzishan Mountain in Nanxi District, Tainan City for the study material.

2.1. Geomorphology

According to the plan of the Tainan City Landslide and Geostrophic Geological Sensitive Area (2014), the Nanxi District belongs to the river valley zone. Owing to the river originating from the Eastern Mountain, it shows a remarkable stream of excavation. At the same time, the cliff end with erosion is produced. A series of river bank terraces are formed under the action of undercut and side erosion; therefore, a small-scale vertical valley development is formed. The study area is located in the east of Meiling Scenic Area with an elevation of 1110 m. To the west, overlooking the Jianan Plain, the northwest side overlooks the Zengwen Reservoir, and the southeast is the Nanhua Reservoir. The terrain of this area has a large height difference in elevation, which is mainly composed of hilly terrain and plain terrain. The average elevation is between 800 and 1300 m. The geology is mainly composed of accumulated soil, and there are faults on both sides—the east and west. The earthquake-induced landslide caused this area soil condition to be very fragile.

2.2. Hydrographic System

The rivers in Tainan City include Bazhangxi, Jiushuixi, Zengwenxi, Yanxi, and Errenxi. The Ziwen River Basin originates from the Alishan Mountains. The drainage area is 1176.6 square kilometers and the longest is 138.5 km. The average slope of the riverbed is 1/200. The main tributaries are Houtunxi, Caixixi, and Guantianxi. It flows through Dongshan, Liujia, Annan, Yujing, Nanhua, Zuozhen, Shanshang, Dain, Guantian, Shanhua Madou, Anding, Xigang, and Qiqi on the Nanxi District of the study area, respectively. The study area is located near Tainan County in which there is Zengwen Reservoir (the largest reservoir in southern Taiwan). The mainstream originates from the Alishan Mountains, flows south to Zengwenxi, and flows southwest through the mountainous area to the Zengwen Reservoir. The strip has a total length of 138.5 km, an average slope of 1/57, and an average annual rainfall of about 2726 mm.

2.3. Geological Structure

The Tainan City Regional Disaster Prevention Plan (2016) is based on the data released by the Central Geological Survey of the Ministry of Economic Affairs in December 2016. It attributes to the historical landslide and ground slide area of about 69.11 square kilometers with landslide or ground slip conditions (with a sloping slope). The area is about 50.4 square kilometers with the buffer zone of 5 m is about 21.99 square kilometers, and the demarcation range is about 0.62 square kilometers. The total area is about 135.45 square kilometers (about 6.18% the total of area city).

In Figure 2, the location map of the landslide and geostrophic geological sensitive area in Tainan City is a plan for the Tainan landslide and geostrophic geological sensitive area (2016). The figure

shows that, to increase the terrain steepness and aspect, the base map is overlaid with topographic shadow maps and the adjacent administrative boundaries.

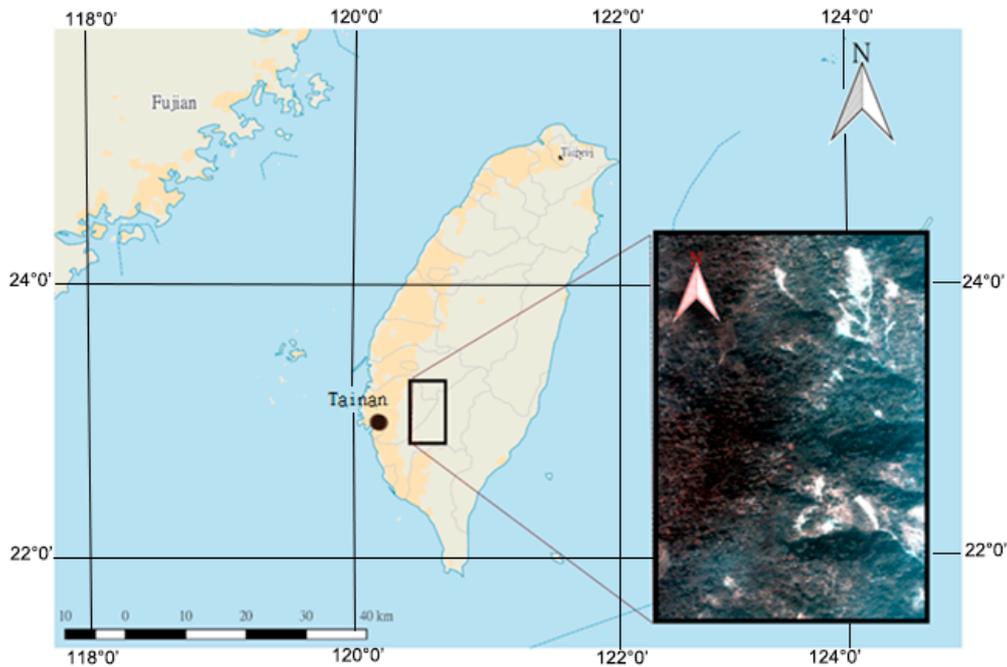


Figure 2. Study area.

2.4. Research Material

The spectral application image used in this study is the hyperspectral image of the Compact Airborne Spectrographic Imager (CASI) of the Bamboo-Waste Mountain in Nanxi District, Tainan City, which was provided by Taiwan Chung-Hsing Measurement in January and April 2016 as shown in Figure 3. The image scanning system CASI-1500 is manufactured by ITRES of Calgary, AB, Canada. The CASI-1500 instrument has a series of spectral wavelengths between 365 nm and 1050 nm, which is equivalent to the visible of near-infrared range. It can acquire 72 bands for this study with a spectral resolution of 3 nm and a spatial resolution of 1 m. Each band has its range and attribute of color, which is presented in Figure 3. Thus, the corresponding number of bands in the latter parts of this study is the same number presented here.

Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	Blue															Green												Red								
nm	366	375	385	394	404	413	423	432	442	452	461	471	480	490	499	509	518	528	537	547	556	566	575	585	594	604	613	623	632	642	651	661	671	680	690	699
Band	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
	Near infrared																																			
nm	709	718	728	737	747	756	766	775	785	794	804	813	823	832	842	851	861	870	880	889	899	908	918	927	937	946	956	965	975	984	994	1003	1013	1022	1032	1041

Figure 3. The hyperspectral corresponding image band of range.

3. Research Method

3.1. Spectrum Similarity Analysis

We carefully selected 240 sampling data for vegetation areas (trees, grass, etc), bareland area, and landslide area, respectively. Spectrum similarity analysis becomes a well-accepted approach to reduce the data dimensionality of hyperspectral imagery. It retrieves several bands of important patterns in some sense by taking advantage of the all high spectral correlation. Verified by classification accuracy, it was expected that, just using a part of original bands, the accuracy is obtained rationally, whereas computational work is significantly reduced [18,19].

Figure 4a shows the entire research step. It includes two parallel approaches. One of the approaches is considering finding the similarity of the image bands width to attain the classification outcomes [20]. Figure 4b shows the similarity of image bands. The vegetation index threshold is found based on clustering analysis. The non-vegetation of the image is attained, which includes the bareland and landslide. All this is part of data normalization. Then, the progress of the similarity spectrum analysis is carried out. Two of the image layers are obtained (D_1 and D_2). The latter part of this paper will introduce the details on how the similarity classification of each pixel is identified.

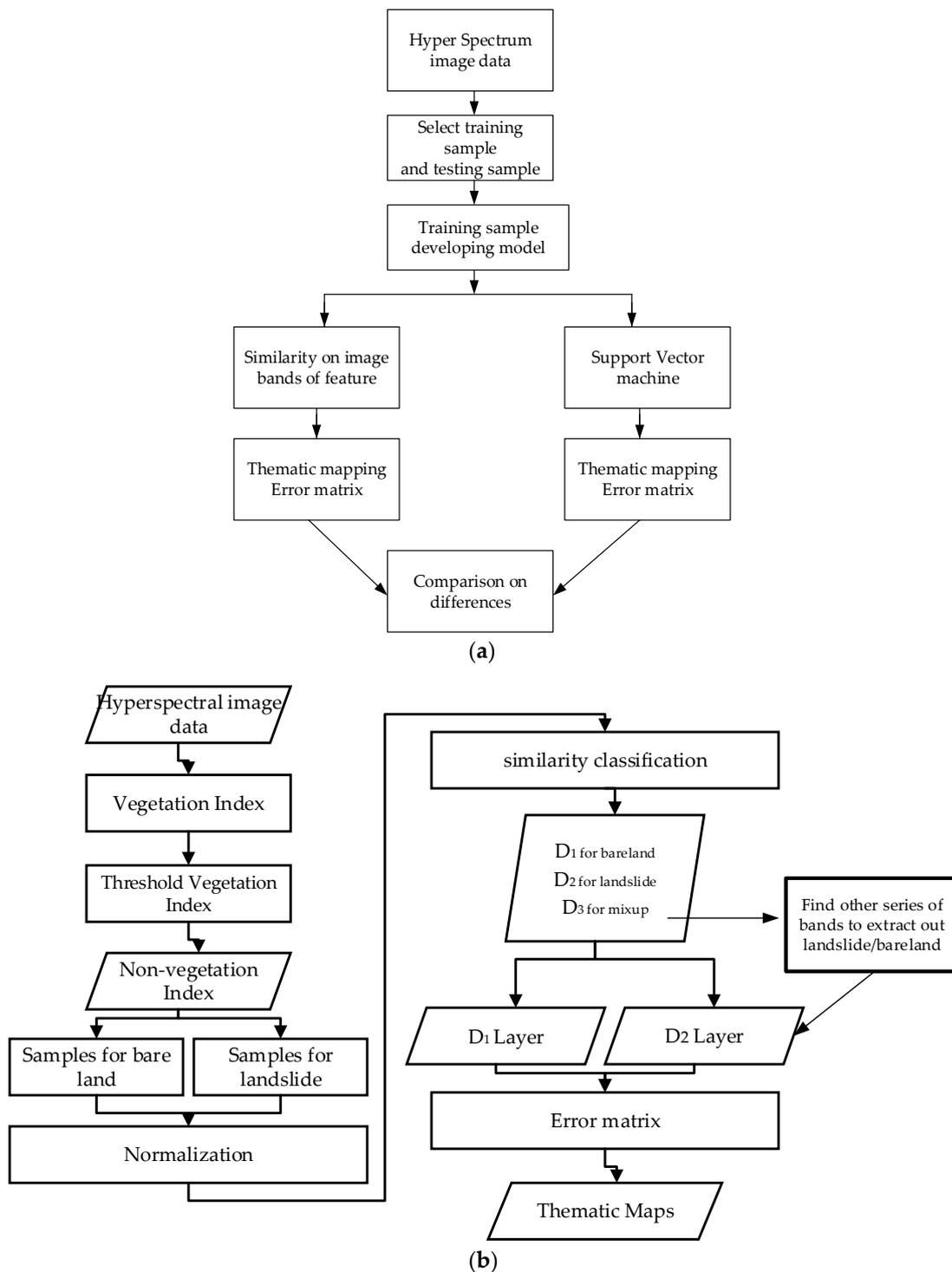


Figure 4. (a) research steps; (b) the steps for similarity classification.

Figure 5a presents the original investigation on the site for observing the location of landslide. All the image data for similarity were carefully checked by in situ investigation and compared to remote sensing data [21]. It was decided to extract several samples as mentioned above for landslide and bareland. Thus, Figure 5b shows the longitude and latitude of the position of study and the accurate place of landslides. This area landslide belongs to block-slide. Block slide is a kind of translational slide. The moving mass of soil and rocks has several related units that move downslope as a relatively coherent mass. The largest size of the landslide is about $8 \times 12 \text{ m}^2$, which is roughly measured by image data.

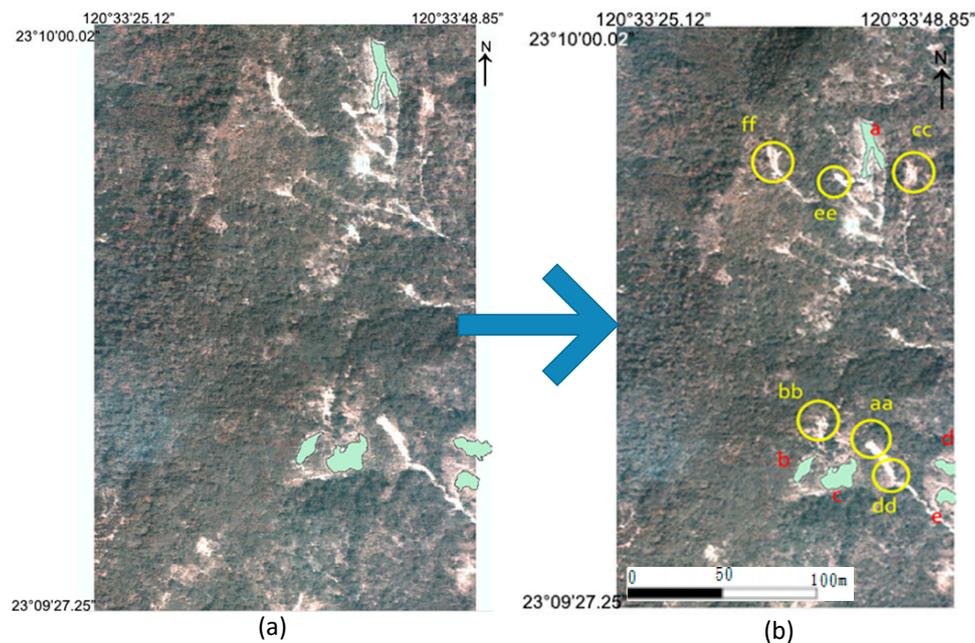


Figure 5. (a) original inventory map: landslide location of in situ investigation; (b) location of training sample, landslide: a, b, c, d, and e; bareland: aa, bb, cc, dd, ee, and ff.

The parallel study was used the SVM (Support Vector Machine) to access the classification on bareland and landslide. The thematic map is compared and the error matrix is also calculated.

3.2. Brief on Support Vector Machine

Support vector machines (SVMs) are well-accepted supervised learning methods used for classification [22]. The study considers the concept of improving statistical learning theory, generally applied as an effective classifier to solve many practical problems [23]. A special feature of this classifiers is to minimize the empirical classification error and maximize the geometric margin, simultaneously. Therefore, it is also known as a maximum margin classifier [24,25].

Linearly separable classes are the simplest cases for the analysis of three various classes (vegetation, landslide, and bareland). Assume the training data with k number of samples are presented as $\{x_i, y_i\}$, where $x \in R^N$ with an n -dimensional space, and $y \in \{+1, -1\}$ is the class label. These training patterns are linearly separable if there exists a vector w (determining the orientation of a discriminating plane) and a scalar b (determine the offset of the discriminating plane from the origin) such that

$$y_i(w \cdot x_i + b) - 1 \geq 0. \quad (1)$$

The hypothesis space is defined by the set of functions given by:

$$f_{w,b} = \text{sign}(w \cdot x + b). \quad (2)$$

If the set of examples is linearly separable, the goal of the SVMs is to minimize the value $\|w_i\|$. It is equivalent to finding the separating hyperplanes for which the distance between the classes of training data. It also measured along a line perpendicular to the hyperplane.

This distance is called the margin. The data points that are closest to the hyperplane are used to measure the margin. Thus, these data points are also called support vectors. Consequently, the number of support vectors should be small.

The problem of minimizing $\|w_i\|$ is solved by applying standard quadratic programming (QP) optimization techniques. It also transforms the problem to a dual space by using Lagrangian multipliers. The Lagrangian is presented by introducing positive Lagrange multipliers λ_i , $i = 1, \dots, k$. The solution of the optimization problem is attained by considering the saddle point of the Lagrange function

$$L(w, b, \lambda) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^k \lambda_i y_i (w \cdot x_i + b) + \sum_{i=1}^k \lambda_i. \quad (3)$$

The solution in Equation (5) needs $L(w, b, \lambda)$ to be minimized with respect to w and b and maximized with respect to $\lambda_i \geq 0$. Therefore, for a two-class problem, the decision rule separates the two classes that can be written as:

$$f(x) = \text{sign} \left(\sum_{i=1}^k \lambda_i y_i (x \cdot x_i) + b \right). \quad (4)$$

A soft margin problem for the case of SVMs is to handle the linearly non-separable data by Vapnik [22]. They concluded that the restriction of each training vector of a given class on the same side of the optimal hyperplane that applies the value. In $\xi_i \geq 0$, the SVM algorithm for the hyperplane maximizes the margin. At the same time, it minimizes a quantity proportional to the number of misclassification errors. This trade-off function between margin and misclassification error is also governed by a positive constant C such that $\infty > C > 0$. Thus, for non-separable data, Label (6) can be written as

$$L(w, b, \lambda, \xi, \mu) = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i - \sum_i \lambda_i \{y_i (w \cdot x_i + b) - 1 + \xi_i\} - \sum_i \mu_i \xi_i, \quad (5)$$

where the μ_i are the Lagrange multipliers introduced to force the ξ_i to be positive. The solution of (7) is determined by the saddle points of the Lagrangian, by minimizing with respect to w , x , and b , and maximizing with respect to $\xi_i \geq 0$ and $\mu_i \geq 0$.

4. Results

As aforementioned, we select 120 of sampling data for training the model of vegetation, bareland, and landslide, respectively. The 40 pieces of data of each (vegetation, bareland, and landslide) categories to build the model. The study also randomly selects 40 pieces of data to verify the model as testing data. The study has been broken into two parts: spectral similarity analysis and support vector machine. As previously mentioned in Figure 1, the landslide mostly occurred on the slope that has the different responses of reflection on hyper-spectrum image data. Compared to the bareland, the ingredient of soil is the same as the landslide; however, most of them are located in the flat area. Thus, the reflection on hyper-spectrum image data must be different to a landslide. This is the objective to classify them by applying the similarity of a spectrum [26].

To introduce the overall accuracy, it can be formulated as

$$Ac = \frac{TP + TN}{\text{All the samples}}, \quad (6)$$

where TP is the true positive and TN is the true negative.

4.1. Spectral Similarity Analysis

The spectrum analysis entire study is divided into the following two steps:

1. Classify the vegetated areas and non-vegetated areas (similar to [18]).
2. Use the clustering analysis to separate bareland areas and landslide areas from non-vegetated areas (similar to [19]).

To achieve this task, the developed program scans all the bands to find the largest discrepancy of vegetation and non-vegetation for discriminating between these two categories. The program calculates and finds that the 34th band has the largest difference. The green lines in Figure 6a are rationally extracted. Figure 6b is generated to extract out vegetation parts (green line). The r -value on the y -axis is the response value of the reflection for various categories (vegetation, bareland, and landslide).

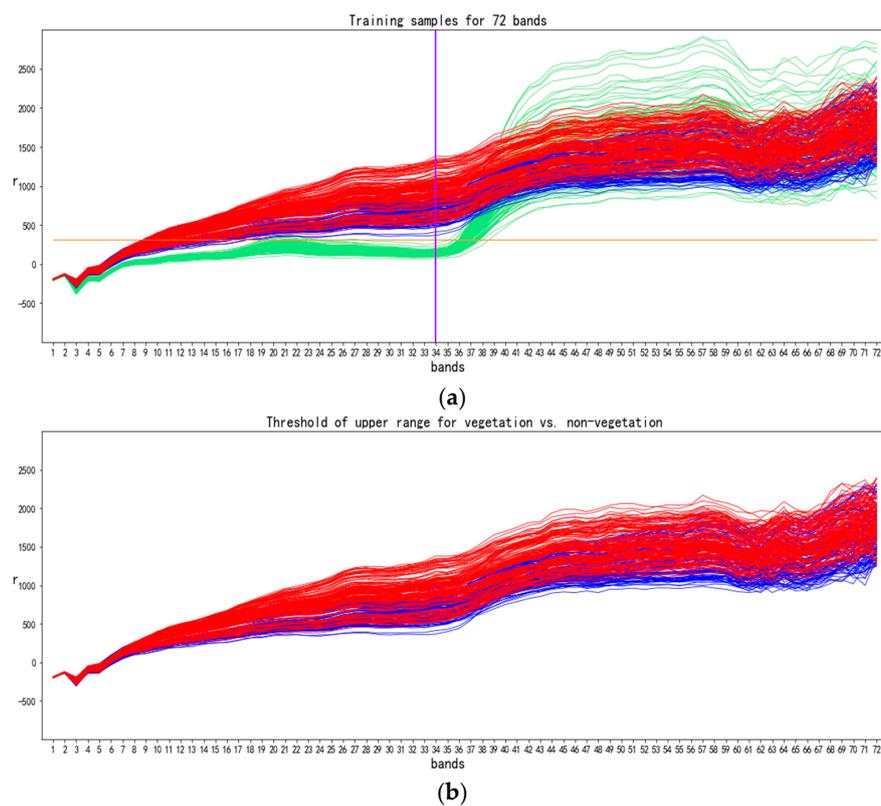


Figure 6. Cont.

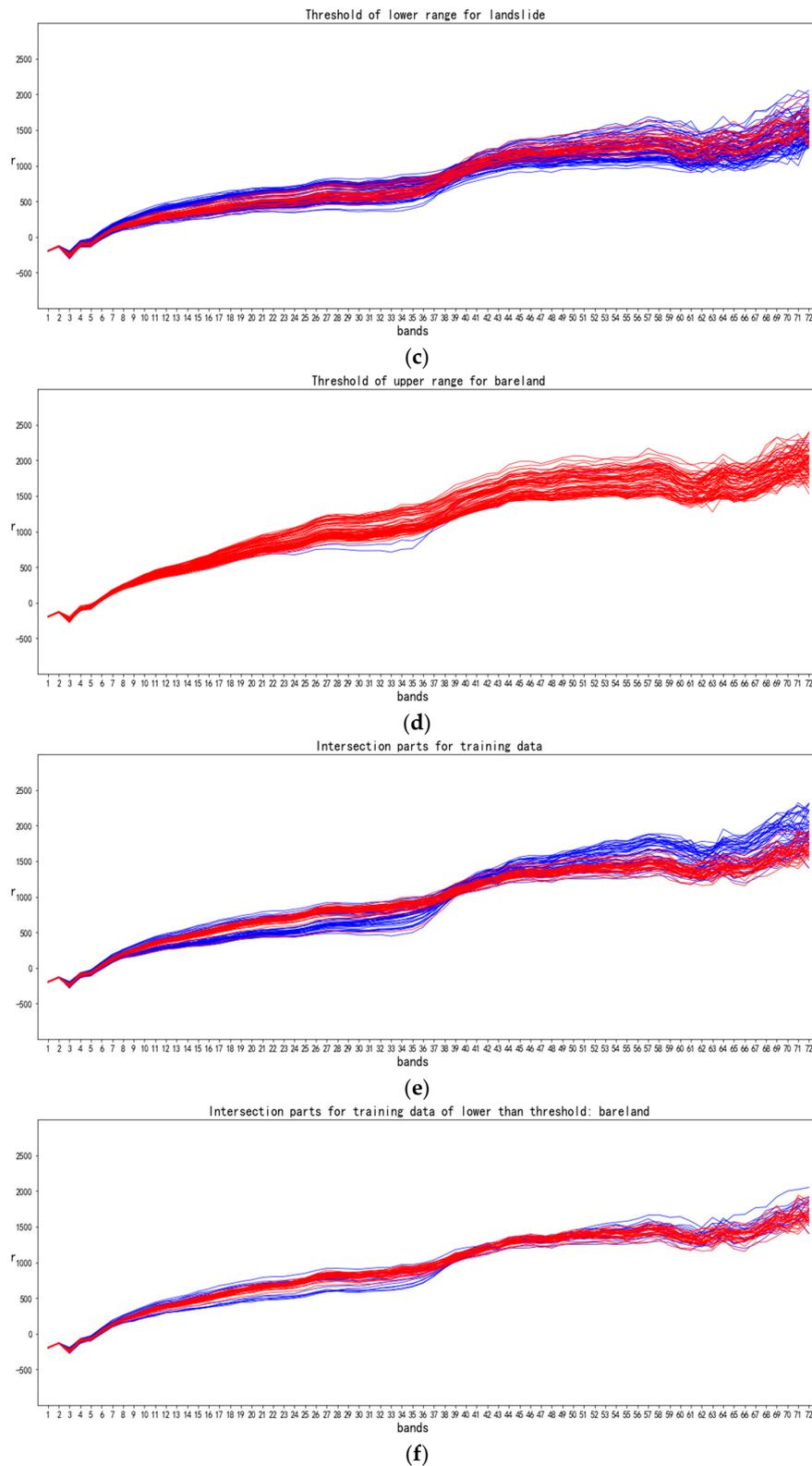


Figure 6. (a) three different categories for classification; (b) threshold of upper range for vegetation and non-vegetation; (c) the lower range on the threshold for landslide; (d) the upper range on the threshold for bareland; (e) intersection parts for training data of larger than threshold: landslide; (f) intersection parts for training data of lower than threshold: bareland (green: vegetation; blue: landslide; red: bareland).

To obtain the best classification outcomes, the 1–72 bands are scanned to find the best part to distinguish the landslide and bareland. The developed program scans the data in Figure 6c to find the maximum discrepancy for landslide and bareland. A single cannot clarify the mixup data for landslide and bareland. Thus, a combination set of bands are requested to approach the goal. It is found that 38 to 42 bands are the best part in the spectrum analysis to attain the classification outcomes. First, the program adds the 38–42 bands of each data as a single band data (transfer the five-dimensional data to the one-dimensional data).

Then, the program generates a parametric r -value as an interval to attain three parts of the data: (a) lower the threshold(landslide), (b) upper the threshold(bareland), and (c) intersection part (mix-up part). Please refer to Figure 6b; Figure 6c; Figure 6d; and Figure 6e. The program gradually increases the r value to approach the optimal classification outcomes for landslide and bareland. For example, the program starts $r = 80$ and $\Delta r = 5$ and finds the error rate between classification on landslide and bareland. That is, the program gains $r = 95$, which is the best allowable value to cut the data into these three parts. The strategy is to approach the largest number of lower the threshold and upper the threshold. The minimal lowest number of intersection part is also requested. The program calculates each data after the summation and sets them as less than 95 as one group and greater than 95 as another. The strategy is the number of data of the largest group to the total number of data must be greater than 40%. The number of data of the smallest group to the total number of data must be smaller than 40%. Because in Data Mining, the portion of the number of data for each decision should be as close as possible. Applying these sets of data can be fairly and uniformly to develop the model.

Then, the program three parts for summing up band 38 to 42 is

$$\left\{ \begin{array}{l} \text{Lower than intersection} \leq 4105 \dots \text{landslide} \\ \text{Greater than intersection} \geq 4828 \dots \text{bareland} \\ \text{intersection} > 4105 \text{ and } < 4828 \dots \text{mixup parts} \end{array} \right. \quad (7)$$

After screening the band data, it is found that the data density variety is not uniform. Hence, different stepwise of a grouping data strategy is needed. In the mix-up parts, the program restarts to find the discrepancy between landslide and bareland. The solution takes a set of band values and uses the clustering technique to search the optimal set of possible outcomes. For instance, we found that the band numbers from 45 to 52 has the largest discrepancy. Then, the program sieves out the 45, 46, 48, 50, and 51 bands are the most useful information. That is, 47, 49, and 52 bands are eliminated from the data set. The program found that the band values in 45, 46, 48, 50, and 51 have the largest discrepancy between landslide and bareland. Then, the intersection ranges of bands of each piece of data are summed into a single value (five multi-band data into one-dimensional data). The summed maximum and maximum values are calculated, and the binary classification is executed. It is found that a finer value of 30 can be gradually increased as a stepwise to each line attribute for each categories (landslide and bareland). Then, the accuracy of each segmentation value is step-by-step calculated as the classification accuracy until the highest accuracy is approached. Each band of the data in the intersection range is clustered based on the rule of a finer interval being less than 30; the other group is greater than 30.

Determination value of 45, 46, 48, 50 and 51

$$\left\{ \begin{array}{l} \text{sum } r : \geq 6995 \dots \text{landslide} \\ \text{sum } r : < 6995 \dots \text{bareland} \end{array} \right. \quad (8)$$

The training data for generating this similarity model have 100% accuracy.

The thematic map (see Figure 7) is generated by inputting are the band data into the program. Green presents the vegetation, red for landslide and white for bareland. The major landslide areas (comparing in Figure 5) are almost found, and bareland is clearly found. The computational time is

fast and a rough result is qualified. We also randomly picked 40 testing data to verify our spectrum analysis model. The error matrix is presented in Table 1. The overall accuracy is about 85%.

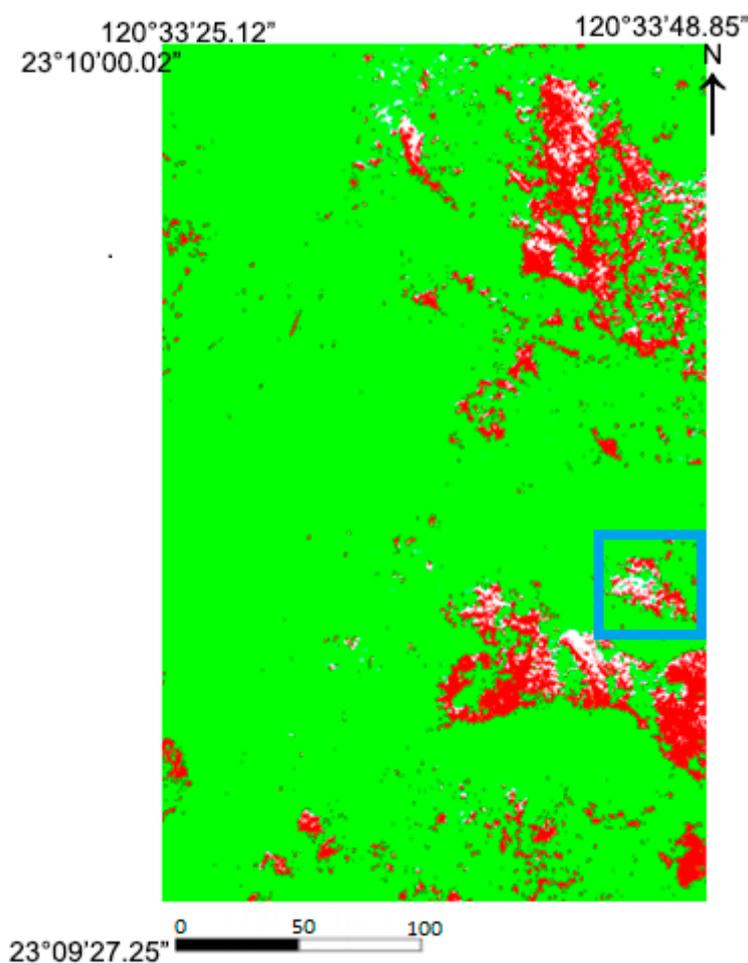


Figure 7. The thematic of spectrum analysis (green: vegetation, red: landslide and white: bareland).

Table 1. Error matrix for spectrum analysis.

Categories	Vegetation	Bareland	Landslide	User Accuracy	Commission Error
Vegetation	40	0	0	100%	0%
Bareland	0	32	8	80%	20%
Landslide	0	10	30	75%	25%
Producer Accuracy	100%	76.19%	78.95%	Overall accuracy = 85%	
Omission error	0%	23.81%	21.05%		

4.2. SVM

As part of the study, the Support Vector Machine is used as a parallel approach to examine the spectrum analysis. The objective of the support vector machine algorithm is to generate a hyperplane in n -dimensional space (n is the number of features) that accurately classifies the data points. The following steps are:

4.2.1. Step1: Normalization

The original data of the collected data sets (such as hyperspectral, multi-spectral, etc.) are normalized, and the values of the attribute data are standardized within the same range. This study converts all attribute values between -1 and 1 , using the formula:

$$d = \frac{d - \min_d}{((\max_d - \min_d) - 0.5) * 2}. \quad (9)$$

4.2.2. Step2: Cross-Validation

This study uses K-Fold Cross-Validation to first split the initial sample into K sub-samples (each sub-sample is independent from each other). A single sub-sample is the data for the validation model with the remaining $K-1$ samples. One of those sets of sub-samples is used for training. After repeating the above procedure K times, the K group classification correct rate will be obtained. In final, the data of the K group for the correct rate average value are estimated.

4.2.3. Step3: Model Selection for a Core Function

The functions of the support vector machine can be divided into four types: linear functions, polynomial functions, radial basis functions, and S functions. The user should select the core function based on different conditions. The parameters are adjusted for different kernel functions that are also different. The user has to adjust the kernel function and parameters according to the situation, which will have a significant impact on the prediction accuracy rate. In this study, the Radial Basis Function kernel (RBF) is taken for consideration. To obtain better model parameters, the Grid Search method repeats the test parameters $C = 4.2$ (penalty parameter) and $g = 0.32$ (gamma function) for possible combination and calculate the correct rate of its parameters (C , g). If it meets its condition, end the repeated test and output its best C and g parameters; otherwise, re-substitute with the new parameters until the combination is found.

This step is to optimize the optimal classification model obtained in the previous step. The testing data of the unknown result are substituted into the classification model construct by the previous step, and the obtained result will be aggregated in which the overall classification accuracy rate is calculated for performing the evaluation. It explores the effectiveness of machine learning under its selection points and different attribute data. The accuracy assessment of this study is divided into two parts: (1) the thematic map and (2) the error matrix. Figure 8 presents the thematic map for the overall condition in three categories. Green presents the vegetation, red for landslide, and white for bareland. Comparing Figure 8 to Figure 7, based on the image data in Figure 5, it presents a clearer and better accurate rate for the thematic map. The error matrix is also calculated in Table 2. The overall accuracy is 98.3%.

Comparing Figures 7 and 8 for the thematic map, the difference is clear. For instance, the blue rectangle part in Figure 7 renders a better interpretation of detecting the bareland. In the inventory map (Figure 5), this part presents as a bareland. The similarity analysis spectrum seems to provide a better prediction. However, SVM has a better interpretation of the integrity on landslide and bareland. The spatial information is a fundamental multi-temporal approach. The method can successfully be applied to several periods of this area or another area. Thus, if the based rule of similarity spectrum can be developed successfully, the approximated location of landslide mapping can be rapidly generated.

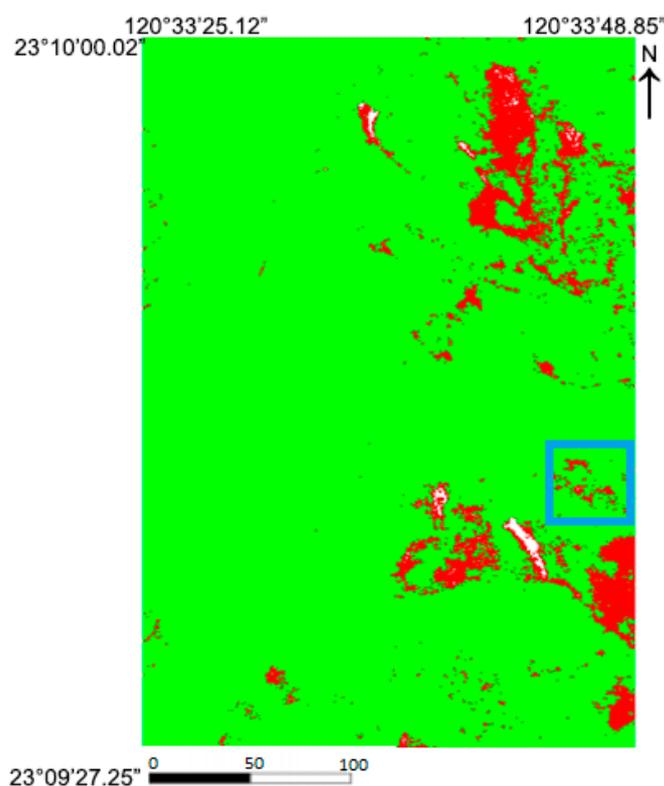


Figure 8. The thematic of support vector machine (green: vegetation, red: landslide and white: bareland).

Table 2. Error matrix for SVM.

Categories	Vegetation	Bareland	Bareland	User Accuracy	Commission
Vegetation	40	0	0	100%	0%
Bareland	0	38	0	100%	0%
Landslide	0	2	40	95.2%	4.8%
Producer	100%	95%	100%	Overall accuracy = 98.3%	
Omission error	0%	5%	0%		

5. Conclusions

The landslide and bareland are the most interesting topics that draw great attention to scientists and researchers. They both have the same soil ingredient but different locations on the hill. Landslides mostly displayed on the hill, which may produce destructive disasters for human beings. Owing to the lack of accurate DEM (Digital Elevation Modeling) map/data in this study, the hyperspectral data have been proved to identify landslide and bareland according to spectral intensities of reflection. A parallel study is designed to compare the spectral analysis approaches.

The study has three major contributions:

- The study proved that the hyperspectral image data can replace the DEM data by considering different land cover categories.
- The spectral similarity analysis can classify 100% of the vegetation area. Most of the landslide and bareland area is also being detected with a satisfactory level of the overall accuracy of 85%.
- The support vector machine is a superior classifier. However, the problem is that each of the training sample data must be supervised data. That is, each piece of in situ sampling data must be carefully labeled. The overall accuracy is 98.3%.

Author Contributions: S.W.: He was responsible for plan and design of this study. He also helped the student write the computer program. T.C.L.: He analyzed the data and discussion. H.L.M.: He wrote the computer program. R.W.C.: She used the program to plot the thematic map.

Funding: This research was funded by grant number 106-2119-M-275-002.

Acknowledgments: The authors expressed their gratitude for the National Science Council (106-2119-M-275-002)-sponsoring this work.

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

References

1. Van Westen, C.J.; Rengers, N.; Soeters, R. Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment. *Nat. Hazards* **2003**, *30*, 399–419. [[CrossRef](#)]
2. Wan, S.; Lei, T.C.; Huang, P.C.; Chou, T.Y. Knowledge Rules of Debris Flow Event: A Case Study for Investigation ChenYu Lan River, Taiwan. *Eng. Geol.* **2008**, *98*, 102–114. [[CrossRef](#)]
3. Wan, S.; Lei, T.C.; Chou, T.Y. A novel data mining technique of analysis and classification for landslide problems. *Nat. Hazards* **2009**, *52*, 211–230. [[CrossRef](#)]
4. Wan, S. A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map. *Eng. Geol.* **2009**, *108*, 237–251. [[CrossRef](#)]
5. Lazzari, M.; Gioia, D. UAV images and historical aerial-photos for geomorphological analysis and hillslope evolution of the Uggiano medieval archaeological site (Basilicata, southern Italy). *Geomat. Nat. Hazards Risk GNHR* **2017**, *8*, 104–119. [[CrossRef](#)]
6. Lazzari, M.; Piccarreta, M. Landslide disasters triggered by extreme rainfall events: The case of Montescaglioso (Basilicata, southern Italy). *Geosciences* **2018**, *8*, 377. [[CrossRef](#)]
7. Lin, C.Y.; Chuang, C.W.; Lin, W.T.; Chou, W.C. Vegetation recovery and landscape change assessment at Chiufenershan landslide area caused by Chichi earthquake in central Taiwan. *Nat. Hazards* **2010**, *53*, 175–194. [[CrossRef](#)]
8. Pradhatsp, B.; Mansor, S.; Lee, S.; Buchroithner, M.F. Application of a data mining model for landslide hazard mapping. *ISPRS* **2008**, *37*, 187–196.
9. Wan, S.; Lei, T.C.; Chou, T.Y. A landslide expert system: Image classification through integration of data mining approaches for multi-category analysis. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 747–770. [[CrossRef](#)]
10. Kirschbaum, D.B.; Adler, R.; Hong, Y.; Lerner-Lam, A. Evaluation of a preliminary satellite-based landslide hazard algorithm using global landslide inventories. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 673–686. [[CrossRef](#)]
11. Pradhan, B.; Lee, S.; Buchroithner, M.F. Remote sensing and GIS-based landslide susceptibility analysis and its cross-validation in three test areas using a frequency ratio model. *Photogramm. Fernerkund.-Geoinform.* **2010**, *1*, 17–32. [[CrossRef](#)]
12. Carrara, A.; Guzzetti, F.; Cardinali, M.; Reichenbach, P. Use of GIS technology in the prediction and monitoring of landslide hazard. *Nat. Hazards* **1999**, *20*, 117–135. [[CrossRef](#)]
13. Chang, C.I. Spectral Information Divergence for Hyperspectral Image Analysis. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Hamburg, Germany, 28 June–2 July 1999; Volume 1, pp. 509–511.
14. Aspinall, R.J.; Marcus, W.A.; Boardman, J.W. Considerations in Collecting, Processing, and Analysing High Spatial Resolution Hyperspectral Data for Environmental Investigations. *J. Geogr. Syst.* **2002**, *4*, 15–29. [[CrossRef](#)]
15. Borgogno Mondino, E.; Giardino, M.; Perotti, L. A neural network method for analysis of hyperspectral imagery with application to the Cassas landslide (Susa Valley, NW-Italy). *Geomorphology* **2009**, *110*, 20–27. [[CrossRef](#)]
16. Aguilera, C.A.; Aguilera, F.J.; Sappa, A.D.; Toledo, R. Learning Cross-Spectral Similarity Measures with Deep Convolutional Neural Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 267–275.
17. Iscen, A.; Avrithis, Y.; Toliass, G.; Furon, T.; Chum, O. Fast Spectral Ranking for Similarity Search. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7632–7641.

18. Hamad, D.; Biela, P. Introduction to Spectral Clustering. In Proceedings of the 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, Syria, 7–11 April 2008.
19. Nadler, B.; Galun, M. Fundamental limitations of spectral clustering. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc.: Vancouver, BC, Canada, 2007; pp. 1017–1024.
20. Fan, C.Y.; Fan, P.S.; Chan, T.Y.; Chang, S.H. Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals. *Expert Syst. Appl.* **2012**, *39*, 8844–8851. [[CrossRef](#)]
21. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.T. Landslide inventory maps: New tools for an old problem. *Earth-Sci. Rev.* **2012**, *112*, 42–86. [[CrossRef](#)]
22. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995; pp. 11–24.
23. Kecman, V. *Learning and Soft Computing, Support Vector machines, Neural Networks and Fuzzy Logic Models*; The MIT Press: Cambridge, MA, USA, 2001.
24. Khandoker, A.H.; Palaniswami, M.; Karmakar, C.K. Support Vector Machines for Automated Recognition of Obstructive Sleep Apnea Syndrome from ECG Recordings. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 37–48. [[CrossRef](#)] [[PubMed](#)]
25. Pal, M. Support vector machines-based modelling of seismic liquefaction potential. *Int. J. Numer. Anal. Methods Geomech.* **2006**, *30*, 983–996. [[CrossRef](#)]
26. Van der Meer, F. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2006**, *8*, 3–17. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).