A One-Class-Classifier-Based Negative Data Generation Method for Rapid Earthquake-Induced Landslide Susceptibility Mapping

Shuai Chen[1], Zelang Miao[1]*, Lixin Wu[1], Anshu Zhang[2], Qirong Li[1] and Yueguang He[3]

[1]School of Geoscience and Info-Physics, Central South University, Changsha, China, [2]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, [3]School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha, China

Machine learning with extensively labeled training samples (e.g., positive and negative data) has received much attention in terms of addressing earthquake-induced landslide susceptibility mapping (LSM). However, the extensive amount of labeled training data required by machine learning, particularly the precise negative data (i.e., non-landslide area), cannot be easily and efficiently collected. To address this issue, this study presents a one-class-classifier-based negative data generation method for rapid earthquake-induced LSM. First, an incomplete landslide inventory (i.e., positive data) was produced with the aid of change detection using before-and-after satellite images and the Geographic Information System (GIS). Second, a one-class classifier was utilized to compute the probability of landslide occurrence based on the incomplete landslide inventory followed by the negative data generation from the low landslide susceptibility areas. Third, the positive data as well as the generated negative data (i.e., non-landslide) were compounded to train a traditional binary classifier to produce the final LSM. Experimental results suggest that the proposed method is capable of achieving a result that is comparable to methods using the complete landslide inventory, and it displays good correspondence with recent landslide events, making it a suitable method for rapid earthquake-induced LSM. The findings in this study would be useful in regional disaster planning and risk reduction.

Keywords: earthquake-induced landslide, landslide susceptibility mapping, one class classifier, incomplete landslide inventory, negative data

1 INTRODUCTION

Many mountainous areas in the world, such as southwest China, are prone to seismic events and, consequently, landslides (Fan et al., 2018; Cao et al., 2019). Seismic landslides are widely distributed and of large scales, and the damage they cause is often great. In addition, large earthquakes may change the local geological structure and create unstable slopes that may slide in the future (Huang and Li, 2014; Yunus et al., 2020). For instance, after the Wenchuan Mw 7.9 earthquake and Jiuzhaigou Mw 6.5 earthquake, many landslides occurred and brought about extensive damages to southwest China (Gorum et al., 2011; Fan et al., 2018). Therefore, timely monitoring of landslide susceptibility after the earthquake is very critical for post-earthquake rehabilitation and reconstruction as well as early disaster monitoring and prevention (Guzzetti et al., 2006).

Multiple methods have been designed for landslide susceptibility mapping (LSM). Field surveys are utilized with great frequency due to their high accuracy and robustness. This method, however, depends heavily on expert experience and is time-consuming and labor-intensive, so it is not suitable for regional/global LSM (Keefer, 2002). Furthermore, it is challenging for field surveys to map landslide susceptibility at full spatial coverage, particularly for some inaccessible areas. In recent years, automated and/or semi-automated LSM has been developed based on remote sensing technology and the Geographic Information System (GIS), and the methods include the heuristic method (Ruff and Czurda, 2008), deterministic method (Jibson, 1993; Jibson et al., 2000; Tsai et al., 2019), and machine learning method (Reichenbach et al., 2018). By use of the heuristic method, investigators rank and weigh the causative factors based on their importance in causing landslides. The result depends on the investigator's understanding of the real causative factors, which may lead to subjective LSM results (Mandal and Mandal, 2018). The deterministic method employs the simplified and physically-based landslide modeling schemes to analyze the stability using simple limit equilibrium models (e.g., the Newmark model) or more sophisticated approaches. The accuracy of this method depends on the reliability of the geotechnical and hydrological input data (Dreyfus et al., 2013). The machine learning method comprehensively considers the correlation between landslide distribution and regional geological environment and thus obtains reliable landslide susceptibility map (Reichenbach et al., 2018). It also has the widest application in LSM among the three methods (Pourghasemi et al., 2018).

A complete and accurate coseismic landslide inventory map (LIM) is an important prerequisite for landslide susceptibility analysis. An ideal coseismic LIM would cover the entire earthquake-affected area, accurately locate all the landslides triggered by the earthquake, and depict the true shapes of landslides in the form of a vector polygon (Harp et al., 2011). The landslide distribution information from such inventories can then be used for the seismic landslide susceptibility analysis and other quantitative analyses (Tanyaş et al., 2017). However, it is difficult or even impossible to obtain a detailed and complete landslide inventory after an earthquake. On the one hand, due to the cloudy and rainy weather, the available optical images are limited and do not have sufficient spatial coverage for the whole earthquake area. On the other hand, a large earthquake often induces many landslides, which are widely distributed and often deeply seated. Existing technology thus faces challenges to mapping all these landslides quickly in the short time following the earthquake.

Besides the detailed LIM, massive labeled training data are also important for the machine learning methods. Most statistical learning methods for predicting landslide-prone distribution depend on data sets with both positive (landslide presence) and negative (landslide absence) data (Conoscenti et al., 2016). The positive data are relatively fixed and are mainly selected from the landslide body cells in the inventory. The negative data are usually uncertain and are randomly selected as individual pixels outside of the landslide bod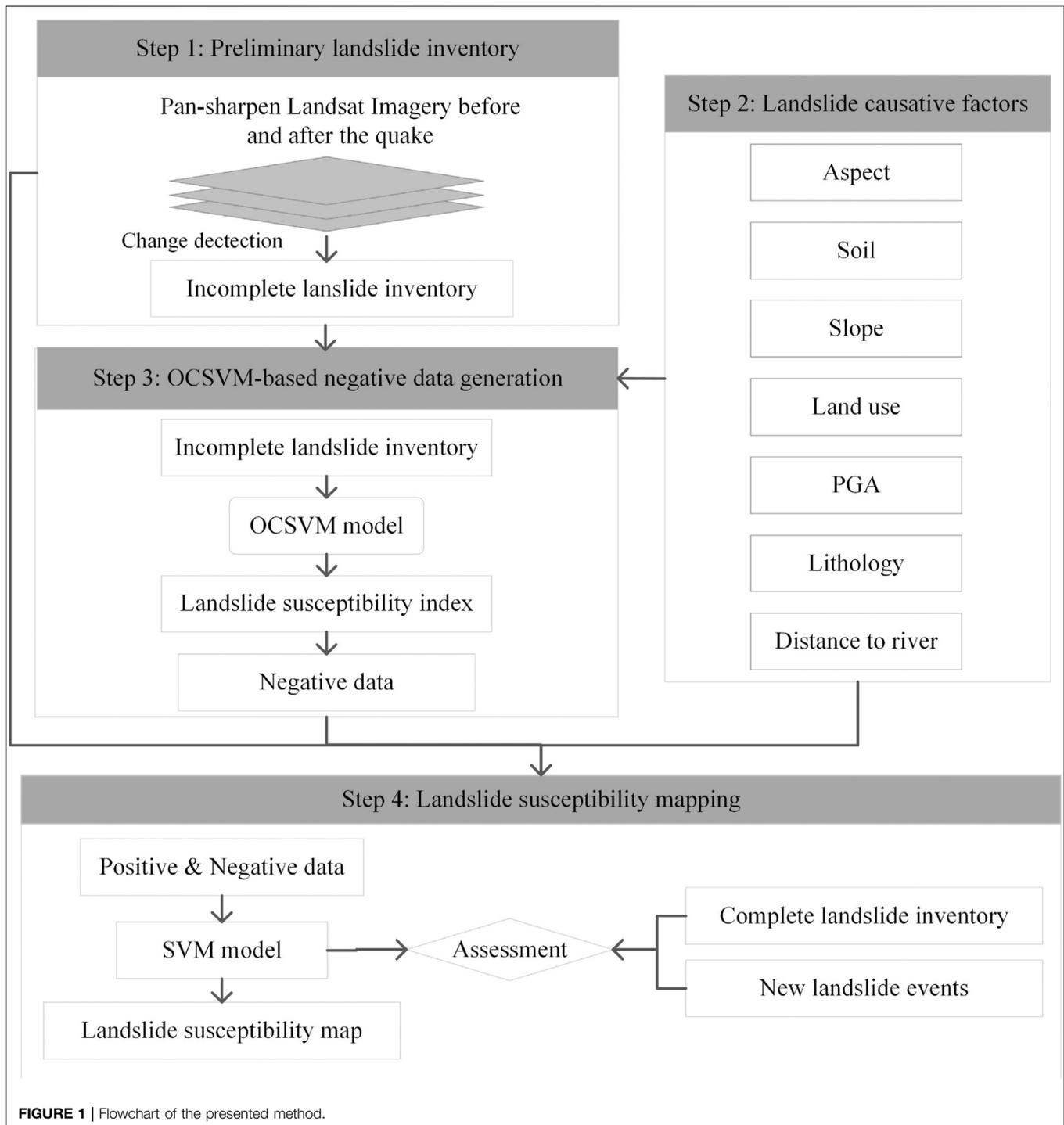y. Then, the LSM can be established using the machine learning models trained on both positive and negative data. To improve the LSM reliability, efforts have been made to optimize the selection strategy of negative data. First, we randomly select negative data from the non-landslide area or the area with a certain distance from the landslide body (Su et al., 2017). This is the most commonly used method for generating negative data, but it requires a complete LIM that covers all landslides. Besides, the earthquake changes the local geological structure, resulting in a lot of shatter mountains or unstable slopes, which still belong to the non-landslide area in LIM but are unsuitable to serve as the negative data. Secondly, we convert positive data into negative data by changing the feature space attribute information. The most representative method of this type is the target space exteriorization sampling (TSES) (Xiao et al., 2010). Since the negative data constructed by this method are from the feature space and the corresponding location cannot be found in the real world, it is difficult to conduct field verification. In addition, this method also requires a complete LIM. Lastly, generate the negative data by clustering analysis. This method clusters the data sets into multiple categories based on feature similarities and automatically generates negative data in the category that contains the fewest landslides. The commonly used clustering methods include the self-organizing map (SOM) (Huang et al., 2017) and similarity-based sampling (SBS) Zhu et al. (2019). Since this method is based on feature similarity, a complete LIM is still required to reflect the true feature space of seismic landslides.

All these negative data generation methods require a complete LIM, which is very difficult to obtain in a short time. It is easy to establish an incomplete LIM using remote sensing and GIS technologies. Thus, if the incomplete LIM can be used for generating reliable negative data, rapid LSM after the earthquake is possible. However, there are only a few studies on the LSM based on incomplete LIM, and this is due to the uncertainty in the generation of negative data (Chen et al., 2020). Therefore, efforts should be made on the generation of negative data based on incomplete LIM.

Based on the aforementioned analysis, this study presents a one-class-classifier-based negative data generation method for earthquake-induced LSM. Using the proposed method, we calculate the landslide susceptibility based on the incomplete LIM, and then automatically generate pseudo labeling of negative data from areas with low landslide susceptibility. After that, the generated negative data and the positive data (i.e., the incomplete LIM) are applied to train the traditional binary classifier to produce the final landslide susceptibility map. **Figure 1** shows the flowchart of the presented method. The left of this study is organized as follows. **Section 2** presents the study area and materials used in this study. **Section 3** introduces the proposed one-class-classifier-based negative data generation method for earthquake-induced LSM. **Section 4** presents the experimental results and discussions, and **Section 5** draws some conclusions.

2 STUDY AREA AND MATERIALS

Wenchuan, situated in Sichuan province, southwest China, has complex terrains and is the site of much intense neotectonic

Step 1: Preliminary landslide inventory

Pan-sharpen Landsat Imagery before and after the quake

Change dectection

Incomplete lanslide inventory

Step 2: Landslide causative factors

Aspect

Soil

Slope

Land use

PGA

Lithology

Distance to river

Step 3: OCSVM-based negative data generation

Incomplete landslide inventory

OCSVM model

Landslide susceptibility index

Negative data

Step 4: Landslide susceptibility mapping

Positive & Negative data

SVM model

Landslide susceptibility map

Assessment

Complete landslide inventory

New landslide events

FIGURE 1 | Flowchart of the presented method.

activity. The region features high ridges, lofty mountains, and crisscrossing gorges and valleys. The elevation fluctuates greatly from 782 m to 5,896 m. The neotectonic activities in this area are intense, and the Longmenshan thrust belt, which is composed of three main faults, namely, the Yingxiu-Beichuan fault, Guanxian-Anxian fault, and Mao-wen fault, runs across the whole of Wenchuan. Among them, the Yingxiu-Beichuan fault is inferred as the main structure of the 2008 earthquake (Li

et al., 2008). The complex terrain and intense structure make this region a area prone to seismic and geological disasters (Wu et al., 2020). For instance, the 2008 Wenchuan Mw7.9 earthquake occurred in this region and induced a large number of coseismic landslides, see **Figure 2**.

The data sources used in this study include pre- and post-earthquake Landsat 7 satellite images (acquisition dates: April 22, 2008, and May 25, 2008), a digital elevation model (DEM), river,
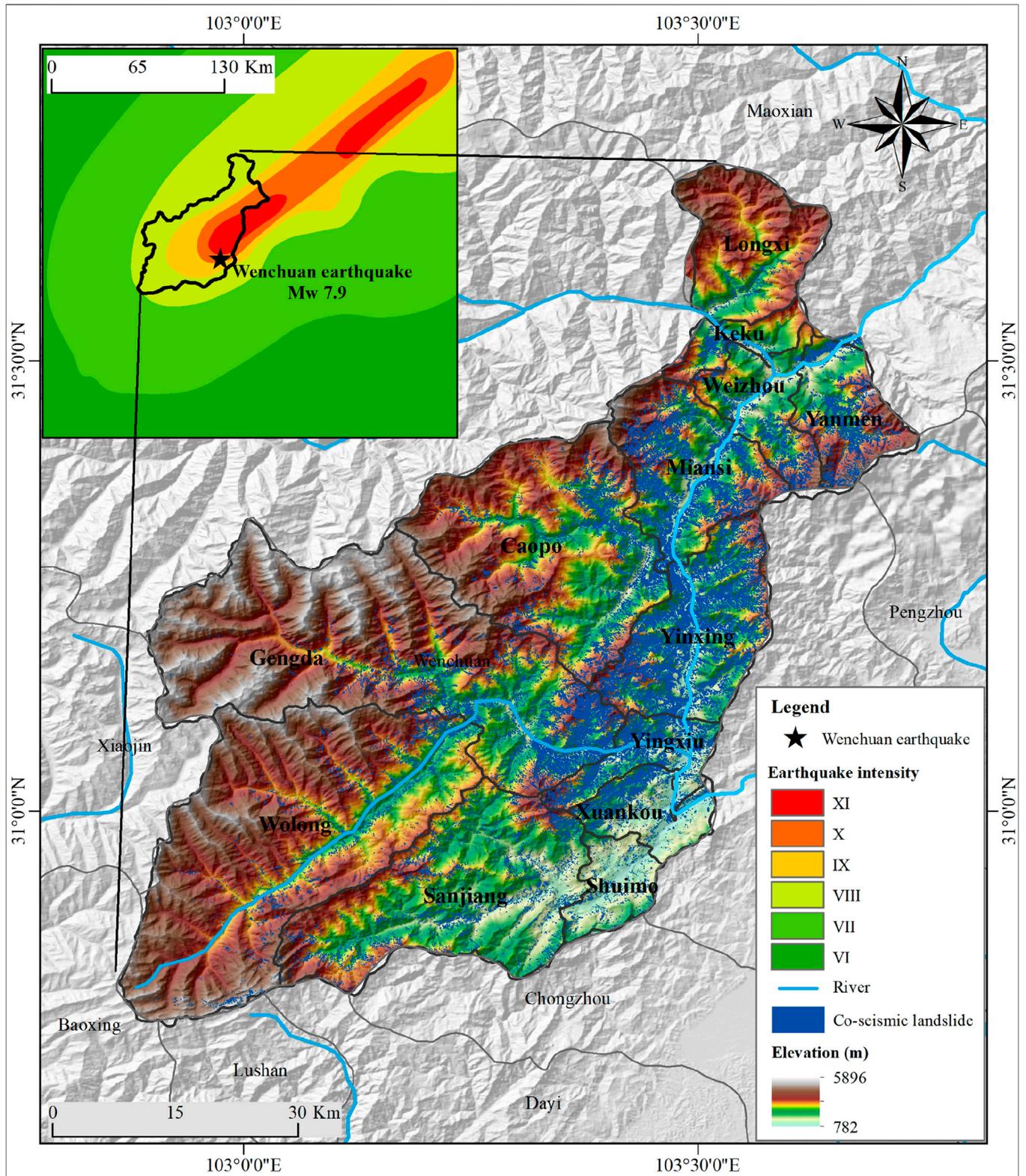
FIGURE 2 | Distribution of coseismic landslides induced by the 2008 Wenchuan earthquake.

Data	Resolution	Application	Data source
Landsat 7	Panchromatic: 15 m; Multispectral: 30 m	Co-seismic landslide	https://earthexplorer.usgs.gov/
DEM	30 m	Causative factor	https://gdex.cr.usgs.gov/gdex
River	Vector	Causative factor	https://www.webmap.cn/
Lithology	1:200,000	Causative factor	http://geocloud.cgs.gov.cn/
Land use	30 m	Causative factor	http://www.resdc.cn/data.aspx/
Soil	30 m	Causative factor	http://www.resdc.cn/data.aspx/
PGA	Vector	Inducing factor	https://earthexplorer.usgs.gov/
Complete LIM	Vector	Evaluation	https://earthquake.usgs.gov/earthquakes/

lithology, land use, soil, peak ground acceleration (PGA), and the complete LIM (see **Table 1**). The slope and aspect were derived from the DEM. The complete LIM was produced by the integration of field survey and rigorous analysis of very-high-resolution satellite images acquired by multiple platforms/sensors (e.g., aerial photo, IKONOS, and QuickBird) (Xu et al., 2014). The post-earthquake images used for the complete landslide inventory ranged from May 23, 2008, to June 13, 2008. The acquisition date of images used for two landslide inventories have a substantial overlap, which compresses the influence of the area and spatial distribution differences of landslides caused by time.

3 METHODOLOGY

3.1 Data Preparation

3.1.1 Acquisition of the Incomplete LIM Through Change Detection

This study compares satellite images before and after the earthquake to extract the landslides by change detection. To reduce the influence of terrain, atmosphere, and sensors, these images are preprocessed through radiation correction, atmospheric correction, orthographic correction, image registration, and a cloud mask. As the landslides will cause surface damage, especially vegetation damage, the normalized difference vegetation index (NDVI) (Ramos-Bernal et al., 2018), sensitive to vegetation change, is adopted as the main feature for change detection. Finally, the change area is separated from the background by the image difference method (Fung, 1990) in which the threshold is determined by **Eq. 3** (Lv et al., 2018).

$$NDVI = \frac{(NIR - R)}{(NIR + R)} \tag{1}$$

$$I_{diff} = \left| NDVI_{post} - NDVI_{pre} \right| \tag{2}$$

$$X_{I_{diff}} = \begin{cases} \text{landslide} & \text{if} \quad I_{diff} > m + k*\sigma \\ \text{background} & n = x + 2y \end{cases} \tag{3}$$

where m and σ are the mean value and standard deviation corresponding to I_{diff}, respectively. k is an adjustable parameter, and we determine the value of k in this study by trial-and-error method and get $k = 1.5$. After we remove the interference of pseudo-change information by morphology operation and manual interaction, the co-seismic LIM is finally obtained.

Figure 3 shows the coseismic LIM obtained from multi-temporal Landsat images. This inventory contains most medium-large landslides but fails to detect landslides in cloudy areas. Besides, limited by the image resolution, this inventory does not include small landslides with an area less than 5,000 m^2. A complete LIM (Xu et al., 2014) was taken as the ground truth to validate the extracted LIM. The completeness of the landslide extraction results is determined by the ratio of the area of the correctly extracted landslides to that of the real coseismic landslides. Specifically, the correctly extracted landslide in this study is 159.46 km^2, and the real coseismic landslide is 359.6 km^2, so the completeness of the result is 49%, which is far less than the real coseismic landslide. Therefore, the established landslide inventory is an incomplete LIM.

3.1.2 Slope Unit Generation

Mapping units, fundamental to LSM (Van Den Eeckhaut et al., 2009; Erener and Düzgün, 2012), include the pixel unit, slope unit (SU), watershed, unique condition unit, and terrain unit (Reichenbach et al., 2018). The pixel unit and slope unit are more widely used than the other three types. The pixel unit is a regular raster unit and capable of processing the resolution differences of data sources using very simple operations. Despite its popularity, the pixel unit is difficult to work with in complex terrain and struggles to distinguish landslide sources from accumulation areas, impacting the performance of LSM. SU represents a slope or a part of a slope. Landslides are geological hazards that develop on slopes; the number of landslides that occur on a slope reflects the slope stability (Sun et al., 2020). Also, SU suppresses the bias introduced by the incomplete LIM (Reichenbach et al., 2018). Based on its advantages, we chose SU as the mapping unit.

This study applies the hydrological analysis of DEM to extract SUs, including the acquisition of positive and negative DEM, extraction of flow direction and accumulated flow, generation of the river network and watershed, and SU generation (Wang et al., 2017). In this process, the accumulative flow threshold corresponding to the river network is the key to obtaining satisfactory SUs. This study sets the optimal accumulative flow threshold through trial and error. To further improve the accuracy of SUs, we use GIS editing tools to manually adjust the boundaries of SUs and eliminate unqualified SUs. Finally, we get 1,351 SUs, see **Figure 4**.

FIGURE 3 | The incomplete LIM extracted from multi-temporal Landsat images.
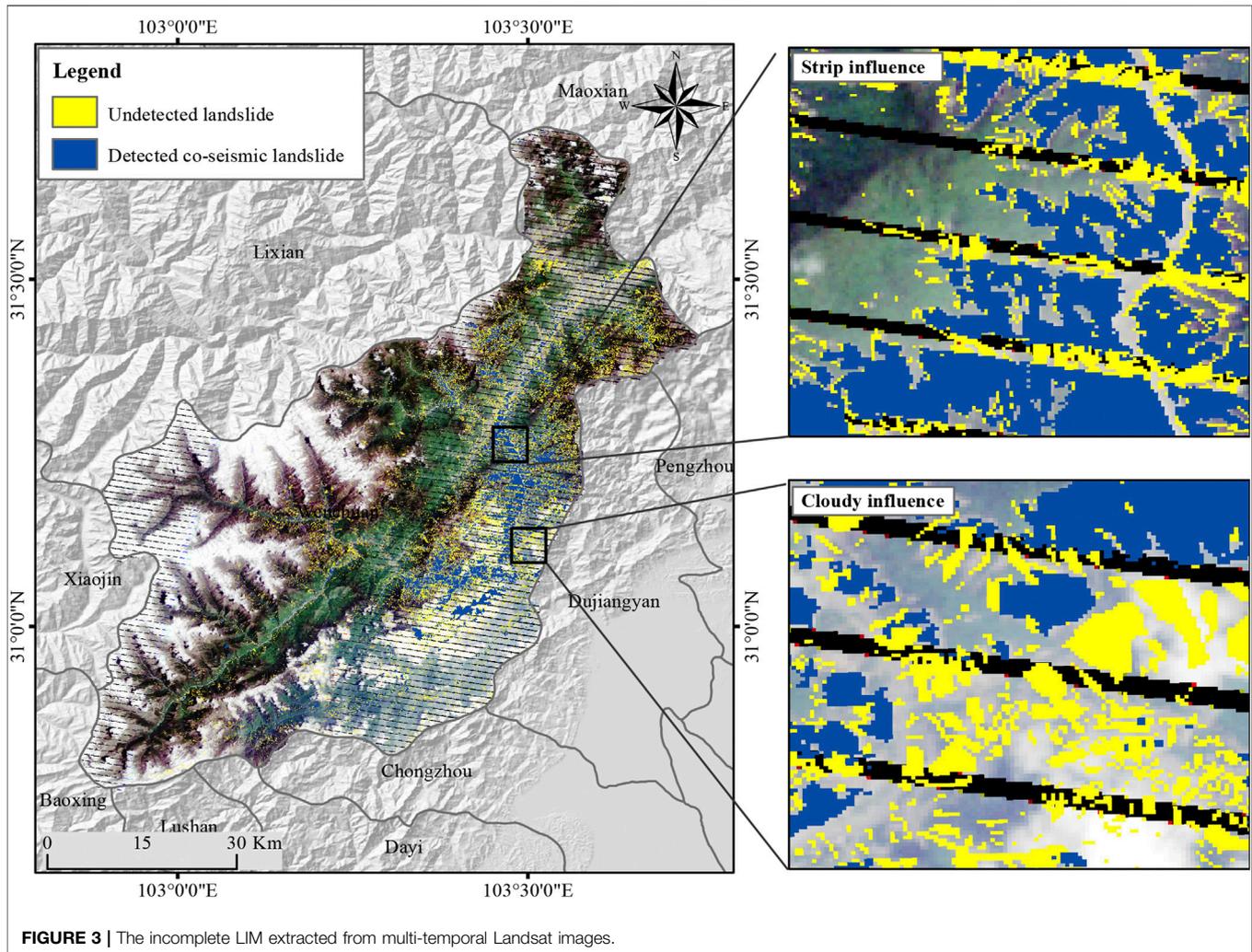
3.1.3 Selection of Causative Factors

A landslide is controlled by a series of causative factors (Pradhan and Lee, 2010), and thus accurately identifying the causative factors is critical for reliable LSM. Through analysis of the publication from 2005 to 2016 (Pourghasemi et al., 2018), this study selected the seven most widely used causative factors for earthquake-induced LSM in Wenchuan: slope angle, slope aspect, distance to river, lithology, soil, land use, and peak ground acceleration (PGA).

3.1.3.1 Slope Angle

Slope angle is a key factor that determines the failure of a landslide. Generally, slope angles between 30° and 40° are considered to be prone to landslides (Xu et al., 2014), as slope angles in this interval can enhance the shear force and surface water penetration, which provide favorable conditions for landslide occurrence. The slope information of the study area is obtained based on the digital elevation model (DEM), and the spatial analysis tool of ArcGIS is used to calculate the ratio of the maximum elevation difference between adjacent grids to the horizontal distance, which is the slope information of this grid. In this paper, the extracted slope information in the study area was divided into nine categories at an interval of 10°: 1) <10°; 2) 10–20°; 3) 20–30°; 4) 30–40°; 5) 40–50°; 6) 50–60°; 7) 60–70°; 8) 70–80°; and 9) >80°, see **Figure 5A**.

3.1.3.2 Slope Aspect

The slope aspect also affects the distribution of coseismic landslides, as slopes with different aspects carry different seismic effects (Zhou et al., 2016). In addition, there are differences in sunshine and rock weathering in different slope aspects, which leads to obvious differences in coseismic landslides in different slope aspects. For example, the spatial distribution of coseismic landslides in Wenchuan shows that the east, southeast, and south are the dominant slope aspects of seismic landslides [1]. Based on the DEM, this paper uses the spatial analysis tool of GIS to identify the downhill direction with the largest change rate from the current grid to the upper value of its adjacent grid direction, namely, the slope aspect, and divides the slope aspect into nine groups: 1) Flat (-1); 2) N (North, 0–22.5° and 337.5–360°); 3) NE (Northeast, 22.5–67.5°); 4) E (East, 67.5–112.5°); 5) SE (Southeast, 112.5–157.5°); 6) S (South,

FIGURE 4 | Slope units of the study area.
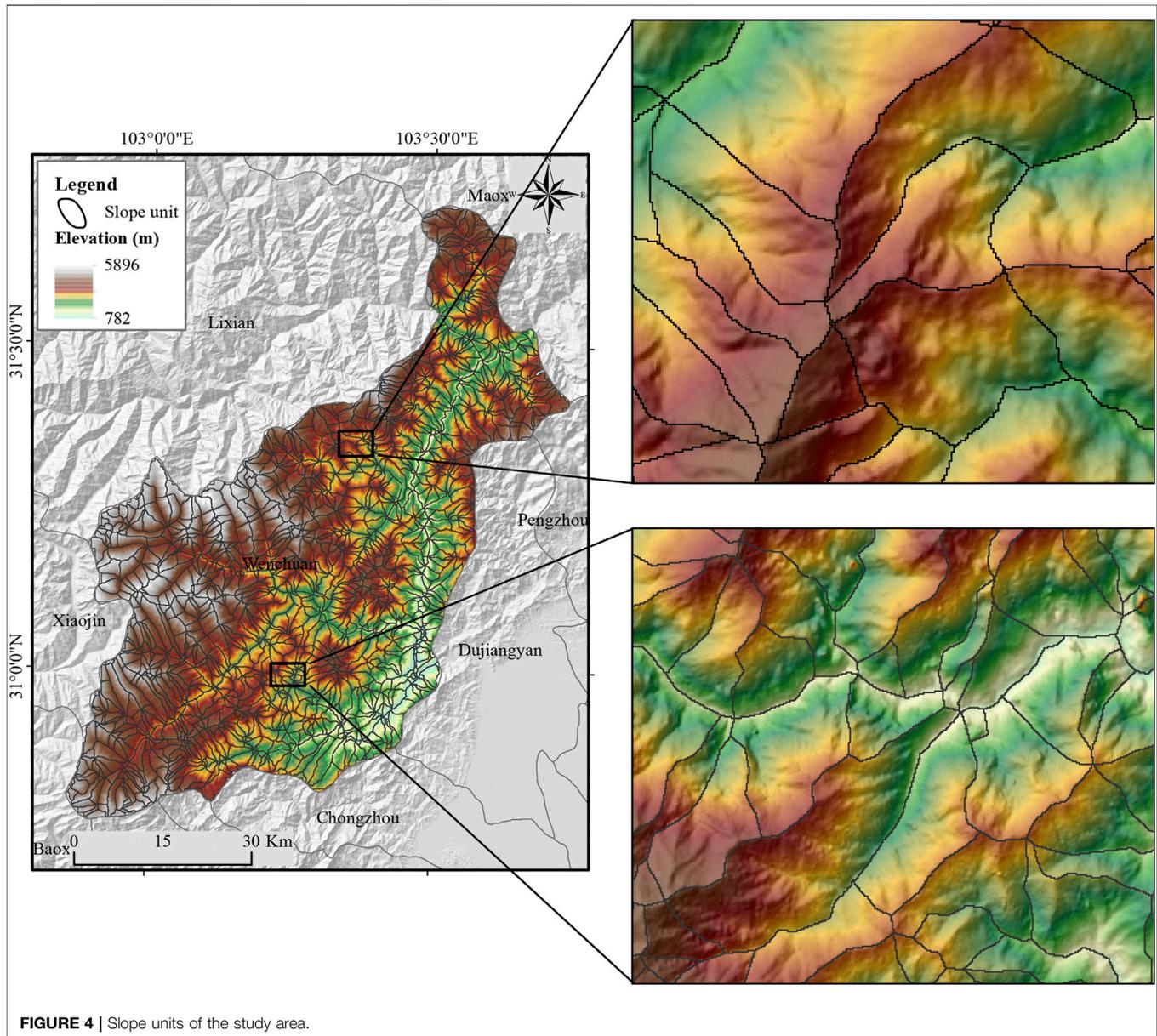
157.5–202.5°); 7) SW (Southwest, 202.5–247.5°); 8) W (West, 247.5–292.5°); and 9) NW (Northwest, 292.5–337.5°); see **Figure 5B**.
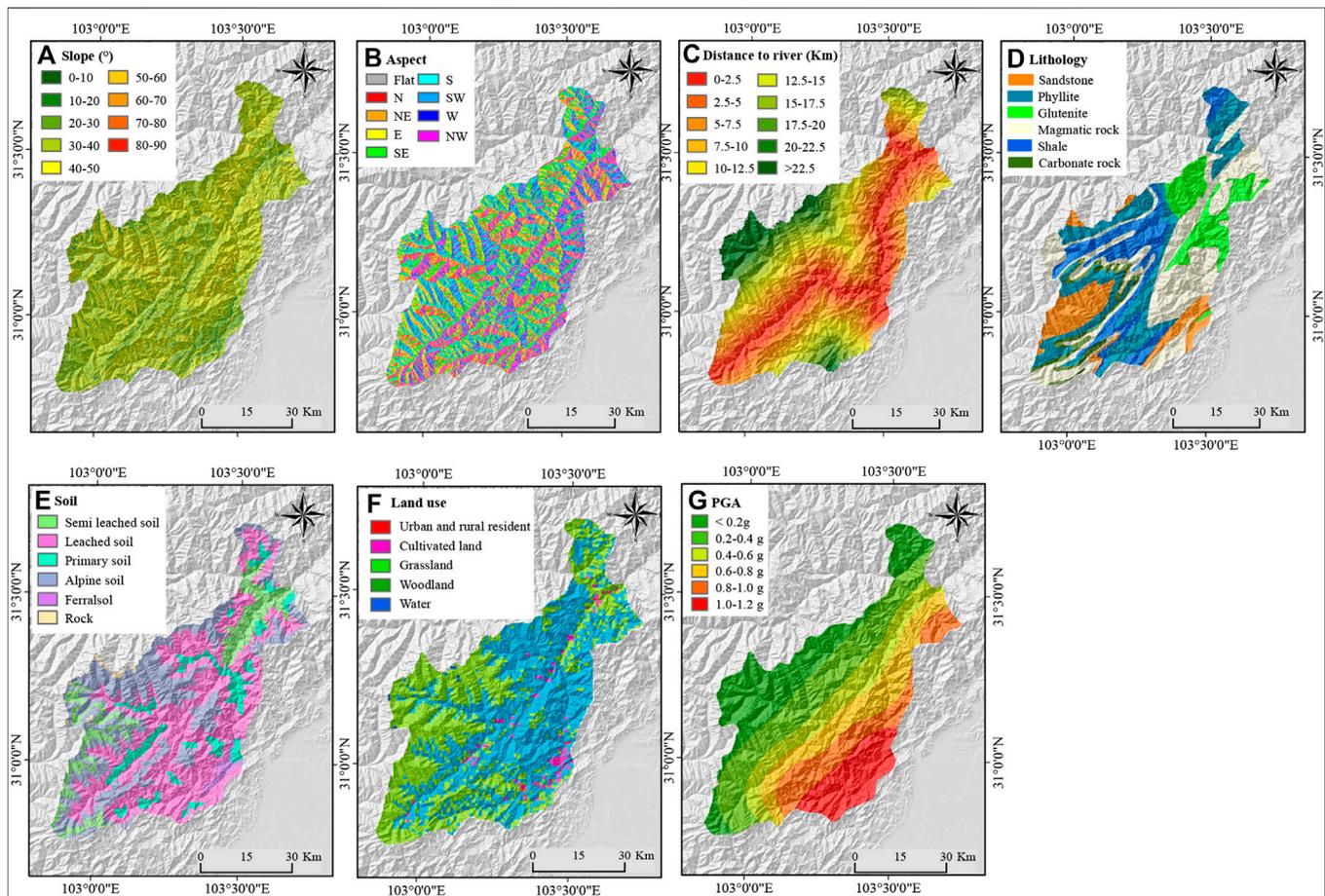
3.1.3.3 Distance to Rivers
The river controls the seismic landslide distribution through the action of groundwater and the erosion of water flow (Kamp et al., 2008). This study considers the distance between slopes and rivers and creates buffers around rivers to analyze the correlation between seismic landslides and their distance to rivers. According to the basic geographic information provided by the National Geographic Information System (NGIS), the initial river network in the study area was extracted. To guarantee the river network quality, we overlayed the initial river network on a cloud-free Landsat-5 mosaic in 2007 of the

study area. We deleted parts of the initial river network that did not exhibit rivers on our Landsat mosaic images. Rivers exhibited on Landsat mosaic image but not on the initial river network were added to the initial river network. Finally, we divide the distance to rivers into 10 groups with a step of 2.5 km: 1) 0–2.5 km; 2) 2.5–5 km; 3) 5–7.5 km; 4) 7.5–10 km; 5) 10–12.5 km; 6) 12.5–15 km; 7) 15–17.5 km; (8)17.5–20 km; 9) 20–22.5 km, and (10) >22.5 km see **Figure 5C**.

3.1.3.4 Lithology
The lithology is considered to be another important factor in dealing with landslide susceptibility assessment and hazards (Reichenbach et al., 2018). Lithology influences the topographic character of the landscape and how seismic energy is transmitted, particularly through elastic and

FIGURE 5 | The landslide causative factors used in this study. **(A)** Slope, **(B)** Aspect, **(C)** Distance to river, **(D)** Lithology, **(E)** Soil, **(F)** Land use, **(G)** PGA.

brittle/elastic properties of the rock, chemical weathering and its control of erosion and slope, fracture development and fault displacement, and seismic wave interactions with topography and lithological structures (Kargel et al., 2016). Using rock mass rating containing lithologic factors to conduct landslide hazard assessment is reasonable but, due to lack of enough field and laboratory data, cannot be applied to a large area (Tang et al., 2011). To determine the relative influence of bedrock lithology on the occurrence of landslides in a large area, the sensitivity index of the landslide occurrence of each rock type is calculated by comparing the landslide density in the area occupied by each rock type. In this paper, the relative influence of rock mass is determined by referring to the existing studies on Wenchuan earthquake landslides. The lithology data in the study area was generated by digitizing the hard copies of the 1:200,000 geological maps released by the China Geological Bureau. After that, the digitized lithology data were re-projected to WGS-84 coordinates and rasterized at 30 m resolution. Based on lithological similarities, the rock types were grouped into six classes, including 1) shale; 2) phyllite; 3) sandstone; 4) glutenite; 5) magmatic rock; and 6) carbonate rock, as shown in **Figure 5D**. The existing studies showed that the seismic

landslides in hard rock layers, including magmatic rock, carbonate rock, and glutenite, were the most developed and that the development density of shale, phyllite, and sandstone were the second (Huang and Li, 2009; Li et al., 2013).

3.1.3.5 Soil
The composition of the soil determines the permeability of the soil, the better the permeability of the soil, the more conducive it is to flood discharge; generally, sand particles are coarser, as they do not as easily produce runoff or small flow. The permeability of clay is poor; it is easy to erode, and geological disasters are occur with relative ease. For instance, granular, nonplastic, and low plasticity soils are more susceptible than fine soils (Maharaj, 1993). The soil data were generated from 1:1,000,000 soil distribution data in China provided by the data processing center of the Chinese Academy of Sciences. Similar to the vectorization process of the lithology data, the digitized data of soil was re-projected to WGS-84 and re-sampled to 30 m. The study area mainly contains six soil types, including 1) semi-leached soil, 2) leached soil, 3) primary soil, 4) alpine soil, 5) ferralsol, and 6) rock, see **Figure 5E**.

3.1.3.6 Land Use

Human activities have changed the land surface and caused great damage to the natural environment, making it easier for landslides to form alongside earthquake disturbances. Studies have shown that there is a strong correlation between the distribution of landslides and land use. However, obtaining land use data at a specific time is exceedingly difficult. Based on the land use map of 2005 collected from the data processing center of the Chinese Academy of Sciences, the initial land use map of 2007 before the Wenchuan earthquake was obtained by combining supervised classification and visual interpretation with a cloud-free Landsat-5 mosaic image acquired in 2007. Supervised classification consists of three steps: 1) to select training samples for five land-use types by visual interpretation, 2) to train the support vector machine using the training samples and classify the whole Landsat-5 mosaic image, and 3) to repeat the classification steps five times and select the land-use map with the highest accuracy. We compare the classified land use map and the initial land use map to produce the most reliable land use map. The study area mainly includes five land-use types: 1) residential area; 2) cultivated land; 3) grassland; 4) the forest land; and 5) water, see **Figure 5F**.

3.1.3.7 PGA

The strong ground motion that causes short-lived disturbances in the balance of forces within hill slopes is the main reason for slope failure (Li et al., 2013). Strong ground motion directly induces coseismic landslides. PGA is simply the maximum peak acceleration on the acceleration time history curve recorded at a site in the earthquake. It is often used as a parameter to describe strong ground motion (Ma and Xu, 2019). Previous studies have shown that there is a positive correlation between PGA and seismic landslide, and the greater the PGA value, the easier it is to trigger a landslide (Xu et al., 2014). After the Wenchuan earthquake, according to the ground motion peak recorded by the seismic sensor, the United States Geological Survey (USGS) established the PGA map of Wenchuan by interpolation method, field amplification correction, and ground motion attenuation model. In this paper, the latest PGA was download from USGS website, and the study area was divided into six groups, including: 1) <0.2 g, 2) 0.2–0.4 g, 3) 0.4–0.6 g, 4) 0.6–0.8 g, 5) 0.8–1 g, and 6) >1 g, see **Figure 5G**.

3.1.4 Multi-Collinearity Analysis of Landslide Causative Factors

Landslide causative factors have intra-correlations, which increases the data dimension and affects the model reliability. To separate the influence of causative factors, we employ the tolerance and variance inflation factor (VIF) (Chen et al., 2019) to conduct a multicollinearity test. Tolerance and VIF are expressed as follows:

$$Tolerance = 1 - R_j^2 \tag{4}$$

$$VIF = \left[\frac{1}{Tolerance}\right] \tag{5}$$

TABLE 2 | Multicollinearity analysis for the landslide causative factors.

Causative factors	Collinearity statistics	
	Tolerance	VIF
Soil	0.907	1.102
Slope	0.913	1.096
Lithology	0.974	1.027
Land-use	0.787	1.27
Aspect	0.979	1.021
Distance to river	0.821	1.218
PGA	0.722	1.385

where R_j^2 is the regression determination coefficient of landslide causative factors. If the VIF is high, the multi-collinearity influence is serious. This study sets VIF < 10 and tolerance > 0.1 (Chen et al., 2019) as the multi-collinearity threshold. **Table 2** presents the tolerance and VIF values of the seven landslide causative factors. The largest VIF and smallest tolerance are 1.385 and 0.722, respectively. The results suggest that there is no multi-collinearity relationship among the seven landslide causative factors, and thus all these factors are used for LSM.

3.2 Negative Data Generation

Traditionally, negative data are randomly selected from the areas outside the LIM. However, non-landslide areas often contain a large number of unstable slopes, which are not identified in the LIM. Besides, there are also many unrecognized landslides in non-landslide areas. The unstable slopes and unrecognized landslides cannot be used to select negative data. To overcome these shortcomings, this study presents a one-class support vector machine (OCSVM) (Schölkopf et al., 2000) based negative data generation method. The OCSVM maps the samples from low dimension to high dimension space through the kernel function and finds the optimal hyperplane between the origin and the high dimension space. Specifically, the distance between the sample and the hyperplane indicates the correlation between the sample and a specific class. The decision function is defined as the following:

$$f(x) = \text{sign}\left((\omega.\Phi(x)) - \rho\right) \tag{6}$$

where $\Phi(x)$ represents the mapping function, weight ω, and threshold ρ of the support vector are obtained by solving the quadratic programming problem.

$$\min \frac{1}{2}\|\omega\|^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - \rho \tag{7}$$

$$\text{s.t.} \ (\omega.\Phi(x_i)) \geqslant \rho - \xi_i \ \xi_i \gg 0$$

where ξ_i is the relaxation variable, and $\nu \in (0, 1)$ controls the proportion of support vectors in the training samples. By introducing the kernel function $k(x)$, the above optimization problem is transformed into a dual form:

$$\min_a \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j k(x_i, x_j) \tag{8}$$

$$\text{s.t.}\, 0 \leqslant a_i \leqslant \frac{1}{\nu N} \sum_{i=1}^{N} a_i = 1$$

where $\rho = \sum_{i=1}^{N} a_i K(x_i, x_j)$ denotes the threshold value, and weight ω determines the optimal hyperplane. The Gaussian kernel is selected as the kernel function to construct OCSVM that is run with the LIBSVM software package (Chang and Lin, 2011). The optimal hyper-parameters are obtained through five-fold cross-validation.

Selecting negative data from non-landslide areas is a commonly used method for LSM, but its feasibility relies on a complete LIM. By contrast, this study uses an incomplete LIM obtained by change detection to generate negative data. Suppose S represent the data sets of the study area. L_i and L_c represent landslides in the incomplete LIM and in the complete LIM, respectively. $N_i = \overline{L_i} \cap S$ and $N_c = \overline{L_c} \cap S$ represent non-landslides in the incomplete LIM and in the complete LIM, respectively. $N_p = \overline{L_i} \cap \overline{L_c}$ represents the pseudo non-landslide, and N_{OCSVM} denotes the extremely low susceptibility areas in the LSM produced by OCSVM.

To analyze the influence of different negative data generation strategies on LSM, we designed four negative data generation strategies:

1. Negative data T_c are randomly selected from N_c (i.e., $T_c \in N_c$)
2. Negative data T_p are randomly selected from N_p i.e., $T_p \in N_p$)
3. Negative data T_i are randomly selected from N_i (i.e., $T_i \in N_i$)
4. Negative data T_{OCSVM} are randomly selected from N_{OCSVM} (i.e., $T_{OCSVM} \in N_{OCSVM}$)

3.3 LSM

The aforementioned steps produce training samples to train a statistical learning algorithm. The support vector machine (SVM) is a common machine learning algorithm that is capable to build a stable and reliable statistical model with a small number of samples (Suykens and Vandewalle, 1999). Therefore, this study selects SVM to compute the landslide susceptibility values for SUs. The decision function is expressed as follows:

$$y = \sum_{i=1}^{n} \omega_i x_i + b \qquad (9)$$

where ω is the support vector weight and b the intercept. To suppress the noise influence, the relaxation variables ξ is introduced, and the corresponding convex quadratic programming of SVM is expressed as follows:

$$\min_{\omega, b} \frac{1}{2} \omega^2 + C \sum_{i=1}^{n} \xi_i \qquad (10)$$

$$\text{s.t.}\, y_i (\omega \cdot x_i + b) \geqslant 1 - \xi_i$$
$$i = 1, 2, 3, \dots n$$

where C represents the penalty coefficient. The optimal hyperplane can be obtained by converting the appeal planning problem into a dual problem that is expressed as follows:

$$\min_{a} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{n} a_i \qquad (11)$$

$$\text{s.t.} \sum_{i=1}^{n} a_i y_i = 0$$

$$0 \leqslant a_i \leqslant C, i = 1, 2, 3, \dots n$$

where $(x_i \cdot x_j)$ represents the kernel function that projects samples to high dimensional space to handle complex nonlinear problems. This study chooses the Gaussian kernel as the kernel function, and the optimal hyper-parameters are obtained by the five-cross validation (Hong et al., 2017). This study implements SVM by the LIBSVM package (Chang and Lin, 2011).

3.4 Model Evaluation

The reliability of LSM is mainly determined by the accuracy of the model. But there is no uniform standard to assess the model accuracy. The common method is to use a set of independent data sets for accuracy assessment. Test sets reserved in the early data preparation phase are widely used for model evaluation (Irigaray et al., 2007; Dou et al., 2019). Considering that the data set used in this study is the incomplete LIM, and the actual complete LIM does not participate in the model construction, the complete LIM is taken as the test set for model evaluation.

Receiver operating characteristics (ROC) is an important tool for model evaluation (Swets, 1988). It calculates the true positive rate and false positive rate of the model according to different discriminant standards and draws a curve with the true positive rate and the false positive rate as the x-axis and the y-axis, respectively. The area under the curve (AUC), generally between 0.5-1, is often used to reflect the model performance. The greater the AUC value, the better the model performance will be.

$$\text{TruePositiveRate} = \text{Sensitivity} = \frac{TP}{TP + FN} \qquad (12)$$

$$\text{TrueNegativeRate} = \text{Specificity} = \frac{TN}{TN + FP} \qquad (13)$$

$$\text{FalsePositiveRate} = 1 - \text{Specificity} = \frac{FP}{TN + FP} \qquad (14)$$

where TN, TP, FP, and FN represent true negative, true positive, false positive, and false negative, respectively.

As LSM reflects the landslide occurrence possibility in the earthquake-affected region (Guzzetti et al., 2006), this study counts the number of new landslides that occurred in the hazardous areas to further verify the model performance. To this end, this study collected 11 landslide events that occurred in the study area after 2008 from the website of the China Geological Survey (referred to as new landslides hereafter). The higher number of new landslides in hazardous areas of LSM, the better the model performance will be. In this study, the hazardous areas include areas with extremely high susceptibility, high susceptibility, and moderate susceptibility levels.

FIGURE 6 | Landslide susceptibility map generated by OCSVM.
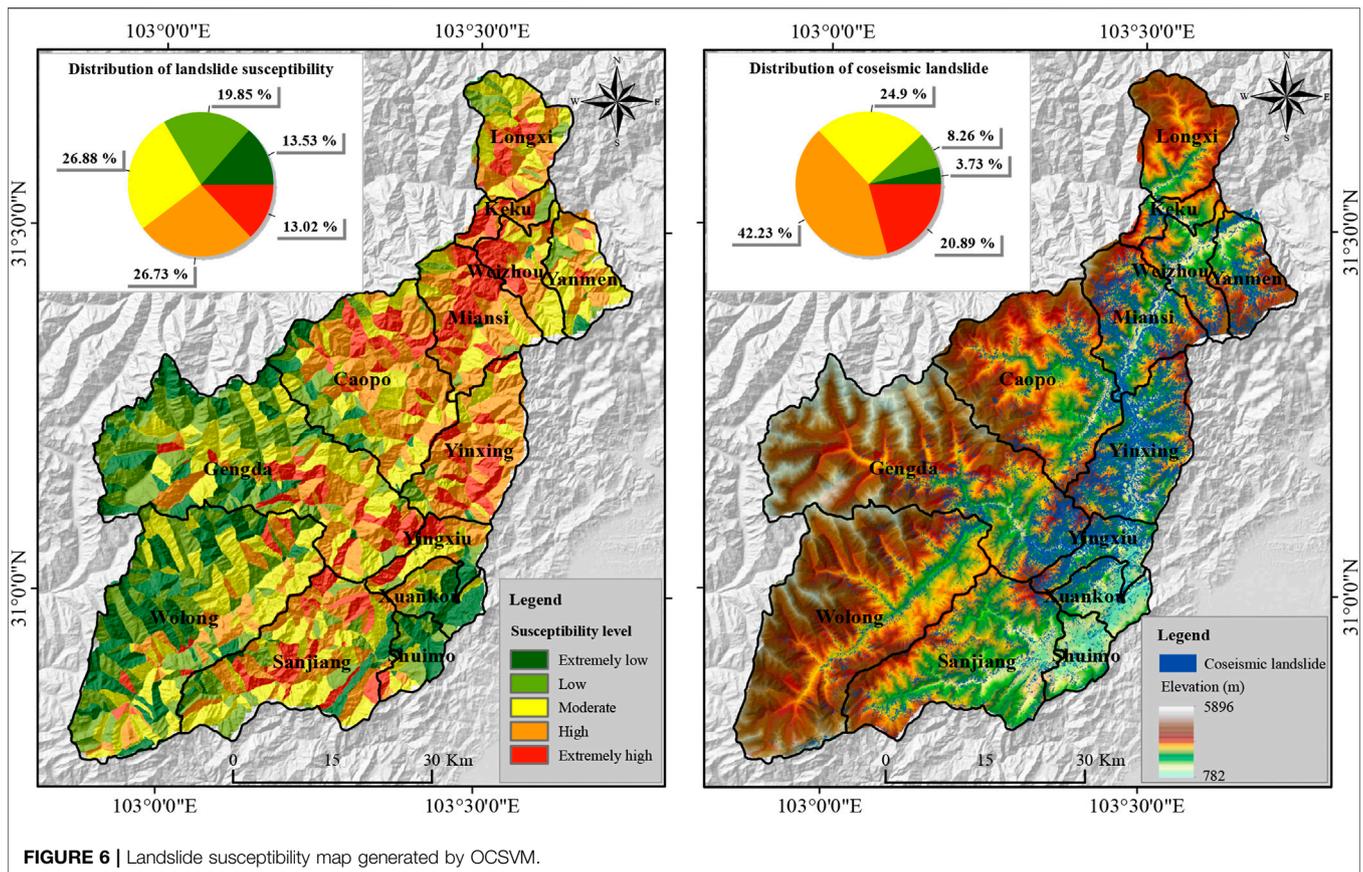
4 EXPERIMENTAL RESULTS

4.1 Evaluation of the Reliability of the Negative Data Generation Method Based on OCSVM

Figure 6 shows the LSM derived from OCSVM. It can be seen that 88.02% of the landslides fall into the hazard area (e.g., extremely high susceptibility, high susceptibility, and moderate susceptibility areas). This suggests that OCSVM can obtain a reliable landslide susceptibility map using positive data only. Meanwhile, extremely low susceptibility areas cover 13.53% of the study area and share 3.73% landslides. In other words, landslides rarely occur in extremely low susceptibility areas derived from OCSVM, which in turn guarantees the quality of the generated labeled negative data sets.

4.2 Accuracy Assessment of LSMs Derived From Different Strategies

Figure 7 shows the landslide susceptibility maps produced by SVM based on different training sample sets. The hazard area percentages of four methods ranges from 46.70% to 60.39%. The generated LSM maps show apparent differences in the spatial distribution of susceptibility area. As both the machine learning algorithm and the positive data are fixed, this difference comes

from negative data generation strategies. This indicates that negative data generation strategy and the LSM performance have a close relationship. To quantitatively evaluate the model performance, **Figure 8** presents ROC curves for four methods. The AUC values derived from ROC curves are 0.8164, 0.7673, 0.7928, and 0.8100, respectively. Among four methods, SVM based on T_c achieves the highest AUC value (i.e., 0.8164), while SVM based on T_p achieved the lowest AUC value (i.e., 0.7673). The AUC value of the presented method is 0.81, which is close to the best performance achieved by T_c. The AUC difference between the presented method and T_c is marginal, suggesting that the presented method based on the incomplete LIM is capable of achieving a comparable performance to that of the traditional method based on the complete LIM. Therefore, the presented method is suitable for rapid earthquake-triggered landslide susceptibility mapping when only the incomplete landslide inventory is efficient and timely in terms of obtaining data after the occurrence of an earthquake.

This study applies the natural break classification to divide the landslide susceptibility into five levels: extremely high, high, moderate, low, and extremely low susceptibility. By overlaying the actual coseismic landslide distribution (i.e., the complete LIM), the areas of different landslide susceptibility levels and corresponding coseismic landslide areas were computed. **Table 3** shows that the evaluation results corresponding to different negative data sets are quite different, and thus the negative
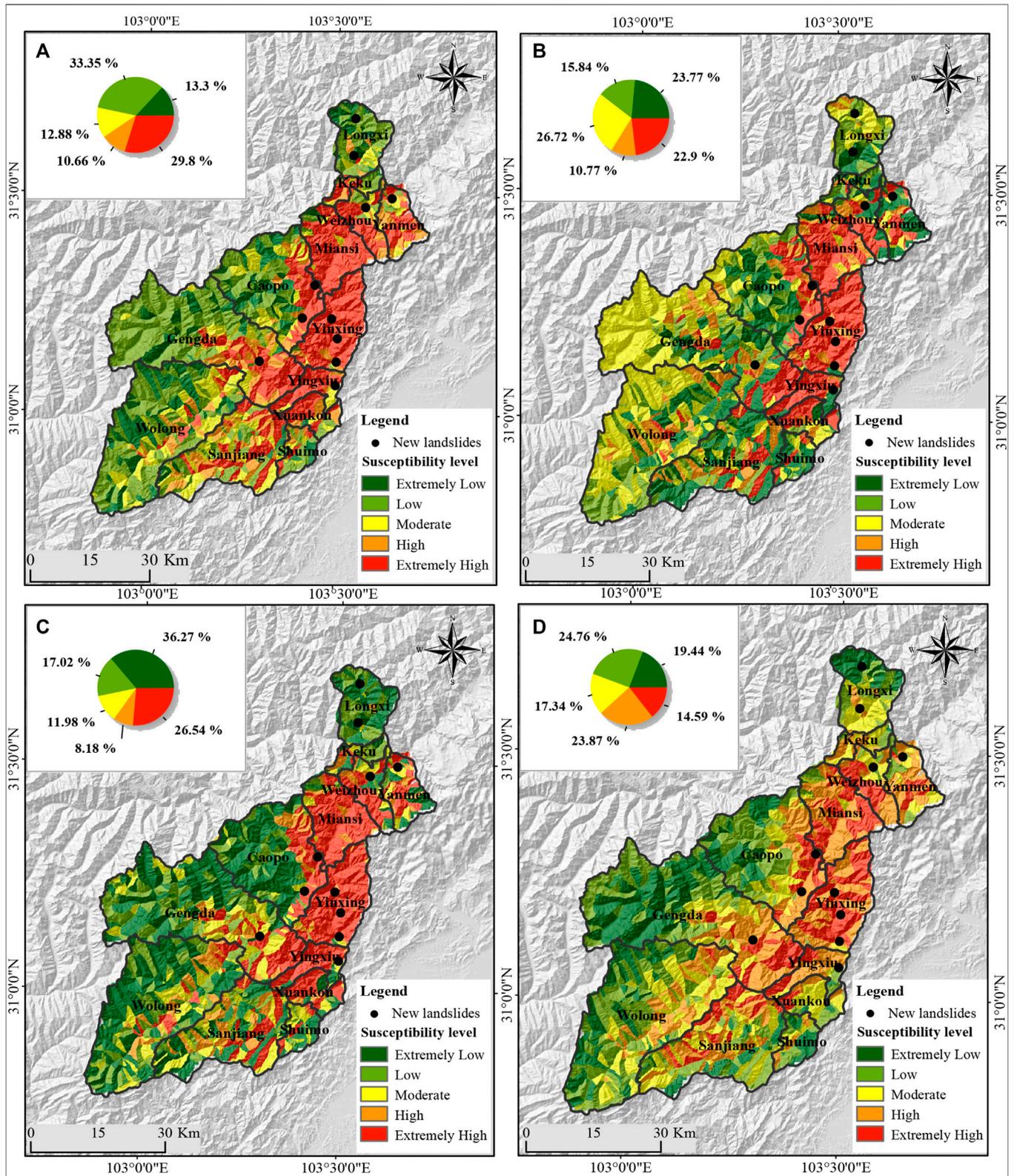
FIGURE 7 | Landslide susceptibility map obtained based on the SVM using different negative data: **(A)** T_c, **(B)** T_p, **(C)** T_i, and **(D)** T_{OCSVM}.
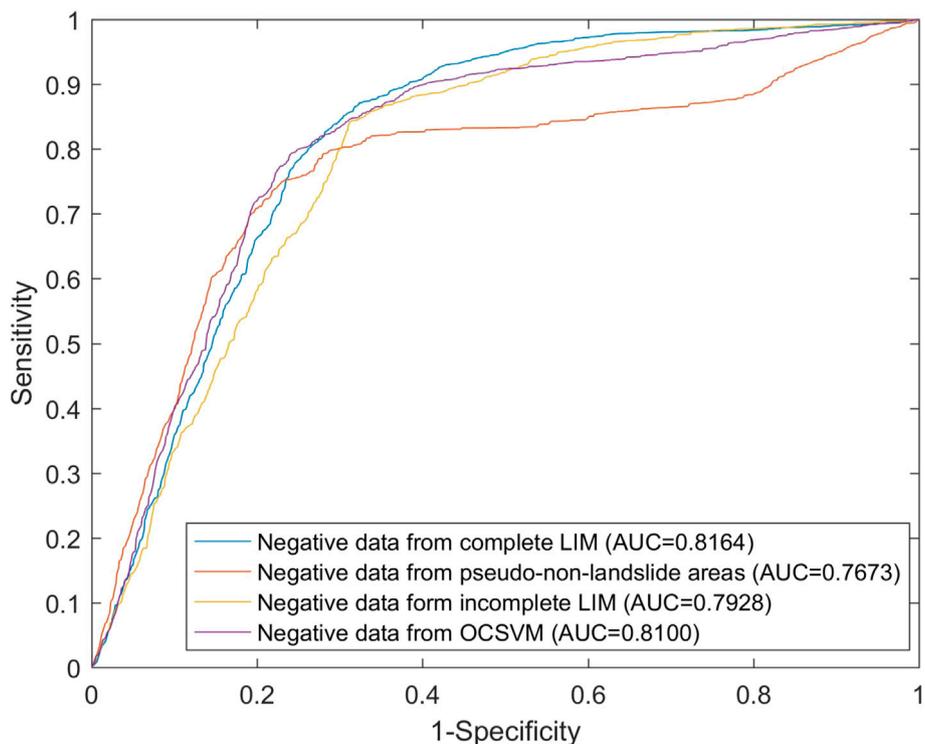
FIGURE 8 | ROC curves of four negative data generation strategies.

TABLE 3 | Landslide susceptibility levels and density of landslides in the study area.

Susceptibility level	Area (km²)	Percent of area	Area of landslide (km²)	Percent of landslide	Density
T_c					
Extremely low	536.28	0.13	6.88	0.02	0.01
Low	1,344.24	0.33	18.19	0.05	0.01
Moderate	519.34	0.13	27.66	0.08	0.05
High	429.85	0.11	36.23	0.1	0.08
Extremely high	1,201.32	0.3	267.48	0.75	0.22
T_p					
Extremely low	958.07	0.24	58.03	0.16	0.06
Low	638.49	0.16	14.14	0.04	0.02
Moderate	1,077.06	0.27	18.86	0.05	0.02
High	434.34	0.11	35.81	0.1	0.08
Extremely high	923.08	0.23	229.61	0.64	0.25
T_i					
Extremely low	1,461.92	0.36	31.85	0.09	0.02
Low	686.22	0.17	14.01	0.04	0.02
Moderate	483.12	0.12	27.33	0.08	0.06
High	329.85	0.08	28.27	0.08	0.09
Extremely high	1,069.93	0.27	254.98	0.72	0.24
T_{OCSVM}					
Extremely low	783.83	0.19	8.33	0.02	0.01
Low	998.04	0.25	22.29	0.06	0.02
Moderate	698.86	0.17	34.14	0.1	0.05
High	962.21	0.24	158.87	0.45	0.17
Extremely high	588.1	0.15	132.82	0.37	0.23

TABLE 4 | The percentage of coseismic landslides in the hazardous areas of towns and villages detected by the LSM using different negative data.

Subarea unit	Percent of coseismic landslide (%)	Coseismic landslides distribution			
		T_c	T_p	T_i	T_{OCSVM}
Yinxing	24.85	1.00	0.99	0.98	1.00
Gengda	14.90	0.90	0.73	0.81	0.87
Miansi	14.13	0.96	0.92	0.96	0.97
Caopo	9.12	0.84	0.75	0.8	0.93
Yingxiu	8.24	0.95	0.89	0.94	0.93
Sanjiang	6.56	0.58	0.48	0.39	0.70
Yanmen	6.50	0.96	0.67	0.71	0.87
Weizhou	4.35	0.88	0.66	0.75	0.95
Wolong	4.24	0.45	0.41	0.55	0.44
Xuankou	3.66	0.85	0.75	0.93	0.83
Keku	1.47	0.59	0.25	0.58	0.98
Shuimo	1.18	0.57	0.45	0.50	0.36
Longxi	0.80	0.38	0.03	0.35	0.98

data generation strategy has an essential influence on the performance of LSM. Specifically, for negative data T_c, 93% of the coseismic landslide exists in 54% of the hazardous areas of LSM, which is consistent with the coseismic landslide distribution. However, for negative data T_p, only 80% of the coseismic landslides occurred in 61% of the hazardous area of LSM. This indicates that the model does not evaluate the landslide susceptibility well. The main reason is that the negative data are from the pseudo-stability area, that is, the negative data belong to the unstable area of the study area. Negative data T_i has considerable uncertainty, due to the incomplete LIM and high concealment of potentially unstable slopes, resulting in significant fluctuations in the results. In this study, 87% of the coseismic landslides are distributed in 47% of the hazardous area of LSM. OCSVM can effectively obtain the stable area of the study area, so the influence of the potentially unstable slopes and the incomplete LIM can be well eliminated. Finally, 92% of coseismic landslides occurred in 56% of the hazardous area of LSM, which is higher than that of other negative data sources generated based on incomplete LIM.

To analyze the site scale accuracy of the LSMs generated from different negative data, we took towns and villages as the subarea unit. We calculated the ratio of coseismic landslide to the hazardous areas in each subarea. **Table 4** shows the percentage of the coseismic landslide distributed in the hazardous areas in each subarea. The landslide susceptibility maps show remarkable differences in towns and villages, especially the unrecognized landslide area in the incomplete LIM. Specifically, for the results of T_c, the percentage of the coseismic landslides in the hazardous area of several towns is less than 60%. Especially in Longxi Town, the coseismic landslide distributed in the detected hazardous area is as low as 38%. The results indicate that although the negative data are from the ideal non-landslide area, there are some unstable slopes affecting the accuracy of LSM. For negative data T_p, the distribution of the landslide susceptibility area is inconsistent with that of a coseismic landslide, especially in Longxi Town, where the percentage of a coseismic landslide is only 3%. For T_i, the

percentage of coseismic landslides in the detected hazardous areas is less than 60%. In Longxi Town, only 35% of the coseismic landslide are distributed in the detected hazardous areas, and the landslide susceptibility of this area was also not well evaluated. Finally, for the negative data T_{OCSVM}, the model estimated results are consistent with the coseismic landslide distribution for most towns. Only the percentages of two towns are less than 60%. Longxi Town is not covered by the incomplete LIM, but 98% of the coseismic landslides that occurred here are in the detected hazardous area. **Table 4** also shows that T_c, T_i, and T_{OCSVM} achieved the best performance in five, two, and seven administrative districts, respectively. T_p delivered the poorest performance among all subareas, suggesting that T_p is not suitable to train statistical learning algorithms for LSM. The third interesting finding is that T_{OCSVM} achieved a comparable or even better performance to that of T_c in most cases. However, in two subareas (i.e., Keku and Longxi), T_{OCSVM} achieved a substantial improvement compared to T_c. Thus, T_{OCSVM} shows a great advantage in areas with a low percent of coseismic landslides.

We also applied 11 new landslides that occurred after 2008 to further verify the performance of LSM. **Figure 7** shows the superposition of new landslides over the landslide susceptibility maps. We counted the number of new landslides that occurred in the hazardous areas to assess the LSM reliability. **Table 5** shows that 9, 7, 9, and 10 new landslides occurred in the hazardous areas produced by T_c, T_p, T_i, and T_{OCSVM}, respectively. Among four negative data generation strategies, T_{OCSVM} achieves the highest match rate with new landslides while T_p the lowest match rate with new landslides. Therefore, the presented negative data generation strategy can effectively assess the landslide susceptibility in earthquake-affected areas.

4.3 The Influence of One-Class Classifier Selection

To test the influence of one-class classifier (OCC) selection on the performance of LSM, another two benchmark OCCs, including SOM and TSES, are selected to compare with OCSVM. SOM is a kind of autonomic learning neural network without tutors. Its hierarchy consists of an input layer and competition layer (Huang et al., 2017). The input layer accepts external input variables, and the competition layer realizes clustering by analyzing and comparing the input variables. The data of one class have similar features, so the negative data can be generated by cluster analysis. In this study, five classes were selected as the final clustering results. TSES directly generates pseudo negative data based on positive data in feature space (Xiao et al., 2010). It exteriorizes positive data to become negative by replacing the value of one of its features with a new one outside the value range of this feature of all positive data.

We produced LSMs using the SVM based on different negative data sets generated by OCSVM, SOM, and TSES, respectively. The results are referred to as LSM-OCSVM, LSM-SOM, and LSM-TSES, see **Figure 9**. LSM-OCSVM is consistent with the distribution of coseismic landslides, suggesting the proposed OCSVM method gets high accuracy landslide susceptibility in
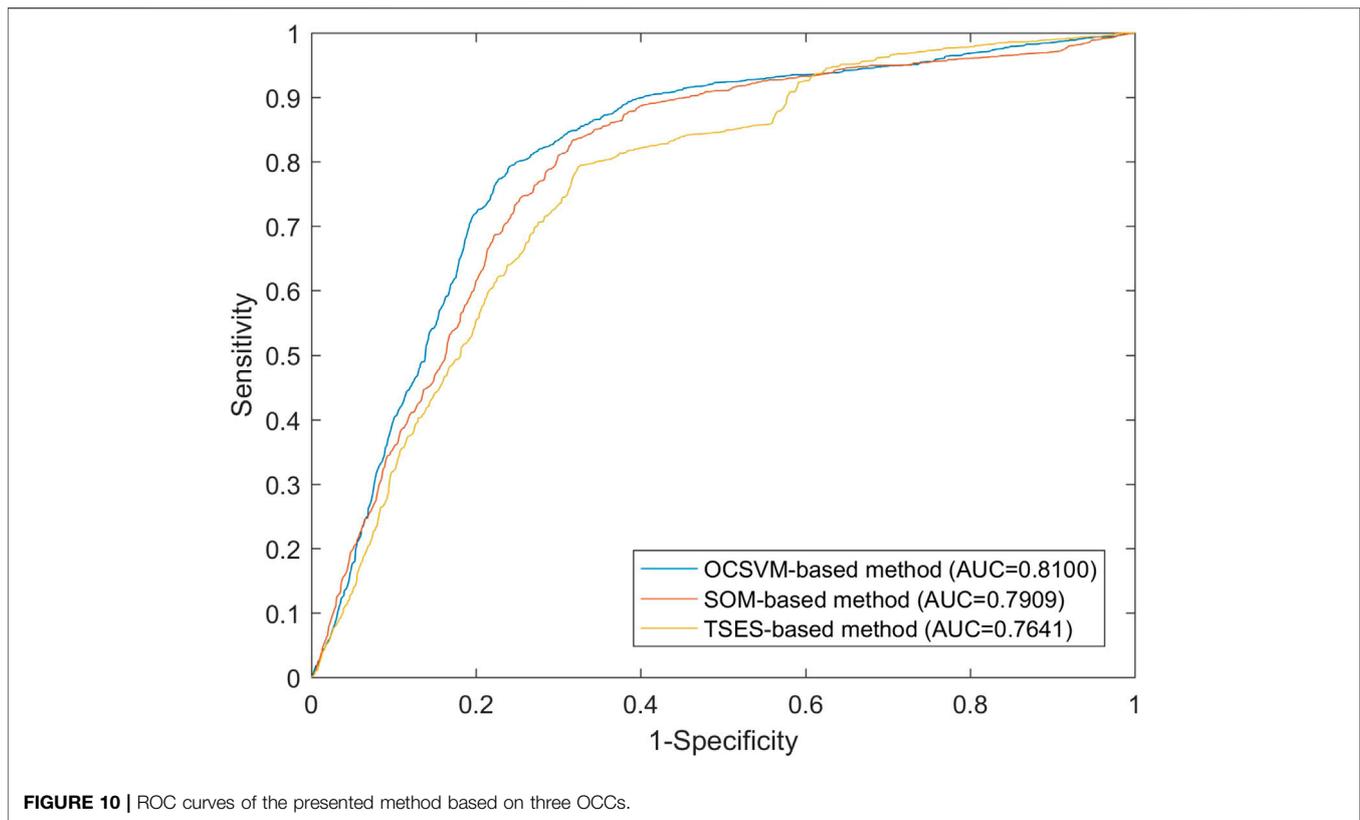
TABLE 5 | LSM and new landslide events.

New landslides	Lon	Lat	T_c	T_p	T_i	T_{OCSVM}
2009.7.25 (Chediguan bridge)	103.48	31.21	Extremely high	Extremely high	Extremely high	Extremely high
2010.5.30 (Suoqiao village)	103.64	31.49	Moderate	Extremely low	Moderate	Moderate
2010.6.12 (Jinbo village)	103.41	31.22	Extremely high	Extremely high	Extremely high	High
2011.7.3 (Dongjienao village)	103.50	31.12	Extremely high	High	High	High
2011.7.3(Fengxiangshu village)	103.50	31.06	High	Extremely high	Extremely high	Moderate
2011.7.3 (Maojiawang)	103.50	31.17	Extremely high	Extremely high	Extremely high	Extremely high
2013.7.22 (Zuwang village)	103.44	31.29	Extremely high	High	Extremely high	Extremely high
2018.4.8 (Aer village)	103.54	31.67	Low	Low	Extremely low	Extremely low
2018.7.26 (Minjiang bridge)	103.57	31.45	High	High	High	Moderate
2019.8.20 (Longtan village)	103.3	31.12	Moderate	Low	Extremely high	High
2020.6.20 (Kuapo village)	103.54	31.59	Extremely low	Extremely low	Extremely low	Moderate
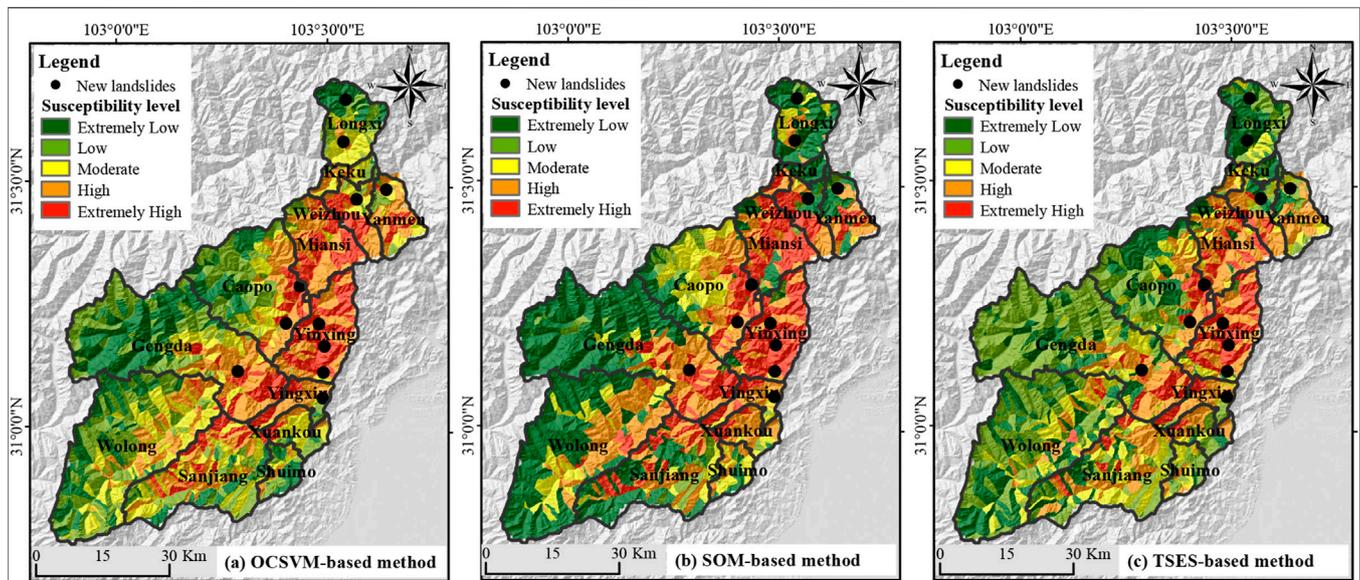
FIGURE 9 | Landslide susceptibility maps produced by: **(A)** OCSVM-based method, **(B)** SOM-based method, and **(C)** TSES-based method.

the post-earthquake area. Most of the LSM-SOM is consistent with the coseismic landslide distribution. However, the results are quite different from the actual situation at the junction of Miansi Town and Yingxiu Town. The TSES-based method underestimates the degree of earthquake disturbance, so many coseismic landslides are in the low susceptibility area.

To quantitatively evaluate the accuracy of the LSM obtained from different negative data generation strategies, we used the complete LIM to draw the ROC curves and calculate the corresponding AUC values. **Figure 10** shows that AUC values of OCSVM, SOM, and TSES are 0.8100, 0.7909, and 0.7641, respectively. Among three OCCs, OCSVM achieves the highest AUC value and TSES the lowest value. This result suggests that the selection of OCC has an influence on the performance of LSM, and thus the OCC should be carefully selected. **Table 6** shows the area percentage of coseismic landslides that occurred in the hazardous areas. The proposed OCSVM method has the highest value with a percentage of 92%, and this was followed by the TSES

method at 86% and the SOM method at 84%. Specifically, 10 of 11 new landslides occurred in the hazardous area of the landslide susceptibility map obtained by the OCSVM negative data generation method. In contrast, the results based on SOM and TSES correspond to eight and six new landslide events, respectively. The main reasons for this result are as follows. Firstly, the result of the negative data generation method based on cluster analysis depends on two aspects. The clustering analysis in this study was carried out based on the frequency ratio of each causative factor, which depends on the LIM in the study area. Therefore, the clustering analysis is also affected by the incomplete LIM. Besides, as the clustering analysis needs to determine the category in advance. To compare this with the method presented in this study, the data sets in the study area were also divided into five categories, and the categories with fewer coseismic landslides were selected as the negative data source. Similarly, there are still a small number of coseismic landslides in the selected category, which affects the evaluation results of the model. Secondly, TSES transforms positive data

FIGURE 10 | ROC curves of the presented method based on three OCCs.

TABLE 6 | The co-seismic landslide and new landslide events in the hazardous area of LSM with different negative data generation methods.

Methods	Hazardous area of LSM	
	Co-seismic landslide (%)	New landslides
SVM-OCSVM	0.92	10
SVM-SOM	0.84	8
SVM-TSES	0.86	6

into negative data by replacing the value of a randomly selected causative factor with a value outside the range. This method is also affected by two factors. The LIM represents the landslide-affected area. On the other hand, the SUs used in this study are small-scale units that cannot accurately reflect the spatial differences of the feature corresponding to coseismic landslides. Therefore, there is still considerable uncertainty in the negative data generation based on inaccurate feature space information. In addition, TSES assumes that the weight of each causative factor is equal, while the existing research shows that the impact of the seismic landslide causative factor is different. Therefore, simply replacing the attribute value of a causative factor cannot generate a reliable negative data set. In summary, the method proposed in this study can effectively obtain the negative data, which has little dependence on the complete LIM and can quickly obtain the landslide susceptibility map after the earthquake.

4.4 DISCUSSION

A reliable and complete LIM is an important data source for rapid assessment of earthquake-induced landslides (Van Westen et al., 2008; Harp et al., 2011). As stated in the introduction, most statistical learning methods for predicting the distribution of landslide-prone areas depend on data sets with both positive (landslide presence) and negative (landslide absence) data (Hong et al., 2019). However, the complete LIM is difficult or impossible to obtain, which hampers the rapid LSM and hazard analysis after an earthquake (Reichenbach et al., 2018). By contrast, incomplete LIM is easy to obtain and is a potential data source for rapid LSM after an earthquake (Xu et al., 2013; Monsieurs et al., 2018; Chen et al., 2020; Du et al., 2020). The results of this paper show that satisfactory results of earthquake-induced LSM can also be obtained based on incomplete LIM provided that the negative data provided by the incomplete LIM is properly processed. Compared with earthquake-induced LSM based on the complete LIM, the approach by using incomplete LIM, proposed in this paper, has the advantages of high efficiency and low economic cost.

This study also finds that the one-class classifier can generate reliable negative data based on incomplete LIM. In this study, a comparative experiment is carried out for the uncertainty of negative data come from incomplete LIM. Compared with using complete LIM, the landslide susceptibility model has higher uncertainty when the negative data come from incomplete LIM, which is basically consistent with the existing research (Huang and Zhao, 2018; Dou

et al., 2020b). However, the negative data generation method based on the one-class classifier can effectively reduce the uncertainty of the negative data and obtain a more stable and reliable model. Incomplete LIM provides negative data that may stem from unrecognized landslides, and false-negative data will reduce the reliability of the landslide susceptibility model. Even within the negative data obtained from complete LIM, there are inevitably unstable areas (Huang et al., 2017; Hong et al., 2019; Zhu et al., 2019), that is, the potential landslide areas in the deformation process. The data for such areas are between the negative data and the positive data, and simply taking them as negative data also increases the uncertainty of the samples. To overcome these challenges, the one-class classifier can construct a screening negative data model based on single class samples, which can eliminate the influence of unrecognized landslide area and potential landslide area and obtain more reliable negative data. Compared with other one-class classifiers, OCSVM can deal with complex nonlinear problems based on small sample data (Schölkopf et al., 2000), which can precisely represent the complex nonlinear relationship between seismic landslides and causative factors.

The performance of machine learning algorithms (e.g., SVM and OCSVM, which are used in this study) varies in accordance with topographic variables, suggesting that the patterns in causative factors are highly complex and variable for the different facets of causative factor attributes (e.g., spatial scale) (Chang et al., 2019). Thus, finding a general approach suitable for all regions and/or topographic variables is unlikely. And instead, evaluation of multiple machine learning algorithms (Merghadi et al., 2020; Wang et al., 2020) and/or ensemble of different machine learning algorithms (Dou et al., 2020a; Pham et al., 2020) should be standard procedure in developing satisfactory LSMs with highly complex landslide causative factors.

Focusing on overcoming the uncertainty of negative data that comes from an incomplete LIM, the method presented in this paper can quickly carry out seismic LSM under data-scarce environments conditions. Inevitably, some open problems require further study. First, we must find out how we can best define an incomplete LIM and explore the relationship between the completeness level of LIM and LSM performance, as these are not discussed in detail in this paper. In addition, the method presented in this paper is mainly based on incomplete LIM and is still a data-driven method. The quality of the results still depends on source data quality (Broeckx et al., 2018). With the rapid development of Synthetic Aperture Radar, 3D deformation information of earthquake-affected area can be obtained quickly after an earthquake, which can be effectively used to identify the hidden landslide area (Xie et al., 2020). More reliable negative data may be obtained by integrating the stable area in 3D deformation and the non-landslide area in LIM. More accurate LSM may be achieved by extending the proposed method based on OCSVM to incorporate such negative data with higher quality.

5 CONCLUSION

This study proposed a negative data generation method based on one class classifier, which can effectively make use of incomplete LIM. Based on the binary classifier established by the generated negative data, we performed LSM for the landslides of the 2008 Wenchuan earthquake. The reliability of the result was analyzed and verified using the coseismic landslides and new landslide events in the study area. The results show the following: 1) the two-class support vector machine is susceptible to the negative data, especially in the absence of complete LIM, and the landslide susceptibility model/mapping obtained from different negative data are quite different; 2) an incomplete LIM is easy to obtain, which is important data for rapid evaluation of landslide susceptibility; 3) the method presented in this study can effectively reduce the uncertainty of negative data and can significantly improve the performance of the landslide susceptibility model even without complete LIM. To sum up, the method in this study can be used to evaluate the landslide susceptibility quickly after the earthquake and provide an important reference for emergency rescue and land planning in post-earthquake regions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ZM, LW, and SC proposed and participated in designing the study; SC, ZM, AZ and QL designed the study, analyzed the data and wrote the manuscript; AZ and YH participated in designing the study and improved the paper; AZ polished the paper. All authors approved the final version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Broeckx, J., Vanmaercke, M., Duchateau, R., and Poesen, J. (2018). A data-based landslide susceptibility map of africa. *Earth-Science Rev.* 185, 102–121. doi:10.1016/j.earscirev.2018.05.002

Cao, J., Zhang, Z., Wang, C., Liu, J., and Zhang, L. (2019). Susceptibility assessment of landslides triggered by earthquakes in the western sichuan plateau. *Catena* 175, 63–76. doi:10.1016/j.catena.2018.12.013

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

Chang, K.-T., Merghadi, A., Yunus, A. P., Pham, B. T., and Dou, J. (2019). Evaluating scale effects of topographic variables in landslide susceptibility models using gis-based machine learning techniques. *Scientific Rep.* 9, 1–21. doi:10.1038/s41598-019-48773-2

Chen, S., Miao, Z., Wu, L., and He, Y. (2020). Application of an incomplete landslide inventory and one class classifier to earthquake-induced landslide susceptibility mapping. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 13, 1649–1660. doi:10.1109/jstars.2020.2985088

Chen, W., Yan, X., Zhao, Z., Hong, H., Bui, D. T., and Pradhan, B. (2019). Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive bayes and rbfnetwork models for the long county area (China). *Bull. Eng. Geol. Environ.* 78, 247–266. doi:10.1007/s10064-018-1256-z

Conoscenti, C., Rotigliano, E., Cama, M., Caraballo-Arias, N. A., Lombardo, L., and Agnesi, V. (2016). Exploring the effect of absence selection on landslide susceptibility models: a case study in sicily, Italy. *Geomorphology* 261, 222–235. doi:10.1016/j.geomorph.2016.03.006

Dou, J., Yunus, A. P., Bui, D. T., Merghadi, A., Sahana, M., Zhu, Z., et al. (2020a). Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* 17, 641–658. doi:10.1007/s10346-019-01286-5

Dou, J., Yunus, A. P., Merghadi, A., Shirzadi, A., Nguyen, H., Hussain, Y., et al. (2020b). Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. *Sci. total Environ.* 720, 137320. doi:10.1016/j.scitotenv.2020.137320

Dou, J., Yunus, A. P., Tien Bui, D., Sahana, M., Chen, C.-W., Zhu, Z., et al. (2019). Evaluating gis-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the lidar dem. *Remote Sensing* 11, 638. doi:10.3390/rs11060638

Dreyfus, D., Rathje, E. M., and Jibson, R. W. (2013). The influence of different simplified sliding-block models and input parameters on regional predictions of seismic landslides triggered by the northridge earthquake. *Eng. Geology.* 163, 41–54. doi:10.1016/j.enggeo.2013.05.015

Du, J., Glade, T., Woldai, T., Chai, B., and Zeng, B. (2020). Landslide susceptibility assessment based on an incomplete landslide inventory in the jilong valley, tibet, Chinese himalayas. *Eng. Geology.* 270, 105572. doi:10.1016/j.enggeo.2020.105572

Erener, A., and Düzgün, H. S. B. (2012). Landslide susceptibility assessment: what are the effects of mapping unit and mapping method?. *Environ. Earth Sci.* 66, 859–877. doi:10.1007/s12665-011-1297-0

Fan, X., Scaringi, G., Xu, Q., Zhan, W., Dai, L., Li, Y., et al. (2018). Coseismic landslides triggered by the 8th August 2017 Ms 7.0 Jiuzhaigou earthquake (Sichuan, China): factors controlling their spatial distribution and implications for the seismogenic blind fault identification. *Landslides* 15, 967–983. doi:10.1007/s10346-018-0960-x

Fung, T. (1990). An assessment of tm imagery for land-cover change detection. *IEEE Trans. Geosci. remote sensing* 28, 681–684. doi:10.1109/igarss.1989.577758

Gorum, T., Fan, X., van Westen, C. J., Huang, R. Q., Xu, Q., Tang, C., et al. (2011). Distribution pattern of earthquake-induced landslides triggered by the 12 may 2008 wenchuan earthquake. *Geomorphology* 133, 152–167. doi:10.1016/j.geomorph.2010.12.030

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

Harp, E. L., Keefer, D. K., Sato, H. P., and Yagi, H. (2011). Landslide inventories: the essential part of seismic landslide hazard analyses. *Eng. Geology.* 122, 9–21. doi:10.1016/j.enggeo.2010.06.013

Hong, H., Miao, Y., Liu, J., and Zhu, A.-X. (2019). Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. *Catena* 176, 45–64. doi:10.1016/j.catena.2018.12.035

Hong, H., Pradhan, B., Bui, D. T., Xu, C., Youssef, A. M., and Chen, W. (2017). Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at suichuan area (China). *Geomatics, Nat. Hazards Risk* 8, 544–569. doi:10.1080/19475705.2016.1250112

Huang, F., Yin, K., Huang, J., Gui, L., and Wang, P. (2017). Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Eng. Geology.* 223, 11–22. doi:10.1016/j.enggeo.2017.04.013

Huang, R., and Li, W. (2014). Post-earthquake landsliding and long-term impacts in the wenchuan earthquake area, China. *Eng. Geology.* 182, 111–120. doi:10.1016/j.enggeo.2014.07.008

Huang, R. Q., and Li, W. L. (2009). Analysis of the geo-hazards triggered by the 12 may 2008 wenchuan earthquake, China. *Bull. Eng. Geol. Environ.* 68, 363–371. doi:10.1007/s10064-009-0207-0

Huang, Y., and Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. *Catena* 165, 520–529. doi:10.1016/j.catena.2018.03.003

Irigaray, C., Fernández, T., El Hamdouni, R., and Chacón, J. (2007). Evaluation and validation of landslide-susceptibility maps obtained by a gis matrix method: examples from the betic cordillera (southern Spain). *Nat. Hazards* 41, 61–79. doi:10.1007/s11069-006-9027-8

Jibson, R. W., Harp, E. L., and Michael, J. A. (2000). A method for producing digital probabilistic seismic landslide hazard maps. *Eng. Geology.* 58, 271–289. doi:10.1016/s0013-7952(00)00039-9

Jibson, R. W. (1993). Predicting earthquake-induced landslide displacements using newmark's sliding block analysis. *Transportation Res. Rec.* 1411, 9–17.

Kamp, U., Growley, B. J., Khattak, G. A., and Owen, L. A. (2008). Gis-based landslide susceptibility mapping for the 2005 kashmir earthquake region. *Geomorphology* 101, 631–642. doi:10.1016/j.geomorph.2008.03.003

Kargel, J. S., Leonard, G. J., Shugar, D. H., Haritashya, U. K., Bevington, A., Fielding, E. J., et al. (2016). Geomorphic and geologic controls of geohazards induced by nepals 2015 gorkha earthquake. *Science* 351, 8353. doi:10.1126/science.aac8353

Keefer, D. K. (2002). Investigating landslides caused by earthquakes–a historical review. *Surv. Geophys.* 23, 473–510. doi:10.1023/a:1021274710840

Li, W.-l., Huang, R.-q., Tang, C., Xu, Q., and van Westen, C. (2013). Co-seismic landslide inventory and susceptibility mapping in the 2008 wenchuan earthquake disaster area, China. *J. Mt. Sci.* 10, 339–354. doi:10.1007/s11629-013-2471-5

Li, X., Zhou, Z., Yu, H., Wen, R., Lu, D., Huang, M., et al. (2008). Strong motion observations and recordings from the great wenchuan earthquake. *Earthquake Eng. Eng. Vibration* 7, 235–246. doi:10.1007/s11803-008-0892-x

Lv, Z. Y., Shi, W., Zhang, X., and Benediktsson, J. A. (2018). Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 11, 1520–1532. doi:10.1109/jstars.2018.2803784

Ma, S., and Xu, C. (2019). Assessment of co-seismic landslide hazard using the Newmark model and statistical analyses: a case study of the 2013 Lushan, China, Mw6.6 earthquake. *Nat. Hazards* 96, 389–412. doi:10.1007/s11069-018-3548-9

Maharaj, R. J. (1993). Landslide processes and landslide susceptibility analysis from an upland watershed: a case study from st. andrew, Jamaica, west indies. *Eng. Geology.* 34, 53–79. doi:10.1016/0013-7952(93)90043-c

Mandal, B., and Mandal, S. (2018). Analytical hierarchy process (ahp) based landslide susceptibility mapping of lish river basin of eastern darjeeling himalaya, India. *Adv. Space Res.* 62, 3114–3132. doi:10.1016/j.asr.2018.08.008

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., et al. (2020). Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. *Earth-Science Rev.* 207, 103225. doi:10.1016/j.earscirev.2020.103225

Monsieurs, E., Jacobs, L., Michellier, C., Basimike Tchangaboba, J., Ganza, G. B., Kervyn, F., et al. (2018). Landslide inventory for hazard assessment in a data-poor context: a regional-scale approach in a tropical african environment. *Landslides* 15, 2195–2209. doi:10.1007/s10346-018-1008-y

Pham, B. T., Nguyen-Thoi, T., Qi, C., Phong, T. V., Dou, J., Ho, L. S., et al. (2020). Coupling rbf neural network with ensemble learning techniques for landslide susceptibility mapping. *Catena* 195, 104805. doi:10.1016/j.catena.2020.104805

Pourghasemi, H. R., Yansari, Z. T., Panagos, P., and Pradhan, B. (2018). Analysis and evaluation of landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). *Arabian J. Geosciences* 11, 193. doi:10.1007/s12517-018-3531-5

Pradhan, B., and Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* 25, 747–759. doi:10.1016/j.envsoft.2009.10.016

Ramos-Bernal, R., Vázquez-Jiménez, R., Romero-Calcerrada, R., Arrogante-Funes, P., and Novillo, C. (2018). Evaluation of unsupervised change detection methods applied to landslide inventory mapping using aster imagery. *Remote Sensing* 10, 1987. doi:10.3390/rs10121987

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Rev.* 180, 60–91. doi:10.1016/j.earscirev.2018.03.001

Ruff, M., and Czurda, K. (2008). Landslide susceptibility analysis with a heuristic approach in the eastern alps (vorarlberg, Austria). *Geomorphology* 94, 314–324. doi:10.1016/j.geomorph.2006.10.032

Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., and Platt, J. C. (2000). "Support vector method for novelty detection," in *Advances in neural information processing systems*, 582–588.

Su, Q., Zhang, J., Zhao, S., Wang, L., Liu, J., and Guo, J. (2017). Comparative assessment of three nonlinear approaches for landslide susceptibility mapping in a coal mine area. *Ijgi* 6, 228. doi:10.3390/ijgi6070228

Sun, X., Chen, J., Han, X., Bao, Y., Zhan, J., and Peng, W. (2020). Application of a gis-based slope unit method for landslide susceptibility mapping along the rapidly uplifting section of the upper jinsha river, south-western China. *Bull. Eng. Geol. Environ.* 79, 533–549. doi:10.1007/s10064-019-01572-5

Suykens, J. A. K., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Process. Lett.* 9, 293–300. doi:10.1023/a:1018628609742

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293. doi:10.1126/science.3287615

Tang, C., Zhu, J., and Qi, X. (2011). Landslide hazard assessment of the 2008 wenchuan earthquake: a case study in beichuan area. *Can. Geotech. J.* 48, 128–145. doi:10.1139/T10-059

Tanyaş, H., Van Westen, C. J., Allstadt, K. E., Anna Nowicki Jessee, M., Görüm, T., Jibson, R. W., et al. (2017). Presentation and analysis of a worldwide database of earthquake-induced landslide inventories. *J. Geophys. Res. Earth Surf.* 122, 1991–2015. doi:10.1002/2017JF004236

Tsai, H.-Y., Tsai, C.-C., and Chang, W.-C. (2019). Slope unit-based approach for assessing regional seismic landslide displacement for deep and shallow failure. *Eng. Geology.* 248, 124–139. doi:10.1016/j.enggeo.2018.11.015

Van Den Eeckhaut, M., Reichenbach, P., Guzzetti, F., Rossi, M., and Poesen, J. (2009). Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the flemish ardennes, Belgium. *Nat. Hazards Earth Syst. Sci.* 9, 507–521. doi:10.5194/nhess-9-507-2009

Van Westen, C. J., Castellanos, E., and Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. *Eng. Geology.* 102, 112–131. doi:10.1016/j.enggeo.2008.03.010

Wang, F., Xu, P., Wang, C., Wang, N., and Jiang, N. (2017). Application of a gis-based slope unit method for landslide susceptibility mapping along the longzi river, southeastern Tibetan plateau, China. *Ijgi* 6, 172. doi:10.3390/ijgi6060172

Wang, W.-D., He, Z., Han, Z., and Li, Y. (2020). "Landslide susceptibility mapping based on the deep belief network: a case study in sichuan province, China," in *Workshop on world landslide forum* (Springer), 201–213.

Wu, R., Zhang, Y., Guo, C., Yang, Z., Tang, J., and Su, F. (2020). Landslide susceptibility assessment in mountainous area: a case study of sichuan–tibet railway, China. *Environ. Earth Sci.* 79, 1–16. doi:10.1007/s12665-020-8878-8

Xiao, C., Tian, Y., Shi, W., Guo, Q., and Wu, L. (2010). A new method of pseudo absence data generation in landslide susceptibility mapping with a case study of shenzhen. *Sci. China Technol. Sci.* 53, 75–84. doi:10.1007/s11431-010-3219-x

Xie, M., Zhao, W., Ju, N., He, C., Huang, H., and Cui, Q. (2020). Landslide evolution assessment based on insar and real-time monitoring of a large reactivated landslide, wenchuan, China. *Eng. Geology.* 277, 105781. doi:10.1016/j.enggeo.2020.105781

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

Xu, C., Xu, X., Yao, X., and Dai, F. (2014). 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* 11, 441–461. doi:10.1007/s10346-013-0404-6

Yunus, A. P., Fan, X., Tang, X., Jie, D., Xu, Q., and Huang, R. (2020). Decadal vegetation succession from modis reveals the spatio-temporal evolution of post-seismic landsliding after the 2008 wenchuan earthquake. *Remote Sensing Environ.* 236, 111476. doi:10.1016/j.rse.2019.111476

Zhou, S., Chen, G., and Fang, L. (2016). Distribution pattern of landslides triggered by the 2014 ludian earthquake of China: implications for regional threshold topography and the seismogenic fault identification. *Ijgi* 5, 46. doi:10.3390/ijgi5040046

Zhu, A.-X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T., et al. (2019). A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. *Catena* 183, 104188. doi:10.1016/j.catena.2019.104188