

Article

Integrating Data Modality and Statistical Learning Methods for Earthquake-Induced Landslide Susceptibility Mapping

Zelang Miao ^{1,2}, Renfeng Peng ^{1,2}, Wei Wang ^{1,2,*} , Qirong Li ³, Shuai Chen ^{1,2}, Anshu Zhang ⁴, Minghui Pu ^{1,2}, Ke Li ^{1,2}, Qinqin Liu ⁵ and Changhao Hu ⁶

¹ School of Geoscience and Info-Physics, Central South University, Changsha 410083, China; zelang.miao@outlook.com (Z.M.); 195012112@csu.edu.cn (R.P.); s.chen1227@csu.edu.cn (S.C.); royalharu@csu.edu.cn (M.P.); ke.lichangsha@csu.edu.cn (K.L.)

² Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Central South University, Changsha 410083, China

³ The Yangtze Three Gorges Technology and Economy Development Co., Ltd., Beijing 100038, China; qirong.li@csu.edu.cn

⁴ Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Kowloon, Hong Kong; anshu.zhang@connect.polyu.hk

⁵ National Institute of Natural Hazards, Ministry of Emergency Management of China, Beijing 100085, China; qinqinliu@ninhm.ac.cn

⁶ Hunan Lianzhi Technology Co., Ltd., Changsha 410000, China; 15111019693@163.com

* Correspondence: wangweicn@csu.edu.cn

Abstract: Earthquakes induce landslides worldwide every year that may cause massive fatalities and financial losses. Precise and timely landslide susceptibility mapping (LSM) is significant for landslide hazard assessment and mitigation in earthquake-affected areas. State-of-the-art LSM approaches connect causative factors from various sources without considering the fusion of different information at the data modal level. To exploit the complementary information of different modalities and boost LSM accuracy, this study presents a new LSM model that integrates data modality and machine learning methods. The presented method first groups causative factors into different modal types based on their intrinsic characteristics, followed by the calculation of the pairwise similarity of modal data. The similarities of different modalities are fused using nonlinear graph fusion to generate a unified graph, which is subsequently classified using different machine learning methods to produce final LSM. Experimental results suggest that the presented method achieves higher performance than existing LSM methods. This study provides a new solution for producing precise LSM from a fusion perspective that can be applied to minimize the potential landslide risk and for sustainable use of erosion-prone slopes.

Keywords: landslide susceptibility mapping; earthquake-induced landslide; data modality; information fusion



Citation: Miao, Z.; Peng, R.; Wang, W.; Li, Q.; Chen, S.; Zhang, A.; Pu, M.; Li, K.; Liu, Q.; Hu, C. Integrating Data Modality and Statistical Learning Methods for Earthquake-Induced Landslide Susceptibility Mapping. *Appl. Sci.* **2022**, *12*, 1760. <https://doi.org/10.3390/app12031760>

Academic Editor: Francesco Fiorillo

Received: 8 January 2022

Accepted: 29 January 2022

Published: 8 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Landslide, one of the most destructive geological hazards in the world, often results in massive casualties and property losses [1–4]. The earthquake is a vital causative factor that triggers numerous landslides throughout the earthquake-affected area [5]. For instance, the 2008 Wenchuan Mw7.9 earth-quake induced more than 20,000 landslides [6,7]. Landslide susceptibility mapping (LSM) refers to predicting potential landslides in an area depending on a range of causative factors [8–12]. It provides a beneficial reference to reflect the spatial distribution and the susceptibility level of landslide hazards, and thus has now become a common tool in addressing landslide risk reduction [13].

A considerable amount of literature related to LSM has been published and can be grouped into three main categories: (1) physically-based methods, (2) knowledge-based methods, and (3) data-based methods [3,14–16]. Physically-based methods typically use

limit-equilibrium to analyze the slope stability [17,18]. For instance, Martin et al. [19] proposed a three-dimensional limit equilibrium slope stability model that is robust and capable to deal with shallow and deep landslides. They applied this method to assess slope stability in Umbria, central Italy, and it yielded satisfactory results. Kukemilks et al. [20] integrated hydrogeological and slope stability models to identify potential landslides. Though this kind of method does not need a landslide inventory, it requires exhaustive geotechnical and engineering geological data that are challenging and expensive to obtain. Thus, physically-based methods are restricted to assessing landslide susceptibility at the site-specific scale, and are unsuitable for regional/global scale. The knowledge-driven methods are based on expert knowledge on the contribution of various causative factors to landslides. In general, experts provide detailed descriptions on causative factors within the landscape and the contribution of each factor to landslides. Such knowledge is then assimilated into a certain knowledge-driven model to produce a landslide susceptibility score for each location (e.g., each pixel) as a weighted sum of the contribution by various causative factors. Representative approaches for knowledge-driven LSM include analytic hierarchy process (AHP) [21] and fuzzy logic (FR) assessment [6]. In the AHP approach, a comparison matrix is constructed to represent the expert judgment on pairwise relative importance of the causative factors for landslides. The weight of each causative factor for the landslide susceptibility score is then derived from the comparison matrix. In the FR approach, a fuzzy membership function for each causative factor is constructed to represent expert opinions on its contribution to landslides. The landslide susceptibility score is then generated for every pixel by integrating the fuzzy membership values of various causative factors. Although the knowledge-driven methods are simple and efficient, their results are sensitive to the correctness of expert knowledge. Accurate expert inputs into the knowledge-driven models, such as the comparison matrix in AHP and the fuzzy membership functions in the FR assessment, not only rely on extensive domain knowledge, but also are specific to regions with certain geological structures and climates. This limits the transferability of knowledge-driven models to other geographical regions and this limitation is almost similar to the physically-based methods. In particular, for urgent landslide susceptibility assessments (e.g., post-earthquake ones), the assessment results can be questionable if the area is unfamiliar to experts. Data-driven methods use machine learning algorithms to analyze the correlation between landslide distribution and causative factors to assess landslide susceptibility. Regional-scale LSM represents a significant leap forward from previous studies, and integrates the Geographic Information System and the modern advent of machine learning (e.g., artificial intelligence). Typical machine learning algorithms include support vector machines [17], logistic regression [22], information value [23], weight of evidence [24], index of entropy [25], frequency ratio [25], artificial neural networks [26], and random forest [27]. Machine learning algorithms are considered more suitable for LSM over large and complex areas than physically-based methods and knowledge-based methods. Data-driven LSM models have their own advantages and disadvantages. These methods are unable to eliminate the differences of causative factors in terms of magnitude and nature, and also have other problems such as low training efficiency and negative influence by several modeling parameters.

For example, the spatial relationship between landslide locations and the parameters that affect them cannot be calculated in machine learning method [26,28,29]. The model requires a large effort to collect and validate the necessary input data, and model preparation takes a long time. In addition, the performance of the model is vulnerable to the study area and has low generalization; therefore, it is difficult to compare the susceptibility classes from different locations [30]. The means to improve the performance of the LSM model have attracted increasing attention. At present, methods related to improving model performance mainly focus on the following aspects: (1) improving the accuracy of non-landslide sample points selection or the appropriate proportion of sample selection [31,32]; (2) assigning weights among different causative factors [3,33–36]; (3) optimizing the internal parameters of assessment models [37–40]; (4) assessment by region after delineation of the

study area [41]; and (5) integrating different assessment models [9,42–45]. In essence, these studies explore and improve the assessment methods or processes, but do not conduct in-depth analysis from the perspectives of the original data from multiple modalities. They link the features from different causative factors directly without considering the possible impact of different data modalities on the performance of LSM. As causative factors of earthquake-induced landslides are various, such as seismology, terrain, geology, hydrology, and human activity, different data modalities reflect different facets of impacts on landslide occurrence and contain complementary information for LSM. Thus, it would be an interesting direction to fuse multiple modal data to further improve the LSM accuracy.

In this paper, through treating the data of different types as heterogeneous data, we distinguish them with different modalities, and subsequently fuse the characteristic data at the data modalities level. Thus, the complementary information of different modalities can be fully considered. We develop a new framework for landslide susceptibility assessment by taking into account the different data modalities and using nonlinear fusion to fuse the different modal data simultaneously. After that, we utilize the fused data to perform the landslide susceptibility mapping. To validate the performance of the presented LSM model, we take Wenchuan as the study area and compare different LSM results calculated by the multi-modal classification (MMC) model and three benchmark landslide susceptibility assessment models.

2. Research Area and Data Sources

2.1. Research Area Overview

Wenchuan is located at the junction of the northwest region of the Sichuan Basin and the eastern margin of the Tibetan Plateau. The study area covers 4038 km² and is bounded between 30°46′ N–31°00′ N and 102°53′ E–103°44′ E (Figure 1). The region is a very typical landform-changing zone of China and includes lofty mountains, high ridges, and crisscross gorges and valleys. Huge geomorphologic sharp-change zones occur from the southeast area with elevations of 800 m to the west area with elevations of over 6000 m. The region has also recorded many large earthquakes in China as it sits on a highly active seismic area. For instance, the Mw 7.9 Wenchuan earthquake on 12 May 2008 occurred in exactly this region. The synergy between earthquake and the complex geological and topographic conditions has induced, is inducing, and will induce a large number of landslides throughout the region. Therefore, it is critical to timely conduct LSM to prevent and reduce earthquake-induced landslide hazards.

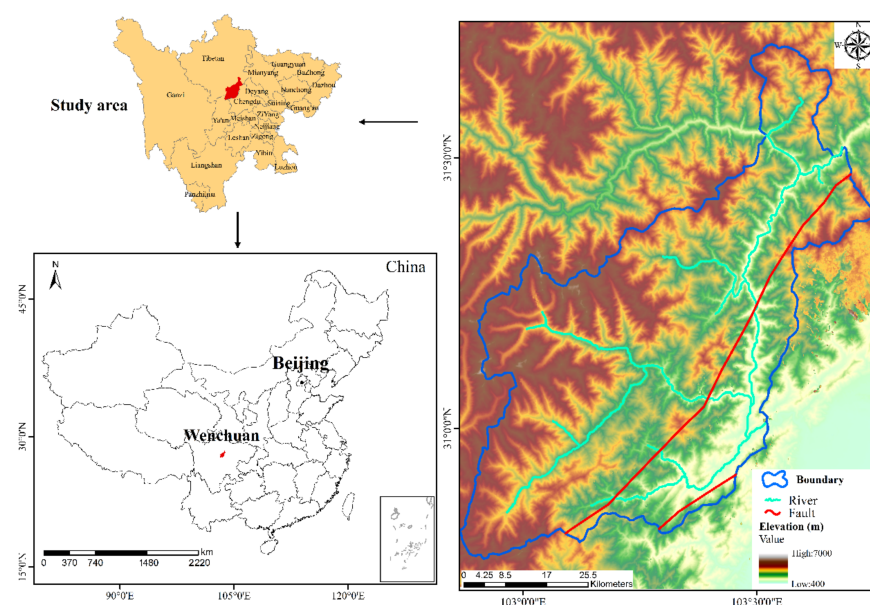


Figure 1. Study area.

2.2. Data Sources Used in This Study

Table 1 lists the sources for data used in this study, and all data are freely accessible. The digital elevation model (DEM) has a spatial resolution of 30 m $\times$ 30 m that is used to generate elevation, slope, aspect, profile curvature, and topographic wetness index (TWI). Distance to faults, distance to roads, and distance to rivers are produced through the buffer analysis function of ArcGIS. The lithology and faults are obtained by digitizing the hard copy of 1:500,000 geological-map. The average annual rainfall is obtained by spatial interpolation based on cumulative annual average rainfall (1981–2010) downloaded from the National Meteorological Science Data Center of China. Continuous causative factors (e.g., elevation, slope, profile curvature, annual average rainfall) are converted to discrete groups using discretization.

Table 1. Data sources used in this study.

Data	Type	Scale	Sources
Digital elevation model (DEM)	Raster	30 m	http://gdex.cr.usgs.gov/gdex/ (accessed on 15 November 2020)
Geological data	Raster	1:500,000	http://geocloud.cgs.gov.cn/#/portal/home (accessed on 15 November 2020)
Annual average rainfall	Datasheet	30 m	http://data.cma.cn/data/detail/dataCode/A.0029.0005.html (accessed on 15 November 2020)
Peak ground acceleration (PGA)	Vector	1:200,000	https://earthquake.usgs.gov/earthquakes/eventpage/usp000g650/shakemap/pga (accessed on 15 November 2020)
Road and river networks	Vector	1:250,000	https://www.webmap.cn/commres.do?method=result25W (accessed on 15 November 2020)
Seismic intensity	Vector	1:200,000	https://earthquake.usgs.gov/earthquakes/eventpage/usp000g650/shakemap/intensity (accessed on 15 November 2020)
Land use	Raster	30 m	http://www.webmap.cn/main.do?method=index (accessed on 15 November 2020)
Soil	Raster	90 m	http://www.resdc.cn/data.aspx?DATAID=145 (accessed on 15 November 2020)
Landform	Raster	90 m	http://www.resdc.cn/data.aspx?DATAID=124 (accessed on 15 November 2020)

This study applies a complete landslide inventory [46] (Figure 2) to train and validate the proposed LSM method. The inventory was produced based on the combination of visual interpretation and field investigation using pre- and post-earthquake multi-source imagery from different platforms (e.g., very high-resolution satellite images, aerial photos, and field surveys). This landslide inventory records detailed landslide information, such as distribution, size, and volume of landslide. To reduce the influence of asynchrony between landslide inventory and causative factors, this study selected DEM, geological data, annual average rainfall, soil, and landform as close as possible to the occurrence time of the Wenchuan earthquake, and accordingly updated land use, road and river networks using very high satellite images acquired from 10 October 2007 to 1 April 2008.

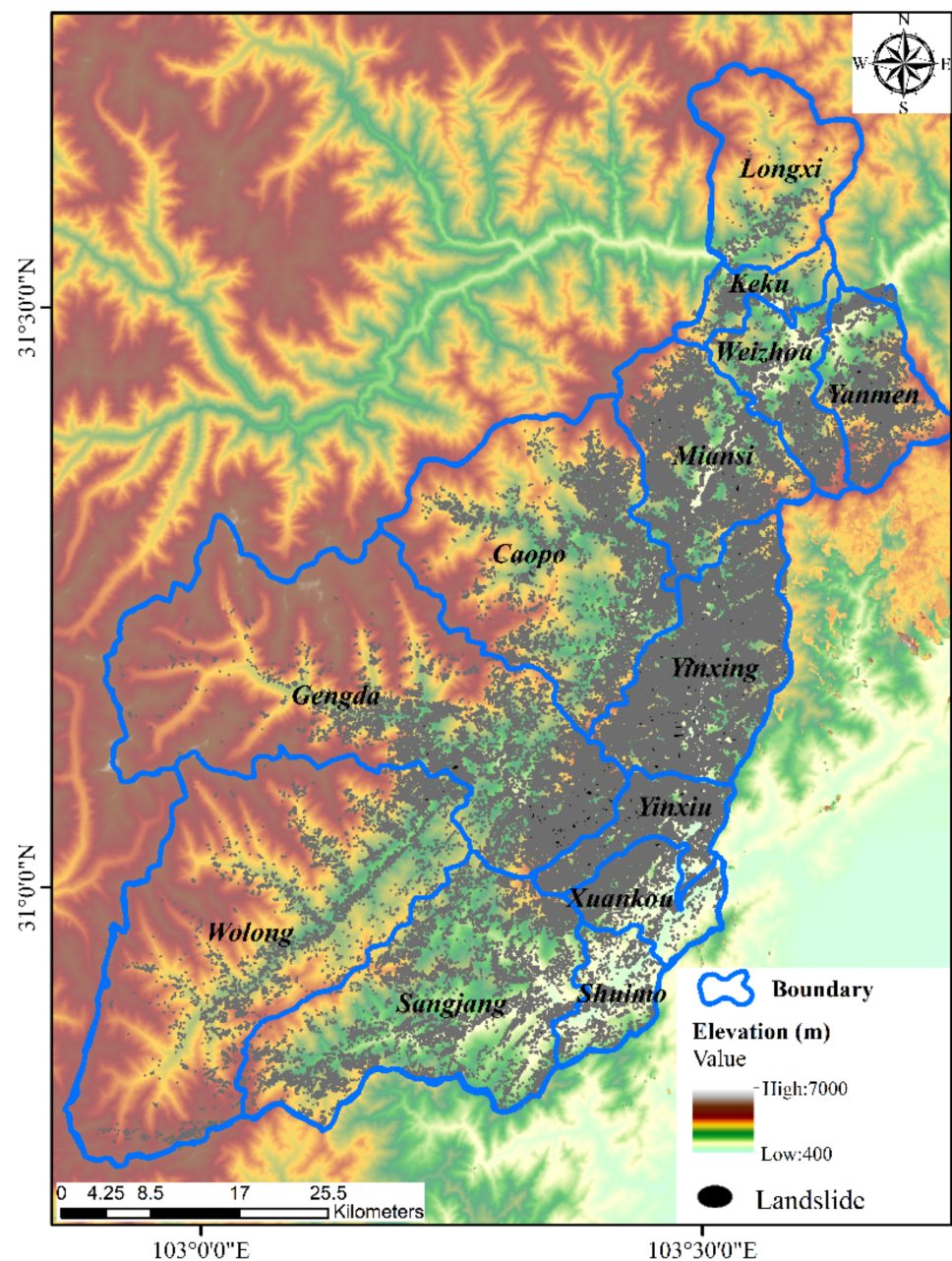


Figure 2. The complete landslide inventory used in this study.

3. Materials and Methods

3.1. Causative Factors

Causative factors control the landslide occurrence and are fundamental to LSM. Due to the complex synergistic inter-action of earthquake and geo-environment, causative factors of seismic landslides vary from one location to another. Until now, there is no unified consensus regarding the selection of seismic landslide causative factors for a specific region [47]. Based on published literature and data availability, this study selects 15 causative factors from 4 data modalities for LSM. Table 2 lists the categories of causative factors. All causative factors were rescaled to spatial resolution of 30 m, and each causative factor is described below.

Table 2. Causative factors categories of landslide.

Causative Factors		Classes	Classification Criteria
Seismology	PGA/(m/s ²)	6	1. <0.2; 2. 0.2~0.4; 3. 0.4~0.6; 4. 0.6~0.8; 5. 0.8~1.0; 6. 1.0~1.2
	Distance to fault/(km)	7	1. 0~3; 2. 3~6; 3. 6~9; 4. 9~12; 5. 12~15; 6. 15~18; 7. >20
	Earthquake intensity	4	1. VIII; 2. IX; 3. X; 4. XI
	Lithology	6	1. sandstone; 2. magmatic rock; 3. phyllite; 4. shale; 5. glutenite; 6. carbonate rock
Terrain	Elevation/(km)	10	1. 0.5~1; 2. 1~1.5; 3. 1.5~2; 4. 2~2.5; 5. 2.5~3; 6. 3~3.5; 7. 3.5~4; 8. 4~4.5; 9. 4.5~5; 10. >5
	Slope (°)	8	1. 0~10; 2. 10~20; 3. 20~30; 4. 30~40; 5. 40~50; 6. 50~60; 7. 60~70; 8. >80
	Aspect	9	1. Flat; 2. N; 3. NE; 4. E; 5. SE; 6. S; 7. SW; 8. W; 9. NW
	Curvature	8	1. 0~3; 2. 3~6; 3. 6~9; 4. 9~12; 5. 12~15; 6. 15~18; 7. 18~21; 8. >21
	Landform	7	1. plain; 2. platform; 3. hill; 4. small undulating mountain; 5. middle undulating mountain; 6. high undulating mountain; 7. very high undulating mountain
Land cover	Land use	5	1. cultivated land; 2. woodland; 3. grassland; 4. water; 5. urban and rural residents
	Soil	6	1. leached; 2. semileached soil; 3. primary soil; 4. alpine soil; 5. ferralsol; 6. rock
	Distance to road/(km)	10	1. 0~0.5; 2. 0.5~1; 3. 1~1.5; 4. 1.5~2; 5. 2~2.5; 6. 2.5~3; 7. 3~3.5; 8. 3.5~4; 9. 4~4.5; 10. >4.5
Hydrological	Distance to river/(km)	11	1. 0~0.5; 2. 0.5~1; 3. 1~1.5; 4. 1.5~2; 5. 2~2.5; 6. 2.5~3; 7. 3~3.5; 8. 3.5~4; 9. 4~4.5; 10. 4.5~5; 11. >5
	Topographic wetness index	9	1. 0.63~2; 2. 2~4; 3. 4~6; 4. 6~8; 5. 8~10; 6. 10~12; 7. 12~14; 8. 14~16; 9. 16~19.45
	Annual average rainfall/(mm)	7	1. <525; 2. 525~625; 3. 625~725; 4. 725~825; 5. 825~925; 6. 925~1025; 7. >1025

3.1.1. Seismology-Related Causative Factors

Peak ground acceleration (PGA): slope instability during earthquakes means that the combined force of the ground acceleration and the gravity of the slope exceeds the adhesion and friction strength of the bedrock in a short time. Normally, the impact of an earthquake on a landslide is measured and quantified by recording the absolute maximum amplitude of the ground acceleration [1]. This study divides PGA into 6 levels with a step of 0.2 g; see Figure 3a.

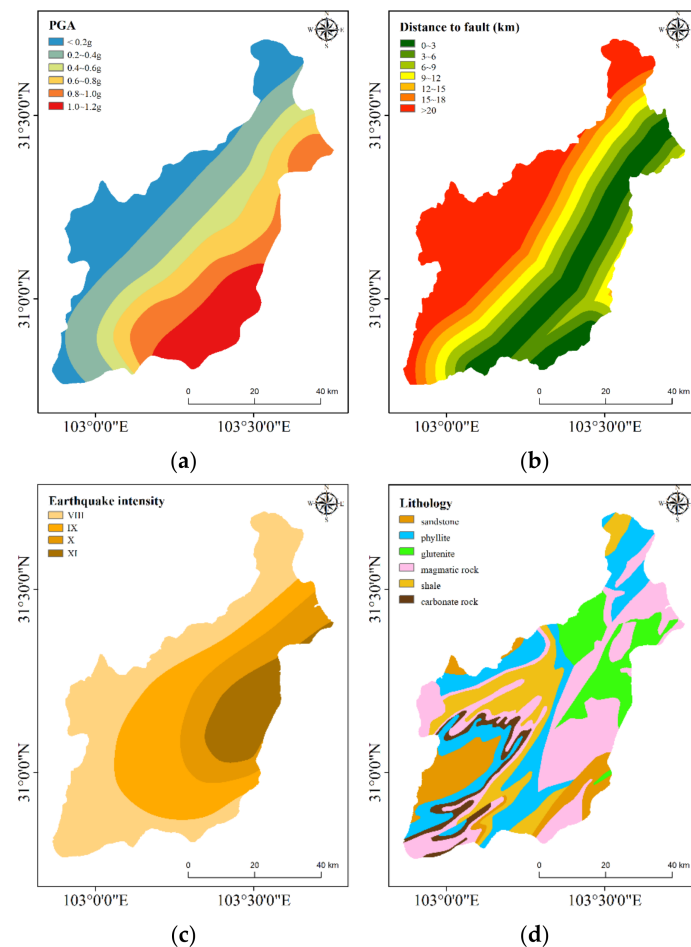


Figure 3. Seismology-related causative factors. (a) PGA; (b) Distance of fault; (c) Earthquake intensity; (d) Lithology.

Distance to fault: fault structure affects the occurrence of geological disasters. Therefore, the distance between the slope and the fault should be considered in slope stability analysis. The internal factor of landslide development and slope instability is that the various weakness planes produced by tectonic movement form different structural combinations with slope or artificial excavation face. This study divides the distance to the fault into 7 levels with a step of 3 km; see Figure 3b.

Seismic intensity: seismic intensity is the main parameter to measure the degree of seismic damage. It refers to the intensity of ground vibration in a certain area within the seismic range. The intensity of seismic activity is not only directly related to slope stability but also has an important influence on the spatial distribution of seismic landslides. The seismic intensity of the study area was divided into 4 levels; see Figure 3c.

Lithology: the lithology is a proxy of the shear strength of the materials constituting slopes, thus directly controlling the state of slope stability. Lithology is the basis to determine slope strength, stress distribution, permeability, and deformation characteristics. There are 6 lithology types in the study area; see Figure 3d.

3.1.2. Terrain-Related Causative Factors

Elevation: elevation is closely related to the occurrence of landslides. Elevation controls terrain slope and surface water catchment capacity. Furthermore, the intensity of human activities varies at different elevations. This study divides the elevation into 10 groups with a step of 0.5 km; see Figure 4a.

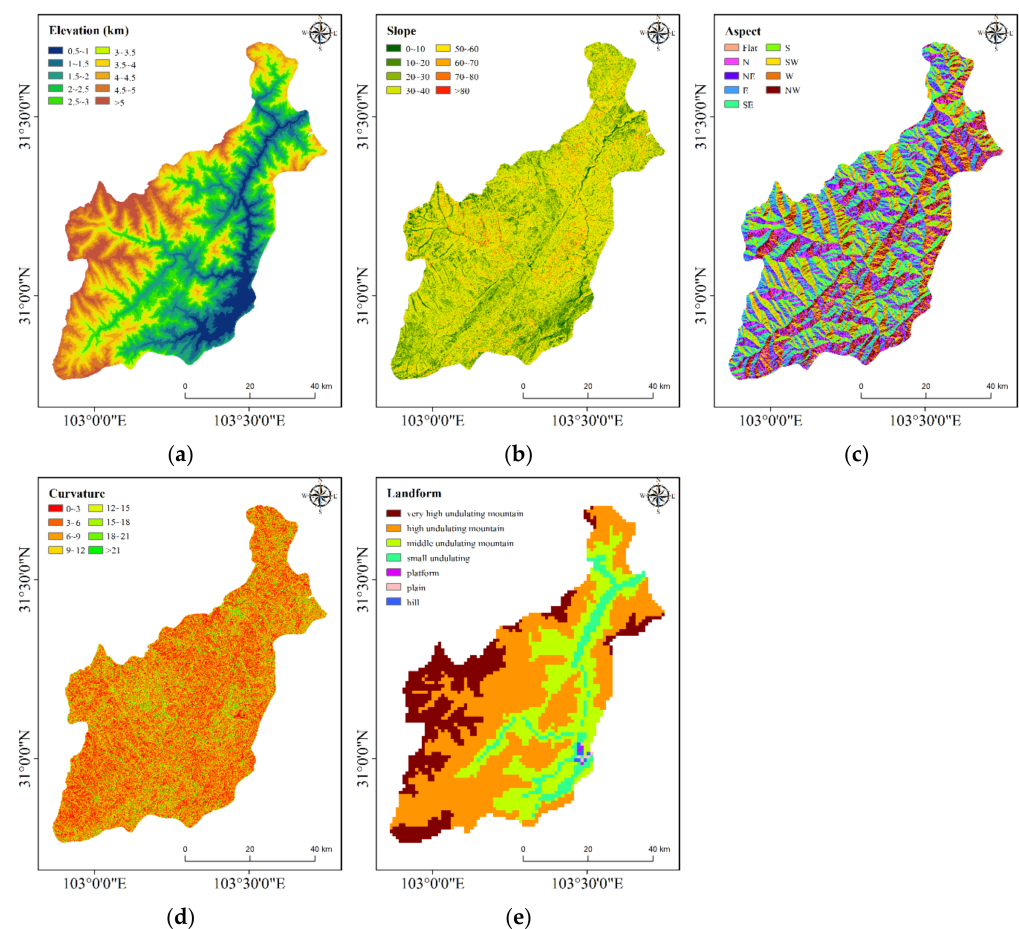


Figure 4. Terrain-related causative factors. (a) Elevation; (b) Slope; (c) Aspect; (d) Carvatu; (e) Landform.

Slope: the slope angle is a very good proxy of the shear stress that promotes instability on slopes. Theoretically, the propensity to landslide occurrence tends to increase with the

increase of slope angle [2]. The slope in the study area was discretized based on a 10° interval and generated 8 categories; see Figure 4b.

Aspect: aspect, referring to the direction of the normal vector of the projected slope on the horizontal plane, is related to soil moisture, surface runoff and vegetation, which indirectly affects landslide development [3]. According to the local terrain, the aspects of the study area were divided into 9 classes at 45° angle intervals, one of which is flat; see Figure 4c.

Profile curvature: profile curvature reflects the sharpness of the slope and represents the ground complexity. Profile curvature affects the movement (acceleration or deceleration) of materials on the slope and plays an indirect role in the transportation and deposition of materials on the slope. This study divides profile curvature into 8 classes according to the classification results; see Figure 4d.

Landform: the distribution of landslides is closely related to geomorphic genesis and topography. Landslides mainly occur in mountainous areas with steep terrain while few landslides occur on the plain. The study area mainly contains 7 different landform types; see Figure 4e.

3.1.3. Land-Cover-Related Causative Factors

Land use: on the one hand, land use reflects anthropogenic actions that can increase landslide activity [6]. On the other hand, the impact of different land use types on seismic landslides reflects the varying degree of vegetation cover on slopes. For instance, landslides are less frequent along slopes with dense and deeply rooted vegetation [7]. According to the data available on China Academy of Sciences, the study area contains 5 different land use types; see Figure 5a.

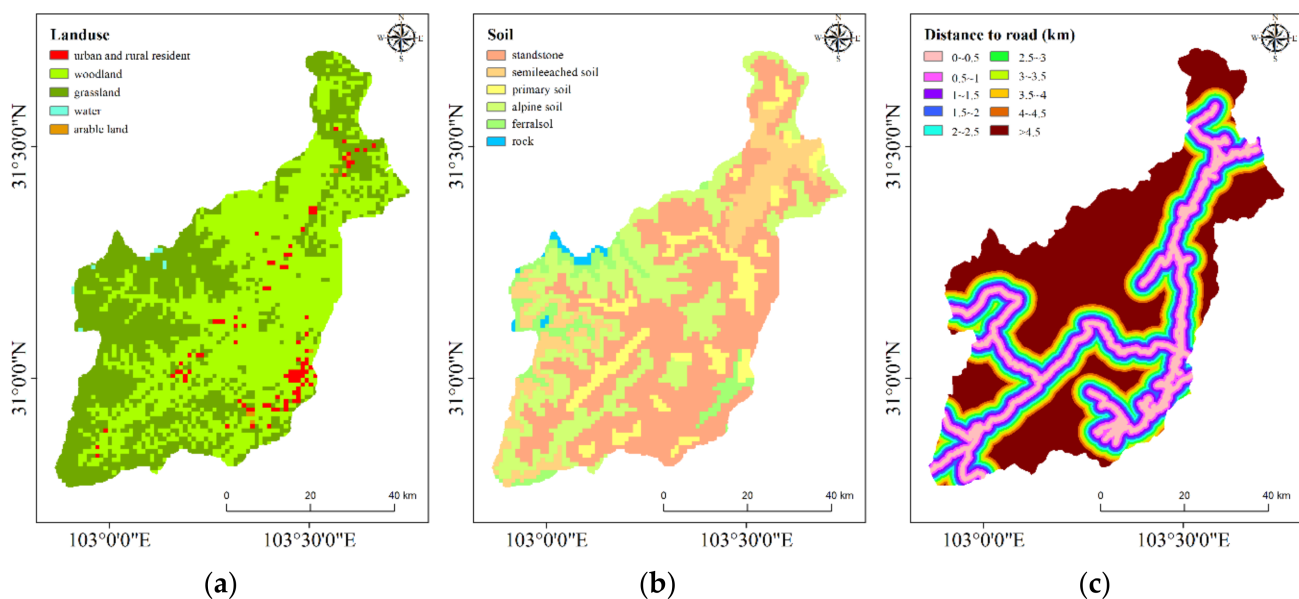


Figure 5. Land-cover-related causative factors. (a) Land use; (b) Soil; (c) Distance to road.

Soil: soils are a critical component of slopes. Different soil types have different mineral composition, densities, and permeability coefficients, which have different effects on the occurrence of landslides. According to the data available from the China Academy of Sciences, the study area contains 6 different land use types; see Figure 5b.

Distance to road: the impact of human activities on landslides is mainly reflected in road construction. In road construction, engineering activities (e.g., such as cutting a slope at the slope top) often change the slope topography, making the slope unstable. The study area was classified into 10 classes with a buffered distance interval of 0.5 km; see Figure 5c.

3.1.4. Hydrological-Related Causative Factors

Distance to river: rivers are key factors river is a key factor in the occurrence probability of landslides. Erosion and cutting by a river can weaken slope stability (Nadim et al. 2006). The distance to a river represents the magnitude of exposure to the erosive action of the river. In accordance with the river distribution of the study area, the buffer zone was established based on a distance of 0.5 km, and the study area was divided into 11 categories; see Figure 6a.

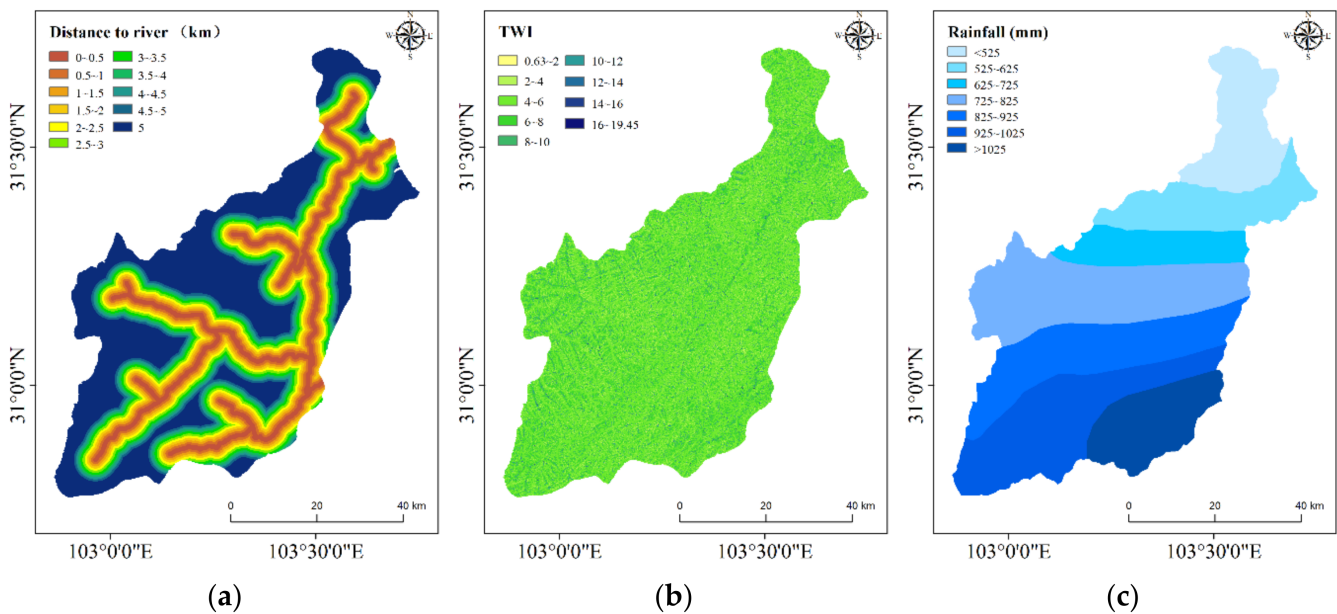


Figure 6. Hydrological-related causative factors. (a) Distance to river; (b) TWI; (c) Rainfall.

Topographic wetness index (TWI): TWI represents the spatial distribution of water in the soil and can describe the topographic influence on the water saturation of the soil. The moisture content in the soil will affect the rock, soil, and vegetation conditions on the slope surface, thereby affecting the landslide occurrence. The study area was divided into 9 sets at 2 intervals according to the range of the TWI; see Figure 6b.

Annual average rainfall: rainfall is an important landslide trigger factor. Surface water seeps into the slope during rainfall, which increases the slope weight and may damage slope stability. The dryness and wetness alternately break the limit equilibrium by increasing pore water pressure through rock cracking, which provides favorable conditions for landslides. This study divides the annual average rainfall into 7 categories; see Figure 6c.

3.2. Methodology

3.2.1. Mapping Unit Selection

A suitable mapping unit is critical for LSM. Five typical mapping units [48] include: (1) grid unit, (2) slope unit (SU), (3) regional unit, (4) homogeneous unit, and (5) sub-watershed unit. Among these five units, grid and slope units are frequently used for LSM due to their simplicity. Though the grid unit is simple and easy to operate, it breaks down the slope integrity and reduces the relationship between landslides and slopes. By contrast to the grid unit, the slope unit boundary is more consistent with the ridge and valley lines, and better reflects the topographic features. Therefore, this study selects SU as the basic mapping unit and then divides the study area into 1351 Sus; see Figure 7.

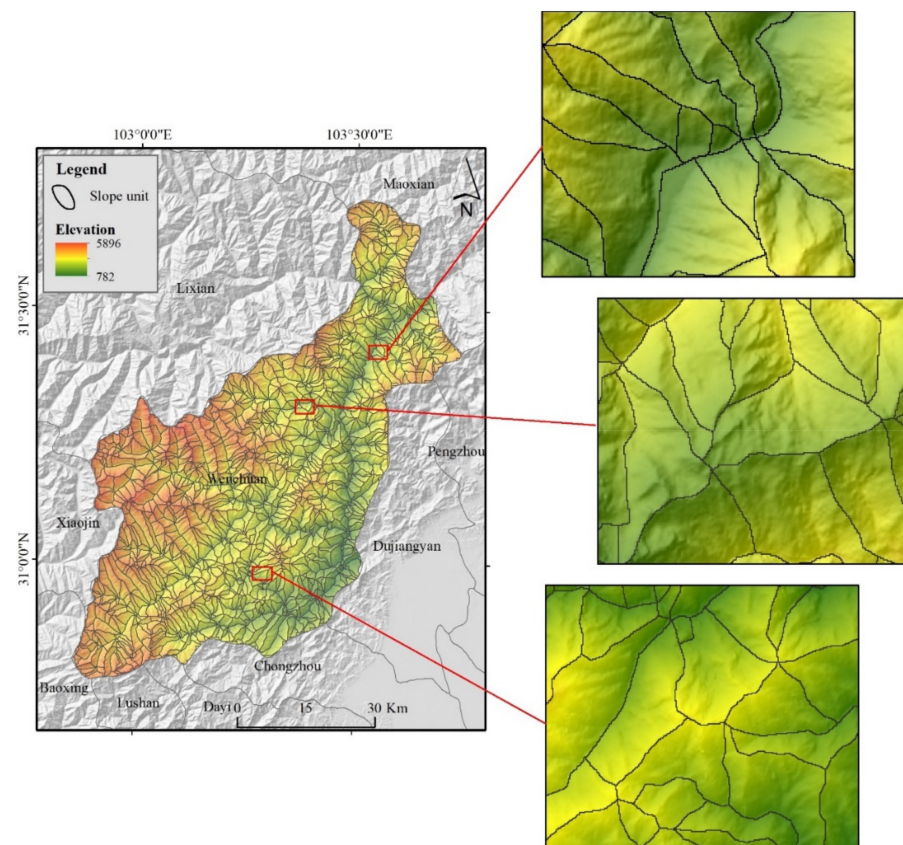


Figure 7. Slope units of the study area.

3.2.2. Independency Analysis

The process of independency analysis is to reduce the influence of the interdependency of causative factors on the performance of the control group. Correlation coefficient analysis is an important method to analyze the independence of causative factors, and the correlation coefficient reflects the correlation between factors. This paper applied GIS statistical analysis tool to establish the correlation coefficient matrix of factors; see Table 3. Based on the threshold algorithm, this study retains those factors with low correlation coefficients.

Table 3. Causative factor categories for landslide.

Factor	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
F1	1														
F2	−0.85	1													
F3	0.04	−0.11	1												
F4	−0.03	−0.06	0	1											
F5	−0.68	0.63	−0.02	0.04	1										
F6	−0.15	0.03	0.12	0.12	0	1									
F7	0.10	−0.10	0.05	0.01	−0.06	−0.09	1								
F8	0	−0.02	0.04	0.06	0.07	0.29	−0.01	1							
F9	−0.48	0.41	0.02	0.03	0.81	0.05	−0.06	0.12	1						
F10	−0.37	0.32	−0.03	0.02	0.48	−0.07	−0.06	0.02	0.40	1					
F11	−0.25	0.25	0	−0.03	0.45	−0.12	0.03	−0.02	0.30	0.21	1				
F12	−0.13	0.08	−0.02	0	0.09	0.10	−0.02	0.07	0.10	−0.08	0.02	1			
F13	−0.30	0.25	0	−0.10	0.67	−0.05	−0.05	0.03	0.65	0.25	0.28	0.12	1		
F14	0.03	0	−0.02	−0.07	−0.07	−0.26	−0.05	−0.22	−0.09	−0.03	0.05	0.02	−0.06	1	
F15	−0.54	0.29	−0.03	0.02	0.16	0.24	−0.03	0.04	0.07	0.04	0.04	0.23	0.14	−0.03	1

F1~F15 represent PGA, distance to fault, earthquake intensity, lithology, elevation, slope, aspect, curvature, landform, land use, soil, distance to road, distance to river, TWI, and rainfall, respectively.

3.2.3. Importance Analysis of Causative Factor

The formation of landslides is the result of the comprehensive action of different causative factors, which have different effects on landslide occurrence. In this paper, the importance of factors is analyzed by mean decrease accuracy from RF according to the complete landslide inventory, and the result is shown in Figure 8. In general, the seismic landslide is a kind of seismic geological disaster directly triggered by the earthquake, and its main causative factors are related parameters of seismology, which in this paper are seismic intensity and distance to fault. The formation of landslides needs certain topographic conditions, that is, there is a steep terrain. Therefore, the terrain slope is another important factor affecting landslides in the Wenchuan earthquake, which is second only to ground motion parameters. Rainfall, especially heavy rainfall, will intensify surface seepage, increase pore water pressure, and weaken rock mass strength. All these effects indicate that rainfall is a factor that cannot be ignored for the formation of landslides. In this paper, rainfall, a hydrological type, is the third major factor after seismology and terrain. Figure 8 also shows that different types of causative factors have different influences on landslides, and a single type of causative factor cannot effectively represent the formation conditions of landslides. Therefore, the comprehensive effects of various types of causative factors should be comprehensively considered.

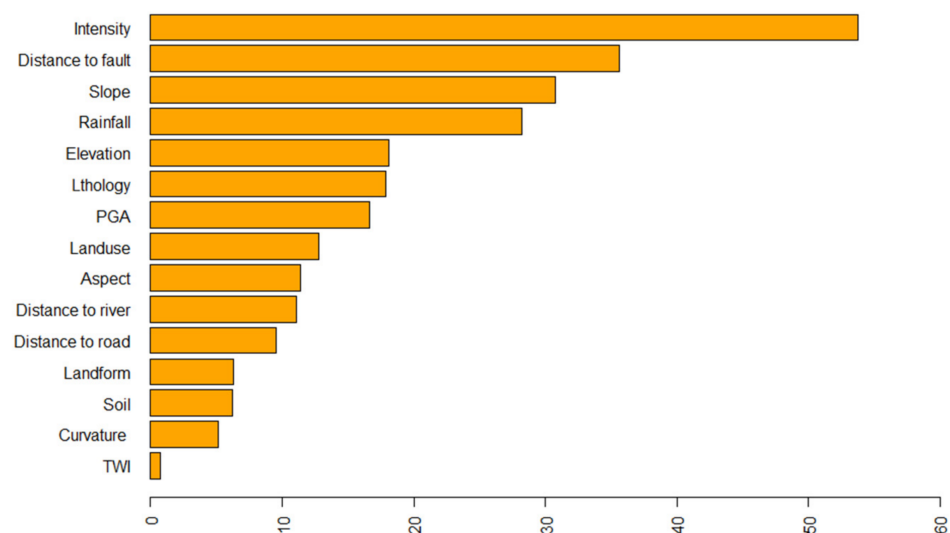


Figure 8. Importance of 15 causative factors.

By full consideration of the results of independency analysis and importance analysis, this study discarded three causative factors, including earthquake intensity, aspect, and soil. The remaining twelve causative factors will be used for the control group.

3.2.4. LSM Based on the Graph Theory and Multi-Modal Classification

This study presents a new landslide susceptibility mapping method based on the graph theory and multi-modal classification. The presented method contains four main steps: (a) feature extraction from each data modality; (b) graph construction based on features of each data modality; (c) nonlinear graph fusion; and (d) landslide susceptibility mapping. Figure 9 illustrates the flowchart of the proposed method.

This study groups 15 causative factors into 4 data modalities: (1) seismic modal, including peak ground acceleration, distance to fault, seismic intensity, and lithology; (2) terrain modal, including elevation, slope, aspect, profile curvature, and landform; (3) land cover modal, including land use, soil, and distance to river; and (4) hydrological modal, including distance to river, topographic wetness index, and annual average rainfall.

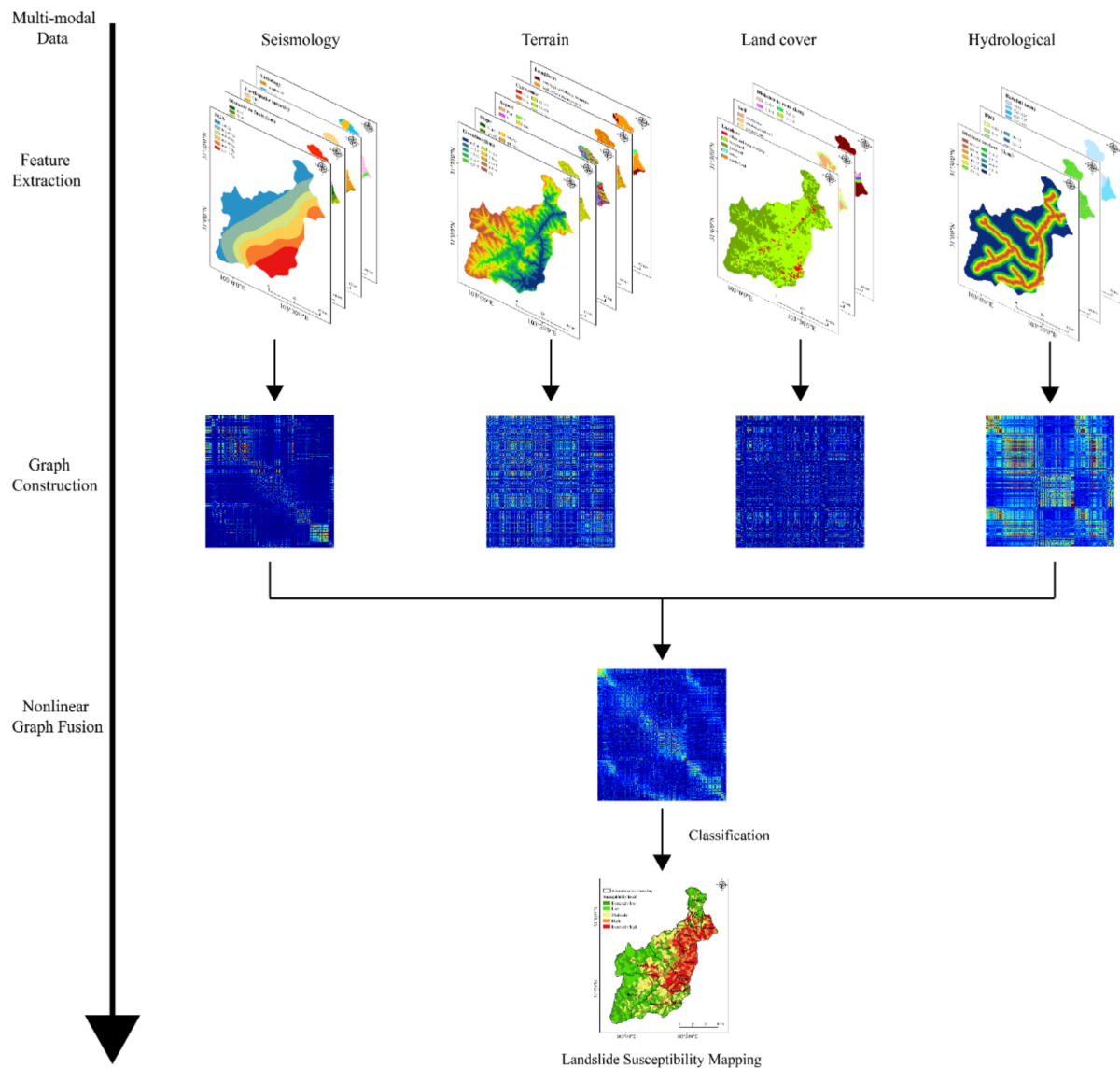


Figure 9. Workflow of LSM based on the graph theory and multi-modal classification.

After the feature extraction, this study applies the graph theory to represent attribute values of slope units (SUs). For each modal data, we build a graph according to its features. Suppose the study area contains n SUs and each SU has l features derived from m modalities. Using features from i -th modality, we construct a graph $G^i = (V^i, E^i)$, $i = 1, \dots, m$ according to model relations between n SUs, where V^i represents n SUs and E^i are weighted by how similar SUs are. W^i is a $n \times n$ similarity matrix that denotes edge weights, $W^i(a, b)$ represents the similarity between SU a and SU b in i -th modality. This study calculates the pairwise Euclidean distance and then converts it to similarity matrix by the exponential kernel. This study applies random forest to calculate the similarity between pairs of subjects. The similarity matrix provides a consistent measure for pairwise similarity of SUs, and thus finds a way to incorporate information from multiple modalities with a unified measure.

The graph construction generates a similarity matrix W^i corresponding to i -th modality. These similarity matrices are then combined to make full use of the complementary information between different modalities. Since different modalities are not necessarily linearly related, a nonlinear graph fusion is utilized for fusion. The similarity matrix W^i is normalized as follows:

$$\tilde{W}^i = \frac{W^i}{D^i}, \quad (1)$$

where $D^i(a, a) = \sum_{b=1}^n W^i(a, b)$ is the diagonal matrix and $\sum_{b=1}^n \tilde{W}^i(a, b) = 1$. The occurrence of self-similarity on the diagonal of W^i renders the normalization algorithm in Equation (1) unstable. To tackle this issue, the study normalizes W^i as:

$$\tilde{w}^i(a, b) = \begin{cases} \frac{w^i(a, b)}{2 \sum_{b=1}^n w^i(a, b)}, & b \neq a \\ 1/2, & b = a \end{cases}, \quad (2)$$

The normalization in Equation (2) not only speeds up the convergence rate, but also guarantees the full rank of the final graph. The sparse matrix $\tilde{S}^i$ of $\tilde{W}^i$ is calculated by k-nearest neighbors (KNN) to measure local affinity as:

$$\tilde{s}^i(a, b) = \begin{cases} \frac{\tilde{w}^i(a, b)}{\sum_{b \in k-NN(a)} \tilde{w}^i} & \text{if } b \in k-NN(a) \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

To keep strong connections and discard weak connections between SUs, the weighted edges between non-neighborhood SUs are set to 0. This is beneficial to improve the robustness to the noise of similarity measures. Given features from m modalities, based on Equations (2) and (3), we obtain m normalized matrices $\tilde{W}^i$ and m sparse matrices $\tilde{S}^i$. After that, a single unified graph is constructed through non-linear iterative cross-diffusion as:

$$\tilde{w}_{t+1} = \tilde{s}^i \times \left(\frac{1}{m-1} \sum_{j=1, j \neq i}^m \tilde{w}_t^j \right) \times (\tilde{s}^i)^T, \quad (4)$$

Let $\tilde{W}_{t=0}^i = \tilde{W}^i$ denote the initial similarity matrix for i -th modality, and $\tilde{S}^i$ is the kernel matrix. The connection information between different modalities is diffused during the iteration. Each iteration will normalize the updated similarity matrix $\tilde{W}_{t+1}^i, i = 1, 2, \dots, m$ as Equation (2). To check the convergence of nonlinear graph fusion, each iteration calculates the relative error $\frac{\|\tilde{W}_t - \tilde{W}_{t-1}\|}{\|\tilde{W}_{t-1}\|}$. If the relative error $\frac{\|\tilde{W}_t - \tilde{W}_{t-1}\|}{\|\tilde{W}_{t-1}\|}$ is smaller than the given threshold, then the iteration will terminate. After T iterations, the final single unified similarity matrix $\tilde{W}_u$ is expressed as:

$$\tilde{W}_u = \frac{1}{m} \sum_{i=1}^m \tilde{W}_T^i, \quad (5)$$

Since the unified similarity matrix $\tilde{W}_u$ keeps all the relationship information between different SUs, entries in $\tilde{W}_u$ will be directly used for LSM. The landslide probability of the study area was calculated based on the aforementioned statistical learning methods (i.e., SVM, LR, KNN, and RF). To intuitively distinguish the hazard degree of the study area, this study utilizes the natural break algorithm to divide the landslide susceptibility into 5 levels: very low susceptibility, low susceptibility, moderate susceptibility, high susceptibility, and very high susceptibility.

4. Results

4.1. Test of the Parameter Influence on the Performance of LSM

The presented method involves two parameters to be set: (1) the number of nearest neighbors k , and (2) the iteration number T for nonlinear graph fusion. To validate the parameter tuning influence on the performance of the presented method, k varies from 100 to 600 with a step of 100. Figure 10a suggests that the influence of k on the area under the curve (AUC) is marginal, which means the effect of k on the performance of LSM is negligible. Thus, this study sets $k = 300$. Figure 10b shows that nonlinear graph fusion converges quickly after very sparse iterations: the cross-diffusion process has converged after 5 iterations. Thus, this study sets $T = 5$.

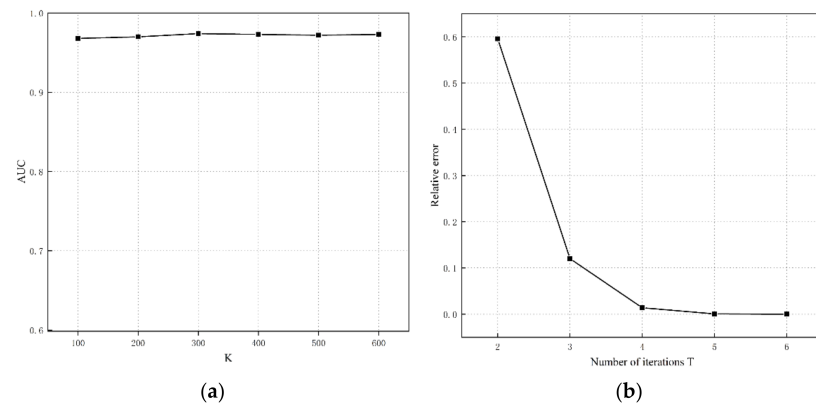


Figure 10. The effect of (a) k and (b) T on the presented method.

4.2. Comparison of Different LSM Methods

This study compares the presented model to three benchmark methods, including support vector machine (SVM), linear regression (LR), KNN, and random forest (RF), and Figure 11 illustrates results of these four LSM methods. The LSM results produced by different methods vary from each other. In the LSM obtained by the MMC model, the landslide risk areas are mainly distributed along the water system, mostly concentrated in the eastern and northeastern part of the study area, and a small amount are distributed in the southeast and western regions; see Figure 12. Among the four LSM methods, the MMC result is the closest to the actual spatial distribution of landslides induced by the Wenchuan earthquake.

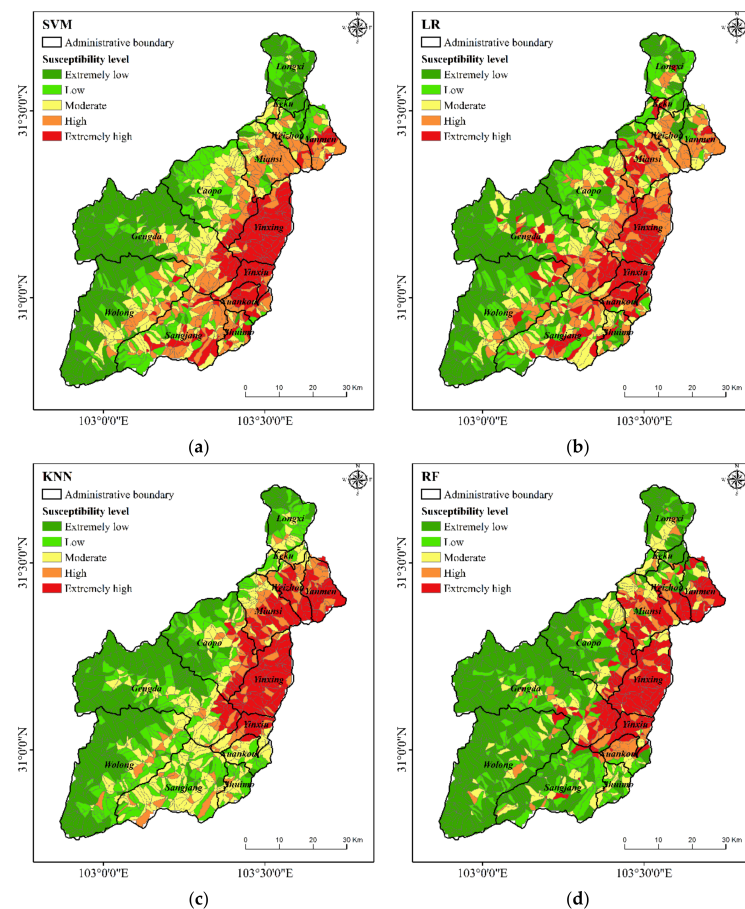


Figure 11. Landslide susceptibility maps produced by different methods: (a) SVM, (b) LR, (c) KNN, and (d) RF.

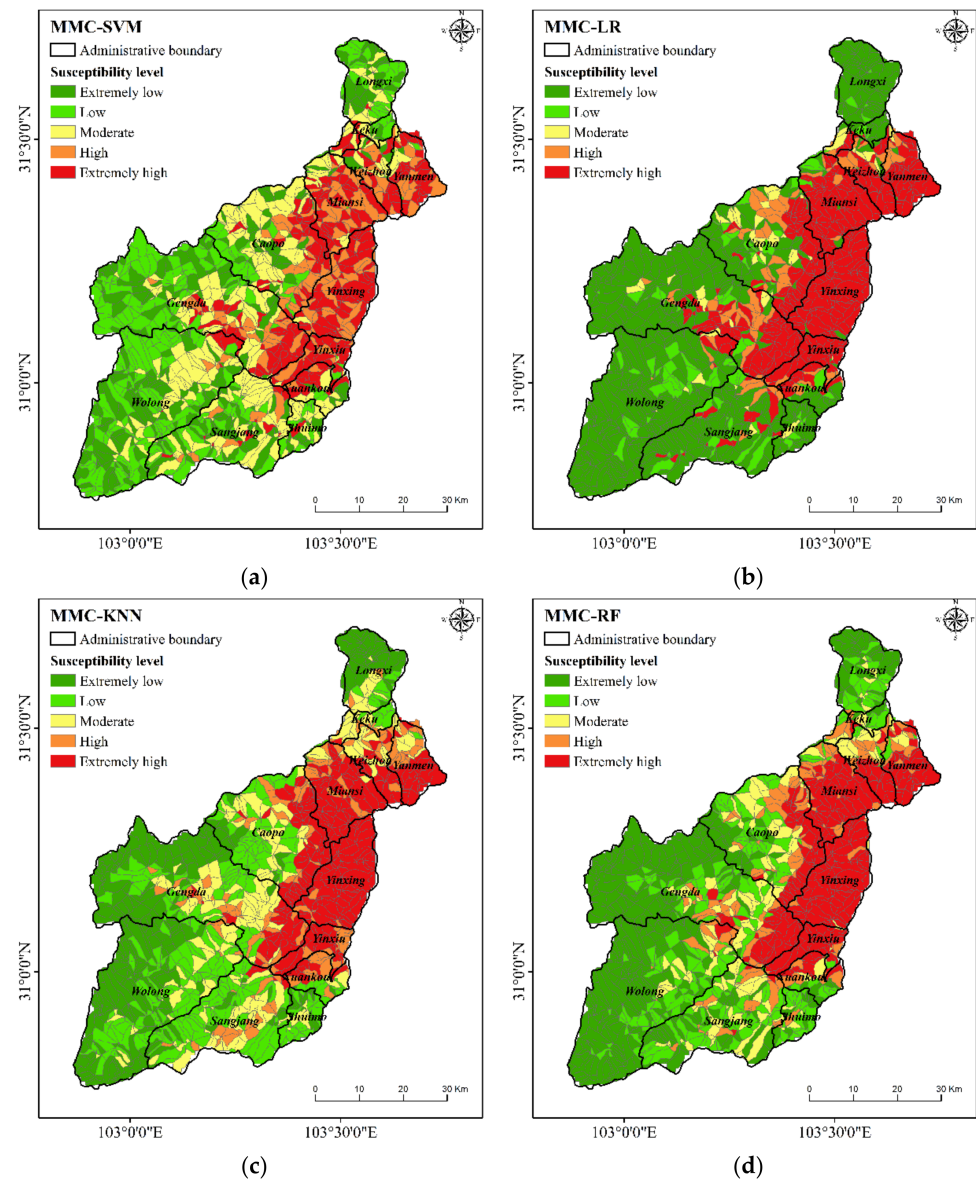


Figure 12. Landslide susceptibility maps produced by different methods considering data modality: (a) MMC-SVM, (b) MMC-LR, (c) MMC-KNN, and (d) MMC-RF.

To quantitatively assess the reliability of the model, we analyzed and verified the reliability of the landslide susceptibility assessment model through the complete landslide inventory data in the study area. Table 4 shows the area of each susceptibility subdivision in the LSM derived from different models, and counts the area of historical landslides in the different subdivisions. In this paper, the area of historical landslides developed in the dangerous zone (containing both very high and high susceptibility areas) is used as the accuracy of model predictions. The statistical results showed that the proportion of landslide area in the hazard zone reached 84% in the MMC modeling, 75% in the KNN model, 72% in the SVM model, and 74% in the LR model. This result shows that the MMC mode has the best performance.

Model evaluation can be used to verify the evaluation performance of LSM. There are two commonly used model evaluation methods, namely, superposition analysis based on complete landslide inventory and comprehensive model evaluation based on independent data sets. The independent data set is mainly established in the preliminary data preparation stage. In this paper, the performance of the model is verified by the above two evaluation methods.

Table 4. Comparison of each susceptibility class for the different modeling.

Algorithm	Susceptibility Level	Area (km ²)	Area Proportion (%)	Landslide Area	Landslide Percentage	Density Percentage
SVM	Very low	1191.12	0.30	14.15	0.04	0.01
	Low	749.00	0.19	32.03	0.09	0.04
	Moderate	746.60	0.18	54.36	0.15	0.07
	High	720.28	0.18	100.21	0.28	0.14
	Very high	631.15	0.16	156.21	0.44	0.25
LR	Very low	1175.66	0.29	12.90	0.04	0.01
	Low	756.84	0.19	27.33	0.08	0.04
	Moderate	717.57	0.18	51.68	0.14	0.07
	High	733.14	0.18	101.20	0.28	0.14
	Very high	656.32	0.16	163.85	0.46	0.25
KNN	Very low	1203.94	0.30	10.83	0.03	0.00
	Low	608.15	0.15	21.47	0.06	0.04
	Moderate	945.85	0.23	64.69	0.18	0.07
	High	594.53	0.15	83.50	0.23	0.14
	Very high	687.27	0.17	176.54	0.49	0.26
RF	Very low	1797.64	0.45	32.35	0.09	0.02
	Low	682.26	0.17	24.43	0.07	0.04
	Moderate	376.85	0.09	29.42	0.08	0.08
	High	434.91	0.11	68.82	0.19	0.16
	Very high	746.49	0.18	201.99	0.57	0.27

In general, landslide inventory mapped after an earthquake is important reference data to verify the model performance. Accurate and reliable LSM should meet the following two requirements: (1) as many landslides as possible should fall into the high susceptibility area, that is, the higher the value, the more accurate the LSM; (2) The proportion of high susceptibility areas in the study area should be as small as possible, that is, the lower the proportion, the more is the reliability of the LSM. The model is considered to have good performance when it meets both requirements. However, this method relies on the classification of susceptibility level, that is, the determination of threshold, and depends more on human subjective factors. Therefore, this method can only represent the reliability of LSM under specific threshold conditions. In contrast, the receiver operating characteristic (ROC) curve is not affected by threshold setting, which is a commonly used method in traditional classification models. By calculating the true positive rate and false positive rate under different thresholds, a curve with true positive rate as X-axis and false positive rate as Y-axis is established. AUC value of ROC is a quantitative indicator of model performance. the higher the AUC value is, the more reliable the model performance is. To further evaluate the prediction accuracy of the model, data on new landslide events after the 2008 Wenchuan earthquake were collected, and the landslide susceptibility level corresponding to the new landslide events was taken as the evaluation index of the model. Generally, the more new landslide events locate in the high susceptibility areas, the more accurate the prediction of the model will be.

Figure 13a presents ROC curves for four LSM methods based on the stack of causative factors. The AUC values for SVM, LR, KNN, and RF are 0.851, 0.874, 0.885, and 0.911, respectively. Among four methods, RF achieves the highest AUC value (0.911) while SVM has the lowest AUC value (0.851). This indicates that LSM results varied in their accuracies depending on the machine learning methods used, and thus the selection of a right machine learning method is critical to obtain a satisfactory LSM for a specific area. Figure 13b shows that AUC values for MCC-SVM, MCC-LR, MCC-KNN, and MCC-RF are 0.985, 0.985, 0.930, and 0.975, respectively. Among four MCC methods, MCC-SVM and MCC-LR produce the highest AUC value (0.985) while MCC-KNN achieves the lowest AUC value (0.930). Overall, all MCC based methods improve the LSM performance compared to the same methods based on the stack feature set. The improvement of AUC values ranges from 4.5% to 13.4%. This suggests that data modality has a significant effect on the LSM performance. Fusion of complementary information from different modalities greatly improves LSM

accuracy compared to using the stack feature set. Thus, the presented method provides a simple and straightforward means to combine complementary information from multiple data modalities, which in turn boosts LSM performance. This verification indicates that the data modality is both statistically viable and provides an improved LSM performance versus stack feature set based models.

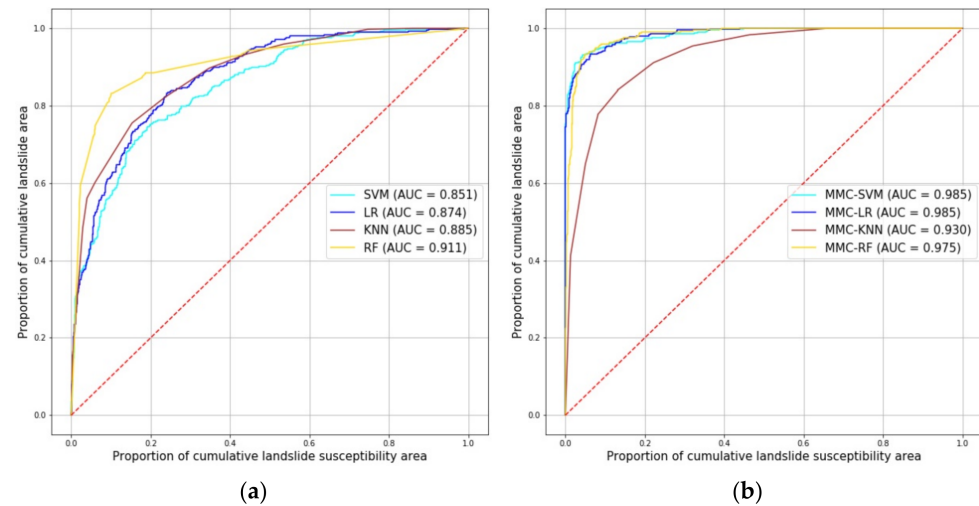


Figure 13. ROC curves of different LSM models: (a) without consideration of data modality, and (b) with the consideration of data modality.

Since LSM is mainly used to evaluate the possibility of subsequent landslides, a new landslide event is also a valuable reference for verifying the performance of LSM generated. Hence, this study collected 10 new landslides that occurred after the Wenchuan earthquake through searching the Internet. Geographic locations of these landslides were determined by the cross-analysis of the news and time series of satellite images from Google Earth, as shown in Table 4. The statistics on landslide susceptibility corresponding to the new landslides and analysis of the reliability of the landslide susceptibility results are presented in Table 5. For the stack feature set, new landslide occurrence in the non-risk areas (i.e., low, and extremely low susceptibility areas) for SVM, LR, KNN and RF was 1, 1, 1, and 3, respectively. For MMC-SVM, MMC-LR, MMC-KNN and MMC-RF, new landslide occurrence in the non-risk areas was 1, 1, 1, and 1, respectively. This suggests that statistical methods by consideration of data modality achieve the same or even higher performance compared to that without consideration of data modality. Thus, LSM methods with the consideration of data modality are also beneficial for assessing the possibility of future landslides.

Table 5. LSM validation using new landslide events.

No.	Time	Latitude	Longitude	Landslide Susceptibility Level							
				SVM	LR	KNN	RF	MMC-SVM	MMC-LR	MMC-KNN	MMC-RF
1	2010/05/30	103.644599	31.496066	M	H	H	M	H	H	M	M
2	2018/07/20	103.574819	31.475016	M	H	EH	L	H	M	EH	H
3	2018/04/02	103.545836	31.678040	EL	EL	EL	EL	L	EL	EL	EL
4	2013/07/22	103.443362	31.296196	M	H	EH	EH	EH	EH	EH	EH
5	2009/07/25	103.488871	31.219105	EH	EH	EH	EH	EH	EH	EH	EH
6	2010/06/12	103.411845	31.220022	H	H	H	EH	H	EH	EH	EH
7	2011/07/03	103.502089	31.120935	EH	H	H	EH	H	EH	EH	EH
9	2011/07/03	103.500687	31.067901	EH	EH	EH	H	EH	EH	H	EH
10	2011/07/03	103.503880	31.174279	EH	EH	EH	EH	H	EH	EH	EH
11	2019/08/20	103.300290	31.120226	M	H	H	L	M	H	M	M

EL, L, M, H, EH represent Extremely low, Low, Moderate, High, Extremely high.

5. Conclusions

Landslide susceptibility mapping (LSM) is imperative to landslide-related disaster reduction and prevention and sustainable uses of slopes. Machine learning algorithms have been widely applied to LSM and many scholars designed various methods to boost LSM performance by combining results of multiple machine learning algorithms. Compared to fusion at the decision level, information fusion for LSM at the data modality level has been seldom reported. To address this issue, this paper exploits data modality to boost the LSM performance. Experimental results suggested that the data modality remarkably affects LSM accuracy, and the presented LSM model effectively integrates the complementary information of different data modalities and achieves more satisfactory results than mainstream LSM models without consideration of data modalities. This study provides new insight on the data modality applications to LSM. This study finds that the performance of machine learning algorithms (e.g., SVM, LR, KNN and RF used in this study) varies in accordance with topographic variables. Such results agree with those of state-of-the-art studies, and suggest that the patterns in causative factors for landslides are highly complex and variable for the different facets of causative factor attributes (e.g., spatial scale). This highlights the importance of considering data modality at the data and modeling level for LSM, as done by the proposed approach in this study. As shown in the experimental results, the proposed MMC approach can improve the accuracy of LSM when used with any of SVM, LR, KNN and RF. Alternatively, assembling the results from different machine learning algorithms at the decision level may further improve the reliability of LSM.

The generation of training data in this study relies on a complete landslide inventory which is beneficial for generating a good amount of training data to get reliable LSM. Extensive labelled training samples, however, are often a constraint as a complete landslide inventory after an earthquake is rarely available for a specific area in a very short time. In such circumstances, the selection of suitable machine learning algorithms is critical. If the incomplete landslide inventory cannot generate sufficient training samples or the number of causative factors is larger than that of training samples, statistical learning algorithms with high bias and low variance (e.g., linear regression, Naive Bayes, and linear SVM) would be recommended. Otherwise, low bias and high variance algorithms (e.g., KNN, decision trees, and kernel SVM) can be considered. Another direction is to develop suitable algorithms (e.g., pseudo training sample generation [49], mixed-effects models [50]) to reduce the influence of the inventory incompleteness on the statistical learning based LSM. The third direction is to exploit the potential of crowdsourcing data (e.g., website, social media) to improve the completeness and the effectiveness of the landslide inventory.

This paper focuses on the multi-modal analysis of causative factors, but the high quality of input causative factor data required is also an important factor that cannot be ignored to produce satisfactory performance of LSM. The rapid development of new sensors undoubtedly provides new sources for high quality data on causative factors (Dou, et al., 2019). However, the higher resolution of the causative data does not necessarily lead to the high accuracy of LSM. The appropriate resolution needs to be further discussed. In addition, the current research directly uses causative factors, which lack the necessary inspection on quality. Therefore, the quality of causative factors requires in-depth analysis to control the influence of its inherent bias error on LSM.

The dataset may have a large number of features that may not all be relevant and significant. For a certain data modality, the feature number may be larger compared to training data size. A large number of features may bog down some learning algorithms, making training time unfeasibly long. Therefore, the selection of causative factors should be used to reduce dimensionality and select important features. In our work, only four modal data are used for landslide classification, and there may have other modal data not used as classification data that may provide additional complementary information to potentially improve the classification performance, such as human engineering activities data. It should be noted that some feature data of one modal may have similar effects on landslide generation in certain areas, which may cause confounding effects when classifying based on

such data. It would be interesting to analyze in depth the impact of the selection of specific characteristic factors in modal data on the methodology used in this paper. In addition, some features extracted from different modal data are not independent from each other, and it makes sense to merge the correlation of features into the proposed classification framework to help classification. In most traditional landslide susceptibility evaluation methods, only feature-level data are simply concatenated for landslide classification. Some studies [51] have also explored the influence of the number of selected feature parameters on the accuracy of landslide evaluation results, and found that not including all features in the classification evaluation may give better result

Author Contributions: Conceptualization, Z.M., R.P. and W.W.; methodology, Z.M., R.P. and Q.L. (Qirong Li); software, R.P.; validation, M.P., S.C. and A.Z.; formal analysis, K.L.; investigation, Q.L. (Qinqin Liu); resources, C.H.; data curation, R.P.; writing—original draft preparation, Z.M.; writing—review and editing, Z.M.; visualization, R.P.; supervision, Z.M.; project administration, Z.M.; funding acquisition, Z.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grant no. 42171084, 4217011932), Innovation-Driven Project of Central South University (grant no. 2020CX036), and the Research Innovation Project for Graduate Students of Central South University (project no. 2021zzts0846).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive and valuable comments that significantly contributed to improving this paper.

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

References

1. Panahi, M.; Gayen, A.; Pourghasemi, H.R.; Rezaie, F.; Lee, S. Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-fuzzy inference system (ANFIS) with various metaheuristic algorithms. *Sci. Total Environ.* **2020**, *741*, 139937. [\[CrossRef\]](#)
2. Li, Z.; Shi, W.; Myint, S.W.; Lu, P.; Wang, Q. Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method. *Remote Sens. Environ.* **2016**, *175*, 215–230. [\[CrossRef\]](#)
3. Ma, Z.; Qin, S.; Cao, C.; Lv, J.; Li, G.; Qiao, S.; Hu, X. The Influence of Different Knowledge-Driven Methods on Landslide Susceptibility Mapping: A Case Study in the Changbai Mountain Area, Northeast China. *Entropy* **2019**, *21*, 372. [\[CrossRef\]](#)
4. Franco, A.; Schneider-Muntau, B.; Roberts, N.J.; Clague, J.J.; Gems, B. Geometry-Based Preliminary Quantification of Landslide-Induced Impulse Wave Attenuation in Mountain Lakes. *Appl. Sci.* **2021**, *11*, 1614. [\[CrossRef\]](#)
5. Lin, W.T. Earthquake-induced landslide hazard monitoring and assessment using SOM and PROMETHEE techniques: A case study at the Chiufenershan area in Central Taiwan. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 995–1012. [\[CrossRef\]](#)
6. Baharvand, S.; Rahnamarad, J.; Soori, S.; Saadatkhah, N. Landslide susceptibility zoning in a catchment of Zagros Mountains using fuzzy logic and GIS. *Environ. Earth Sci.* **2020**, *79*, 204. [\[CrossRef\]](#)
7. Hong, H.; Pradhan, B.; Jebur, M.N.; Bui, D.T.; Xu, C.; Akgun, A. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environ. Earth Sci.* **2015**, *75*, 40. [\[CrossRef\]](#)
8. Nadim, F.; Kjekstad, O.; Peduzzi, P.; Herold, C.; Jaedicke, C. Global landslide and avalanche hotspots. *Landslides* **2006**, *3*, 159–173. [\[CrossRef\]](#)
9. Xiao, T.; Segoni, S.; Chen, L.; Yin, K.; Casagli, N. A step beyond landslide susceptibility maps: A simple method to investigate and explain the different outcomes obtained by different approaches. *Landslides* **2020**, *17*, 627–640. [\[CrossRef\]](#)
10. Papathanassiou, G.; Valkaniotis, S.; Ganas, A.; Pavlides, S. GIS-based statistical analysis of the spatial distribution of earthquake-induced landslides in the island of Lefkada, Ionian Islands, Greece. *Landslides* **2013**, *10*, 771–783. [\[CrossRef\]](#)
11. Dekavalla, M.; Argialas, D. Evaluation of a spatially adaptive approach for land surface classification from digital elevation models. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1978–2000. [\[CrossRef\]](#)
12. Qin, C.-Z.; Bao, L.-L.; Zhu, A.X.; Wang, R.-X.; Hu, X.-M. Uncertainty due to DEM error in landslide susceptibility mapping. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1364–1380. [\[CrossRef\]](#)

13. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378. [\[CrossRef\]](#)
14. Keefer, D.K.; Larsen, M.C. Assessing Landslide Hazards. *Science* **2007**, *316*, 1136–1138. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Yeon, Y.-K.; Han, J.-G.; Ryu, K.H. Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Eng. Geol.* **2010**, *116*, 274–283. [\[CrossRef\]](#)
16. Zêzere, J.L.; Pereira, S.; Melo, R.; Oliveira, S.C.; Garcia, R.A.C. Mapping landslide susceptibility using data-driven methods. *Sci. Total Environ.* **2017**, *589*, 250–267. [\[CrossRef\]](#)
17. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [\[CrossRef\]](#)
18. Liu, Y.; Deng, Z.; Wang, X. The Effects of Rainfall, Soil Type and Slope on the Processes and Mechanisms of Rainfall-Induced Shallow Landslides. *Appl. Sci.* **2021**, *11*, 1652. [\[CrossRef\]](#)
19. Mergili, M.; Marchesini, I.; Rossi, M.; Guzzetti, F.; Fellin, W. Spatially distributed three-dimensional slope stability modelling in a raster GIS. *Geomorphology* **2014**, *206*, 178–195. [\[CrossRef\]](#)
20. Kukemilks, K.; Wagner, J.-F.; Saks, T.; Brunner, P. Physically based hydrogeological and slope stability modeling of the Turaida castle mound. *Landslides* **2018**, *15*, 2267–2278. [\[CrossRef\]](#)
21. Yang, Z.-H.; Lan, H.-X.; Gao, X.; Li, L.-P.; Meng, Y.-S.; Wu, Y.-M. Urgent landslide susceptibility assessment in the 2013 Lushan earthquake-impacted area, Sichuan Province, China. *Nat. Hazards* **2015**, *75*, 2467–2487. [\[CrossRef\]](#)
22. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [\[CrossRef\]](#)
23. Che, V.B.; Kervyn, M.; Suh, C.E.; Fontijn, K.; Ernst, G.G.J.; del Marmol, M.A.; Trefois, P.; Jacobs, P. Landslide susceptibility assessment in Limbe (SW Cameroon): A field calibrated seed cell and information value method. *CATENA* **2012**, *92*, 83–98. [\[CrossRef\]](#)
24. Ilia, I.; Tsangaratos, P. Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map. *Landslides* **2016**, *13*, 379–397. [\[CrossRef\]](#)
25. Mondal, S.; Mandal, S. Landslide susceptibility mapping of Darjeeling Himalaya, India using index of entropy (IOE) model. *Appl. Geomat.* **2019**, *11*, 129–146. [\[CrossRef\]](#)
26. Abbaszadeh Shahri, A.; Spross, J.; Johansson, F.; Larsson, S. Landslide susceptibility hazard map in southwest Sweden using artificial neural network. *CATENA* **2019**, *183*, 104225. [\[CrossRef\]](#)
27. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831. [\[CrossRef\]](#)
28. Shi, D.; Yang, Q.; Zhu, Q.; Jupp, D.L.B.; Long, Y. Uncertainties and errors in algorithms for elevation gradients. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 296–320. [\[CrossRef\]](#)
29. Miao, S.; Zhu, Q.; Zhang, B.; Ding, Y.; Zhang, J.; Zhu, J.; Zhou, Y.; He, H.; Yang, W.; Chen, L. Knowledge-guided consistent correlation analysis of multimode landslide monitoring data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2255–2271. [\[CrossRef\]](#)
30. Huabin, W.; Gangjun, L.; Weiya, X.; Gonghui, W. GIS-based landslide hazard assessment: An overview. *Prog. Phys. Geogr. Earth Environ.* **2005**, *29*, 548–567. [\[CrossRef\]](#)
31. Pourghasemi, H.R.; Kornejady, A.; Kerle, N.; Shabani, F. Investigating the effects of different landslide positioning techniques, landslide partitioning approaches, and presence-absence balances on landslide susceptibility mapping. *CATENA* **2020**, *187*, 104364. [\[CrossRef\]](#)
32. Wang, L.-J.; Sawada, K.; Moriguchi, S. Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. *Comput. Geosci.* **2013**, *57*, 81–92. [\[CrossRef\]](#)
33. Lee, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* **2005**, *26*, 1477–1491. [\[CrossRef\]](#)
34. Zêzere, J.L.; Reis, E.; Garcia, R.; Oliveira, S.; Rodrigues, M.L.; Vieira, G.; Ferreira, A.B. Integration of spatial and temporal data for the definition of different landslide hazard scenarios in the area north of Lisbon (Portugal). *Nat. Hazards Earth Syst. Sci.* **2004**, *4*, 133–146. [\[CrossRef\]](#)
35. Lee, S.; Ryu, J.-H.; Won, J.-S.; Park, H.-J. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Eng. Geol.* **2004**, *71*, 289–302. [\[CrossRef\]](#)
36. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. 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* **2013**, *65*, 135–165. [\[CrossRef\]](#)
37. Tian, Y.; Xu, C.; Hong, H.; Zhou, Q.; Wang, D. Mapping earthquake-triggered landslide susceptibility by use of artificial neural network (ANN) models: An example of the 2013 Moxian (China) Mw 5.9 event. *Geomat. Nat. Hazards Risk* **2019**, *10*, 1–25. [\[CrossRef\]](#)
38. Li, X.Z.; Kong, J.M. Application of GA-SVM method with parameter optimization for landslide development prediction. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 525–533. [\[CrossRef\]](#)

39. Jaafari, A.; Panahi, M.; Pham, B.T.; Shahabi, H.; Bui, D.T.; Rezaie, F.; Lee, S. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *CATENA* **2019**, *175*, 430–445. [[CrossRef](#)]
40. Chen, W.; Panahi, M.; Pourghasemi, H.R. Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *CATENA* **2017**, *157*, 310–324. [[CrossRef](#)]
41. Wang, Y.; Feng, L.; Li, S.; Ren, F.; Du, Q. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. *CATENA* **2020**, *188*, 104425. [[CrossRef](#)]
42. Lagomarsino, D.; Tofani, V.; Segoni, S.; Catani, F.; Casagli, N. A Tool for Classification and Regression Using Random Forest Methodology: Applications to Landslide Susceptibility Mapping and Soil Thickness Modeling. *Environ. Model. Assess.* **2017**, *22*, 201–214. [[CrossRef](#)]
43. Shirzadi, A.; Bui, D.T.; Pham, B.T.; Solaimani, K.; Chapi, K.; Kavian, A.; Shahabi, H.; Revhaug, I. Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ. Earth Sci.* **2017**, *76*, 60. [[CrossRef](#)]
44. Goetz, J.N.; Guthrie, R.H.; Brenning, A. Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology* **2011**, *129*, 376–386. [[CrossRef](#)]
45. Kanungo, D.P.; Arora, M.K.; Sarkar, S.; Gupta, R.P. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Eng. Geol.* **2006**, *85*, 347–366. [[CrossRef](#)]
46. Xu, C.; Xu, X.; Yao, X.; Dai, F. Three (nearly) complete inventories of landslides triggered by the May 12, 2008 Wenchuan Mw 7.9 earthquake of China and their spatial distribution statistical analysis. *Landslides* **2014**, *11*, 441–461. [[CrossRef](#)]
47. Bornaetxea, T.; Rossi, M.; Marchesini, I.; Alvioli, M. Effective surveyed area and its role in statistical landslide susceptibility assessments. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2455–2469. [[CrossRef](#)]
48. Van Den Eeckhaut, M.; Reichenbach, P.; Guzzetti, F.; Rossi, M.; Poesen, J. Combined landslide inventory and susceptibility assessment based on different mapping units: An example from the Flemish Ardennes, Belgium. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 507–521. [[CrossRef](#)]
49. Chen, S.; Miao, Z.; Wu, L.; He, Y. Application of an Incomplete Landslide Inventory and One Class Classifier to Earthquake-Induced Landslide Susceptibility Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1649–1660. [[CrossRef](#)]
50. Steger, S.; Brenning, A.; Bell, R.; Glade, T. The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides* **2017**, *14*, 1767–1781. [[CrossRef](#)]
51. Pradhan, B. 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.* **2013**, *51*, 350–365. [[CrossRef](#)]