Article

A Novel Ensemble Approach for Landslide Susceptibility Mapping (LSM) in Darjeeling and Kalimpong Districts, West Bengal, India

Jagabandhu Roy [1]**, Sunil Saha** [1]**, Alireza Arabameri** [2,*]**, Thomas Blaschke** [3] **and Dieu Tien Bui** [4,*]

[1] Department of Geography, University of GourBanga, West Bengal 732103, India; jagabandhuroy1991@gmail.com (J.R.); sunilgeo.88@gmail.com (S.S.)
[2] Department of Geomorphology, Tarbiat Modares University, Tehran 14115-111, Iran
[3] Department of Geoinformatics – Z_GIS, University of Salzburg, 5020 Salzburg, Austria; thomas.blaschke@sbg.ac.at
[4] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
*   Correspondence: alireza.arab@ut.ac.ir (A.A.); buitiendieu@duytan.edu.vn (D.T.B.)

Abstract: Landslides are among the most harmful natural hazards for human beings. This study aims to delineate landslide hazard zones in the Darjeeling and Kalimpong districts of West Bengal, India using a novel ensemble approach combining the weight-of-evidence (WofE) and support vector machine (SVM) techniques with remote sensing datasets and geographic information systems (GIS). The study area currently faces severe landslide problems, causing fatalities and losses of property. In the present study, the landslide inventory database was prepared using Google Earth imagery, and a field investigation carried out with a global positioning system (GPS). Of the 326 landslides in the inventory, 98 landslides (30%) were used for validation, and 228 landslides (70%) were used for modeling purposes. The landslide conditioning factors of elevation, rainfall, slope, aspect, geomorphology, geology, soil texture, land use/land cover (LULC), normalized differential vegetation index (NDVI), topographic wetness index (TWI), sediment transportation index (STI), stream power index (SPI), and seismic zone maps were used as independent variables in the modeling process. The weight-of-evidence and SVM techniques were ensembled and used to prepare landslide susceptibility maps (LSMs) with the help of remote sensing (RS) data and geographical information systems (GIS). The landslide susceptibility maps (LSMs) were then classified into four classes; namely, low, medium, high, and very high susceptibility to landslide occurrence, using the natural breaks classification methods in the GIS environment. The very high susceptibility zones produced by these ensemble models cover an area of 630 km^2 (WofE& RBF-SVM), 474 km^2 (WofE& Linear-SVM), 501km^2 (WofE& Polynomial-SVM), and 498 km^2 (WofE& Sigmoid-SVM), respectively, of a total area of 3914 km^2. The results of our study were validated using the receiver operating characteristic (ROC) curve and quality sum (Qs) methods. The area under the curve (AUC) values of the ensemble WofE& RBF-SVM, WofE & Linear-SVM, WofE & Polynomial-SVM, and WofE & Sigmoid-SVM models are 87%, 90%, 88%, and 85%, respectively, which indicates they are very good models for identifying landslide hazard zones. As per the results of both validation methods, the WofE & Linear-SVM model is more accurate than the other ensemble models. The results obtained from this study using our new ensemble methods can provide proper and significant information to decision-makers and policy planners in the landslide-prone areas of these districts.

Keywords: landslide; machine learning models; remote sensing; ensemble models; validation

1. Introduction

Mountainous regions are threatened by the common natural disaster of landslides. Like hurricanes, floods, droughts, earthquakes, soil erosion, and tsunamis, landslides are important environmental disasters which cause damage and destruction to residential areas, roads, agricultural fields, gardens, and grasslands. Spatially predicting landslide-prone areas may play an important role in disaster management, and it can be considered the standard tool for decision-making in different areas [1]. In geological engineering, landslides are defined as the downward movement of material mass on a slope [2]. Worldwide, mountainous areas are profoundly affected by landslides due to the instability of slopes and masses [3]. For example, the Indian Himalayan mountain regions such as Jammu & Kasmir, Himachal Himalaya, Kumayun, Darjeeling, Sikkim, and north-eastern hilly regions are severely affected by landslides [4]. In the Darjeeling Himalayan region, landslides have a severe environmental impact on socio-economic development. Every year the Darjeeling and Kalimpong districts are frequently heated by landslides due to heavy monsoon rainfall and seismic activity [5]. During July–August in 1993, May in 2009, and September in 2011, the Darjeeling and Kalimpong districts were severely affected by extreme landslides [6]. Furthermore, some major towns in these districts, such as Darjeeling, Mirik, Kurseong, and Kalimpong were hit by landslides during June–July in 2015 due to heavy rainfall, causing fatalities and damage to properties. Therefore, it is necessary to address the landslide risk faced by this particular region to reduce the impact of this environmental disaster. Information regarding the magnitude, character, and probability of landslides can be used to reduce the impact of landslide hazards and for sustainable environmental development and future planning [7]. Therefore, landslide potentiality zoning is an important step for sustainable land management, not only for this particular region but also for other mountainous regions all over the world. The chance of landslide occurrence depends on various conditioning factors rather than a single factor. For preparing the landslide susceptibility map, two things are important: Firstly, landslide inventory data which is considered the dependent variable, and, secondly, landslide conditioning factors which are considered independent variables. In this study, the landslide inventory map was prepared using data collected from Google Earth imagery, a global positioning system (GPS), and extensive field investigations. The landslide conditioning factors or environmental factors, such as slope, aspect, altitude, curvature, geology, soil, land use, normalized differential vegetation index (NDVI), distance from drainage, distance from fault, distance from road, topographic wetness index (TWI), and stream power index (SPI) were selected based on the findings of previous literature (Yilmaz [8], Abedini et al. [9], Regmi et al. [10], Chawla et al. [11], Shahabi and hasim [12], Roy and Saha [13], Pradhan [14], Pourghasemi et al. [15], Pham et al. [16], and Goetz et al. [17]). The landslide inventory data and aforementioned landslide conditioning factors were used to prepare the LSMs with the help of the remote sensing data (RS) and geographical information systems (GIS). Nowadays, most researchers argue that machine learning algorithms using remote sensing and geographical information systems are reliable and appropriate methods for assessing landslide hazards. During recent decades, many studies on landslide susceptibility mapping have been conducted in various parts of the world. Researchers have applied different approaches to produce landslide susceptibility maps, such as statistical models, probabilistic models, knowledge-driven models, and machine learning models using geographical information systems and remote sensing techniques like the analytical hierarchy process (AHP) and bivariate statistics [9,18], logistic regression (LR), artificial neural networks (ANN), frequency ratio (FR), naive bayes classifier, auto logistic modeling, static methods, multivariate adaptive regression, two-class kernel logistic regression, SVM, artificial neural network kernel, logistic regression and logistic tree, random forest, and decision tree methods [19]. Ensemble techniques have been shown to achieve better results than a single method. In this article, WofE was ensembled with four kernels (radial basis function (RBF), linear kernel, polynomial kernel, and sigmoid kernel) of SVM to predict probable landslide hazard areas and for a comparison of the results. The Darjeeling and Kalimpong districts are parts of the eastern Himalayan region in India. Both districts are mostly covered in hilly terrain. Every year, these districts are affected by landslides, which cause destruction

to the roads, residential areas, tea gardens, and forests, leading to numerous fatalities. Therefore, these regions were selected as the study area to raise awareness among the public and government so necessary steps can be taken to mitigate the landslide hazard.

2. Materials and Methods

2.1. Study Area

The Darjeeling and Kalimpong districts are situated in the eastern Himalaya region of India and are mainly covered in hilly and rugged mountainous terrain. Combinedly, these districts cover an area of 3914 square km. The research site is bounded within the 26°27″ to 27°13″N latitudes and 87°59″E to 88°53″E longitudes (Figure 1). The altitude of the study area ranges from 15 m to 3602 m above the mean sea level. Climatically, the region is influenced by the south-west and north-east Indian monsoon. The summer season is very wet, and the winter season is dry and cold. The temperature of this region can drop close to zero degrees. According to the Indian metrological department, the rainfall of this region ranges from 1877 mm to 2333 mm. Geologically, the region is composed of Precambrian (Darjeeling gneiss, Daling series), Permian (Damuda series), Miocene (Swaliks), and recent Pleistocene (Alluvium) lithologies, as shown in Table 1 [20]. The Gorubathan and Rangamati surface are tectonic landscape of these regions. The majority of the study area is covered in Triassic rock. Regarding its geomorphology, the research site is composed of active flood plain, alluvial plain, folded ridge, highly dissected hill slope, intermontane valley, and piedmont fan plain [11]. Pedologically, the region is characterized by various soil texture classes; namely, gravelly-loamy, fine-loamy to coarse-loamy, gravelly-loamy to loamy skeletal, and gravelly-loamy to coarse-loamy [21]. Several rivers—namely, the Mahananda, Tista, Mechi, Balason, Jaldhaka, Rammam and Rangit—flow across these districts originating in the mountainous areas. The Darjeeling and Kalimpong districts are famous for national and international tourism. Some attractive tourist places in these regions are Tiger Hill, Rock Garden, Mahakal Temple, Dhirdham Temple, Batasia Loop, Ghoom Monastery, and Happy Valley Tea Garden. The major economic activities of these regions are tea plantation, horticulture, and tourism. The healthy and tasty tea of this region is famous worldwide. Siliguri, Darjeeling, and Kalimpong are the major towns and headquarters within our study area. the total population comprises 18,46,823 people, of which 50.75 % are males, and 49.25% are females. The population density of the research site is 586 people/km^2, which is comparatively higher than the mean Indian population density [22]. The length of the national highway, state highway, and other main district highway has increased from 100 to 111 km, 80 to 191 km, and 37 to 79 km from 2001 to 2011. Different cultural communities are present in the study area, such as Nepali, Lepcha, Bhutia, and Rai.

Table 1. Geological successions of Darjeeling Himalaya.

Age	Series	Lithological Characteristics
Recent (Holocene) Pleistocene	Sub-aerial formations (soil, alluvia, colluvial) Raised Terraces	Younger flood plain deposits of the rivers composed of sand, gravel, pebble, etc. and soil covering the rocks sandy, clay, gravel, pebble, boulders etc. representing older fluvial deposits
Miocene	Siwalik	Micaceous sandstones with slaty bands, seams of graphitic coal, silts and minor bands of limestone
Permian	Damuda Series (Lower Gondwana)	Quartzitic sandstones with slaty bands, carbonaceous shales, seams of graphitic coal, lamprophyre sills and minor bands of limestone
Precambrian	1) Darjeeling gneiss 2) Daling gneiss	Golden-silvery micaschists; Carbonaceousmicaschists; Granatiferousmicaschists and coarse grained gneisses. Slates (greenish to grey with perfect slaty cleavage). Phyllites surrounded by pebbles of quartz, Chlorite-schists with bands of grilty schist's injected with gneiss (crinkled). Granites, pagmatites's and quartz veins, with tourmaline and iron as accessories

Source: Mallet (50); Gansser, (51); Pawde and Saha, (52).
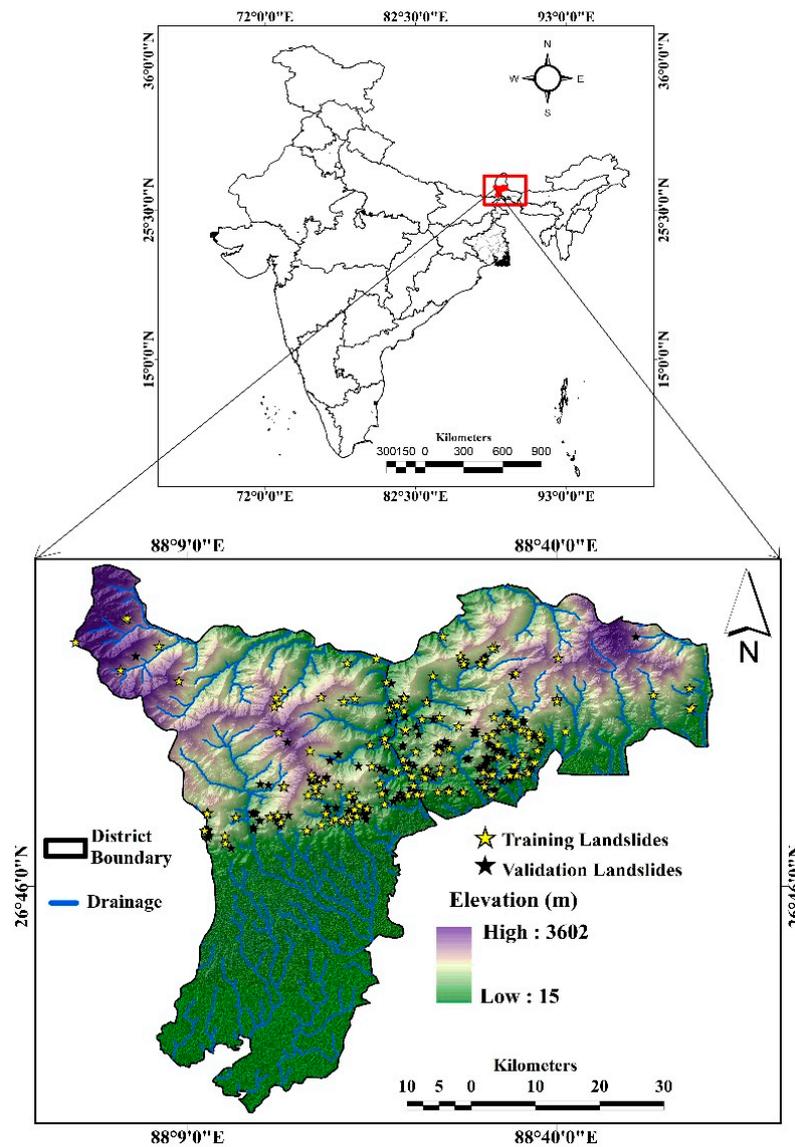
Figure 1. Study area and landslide location map.

2.2. Methodology

The methodology of the present study is depicted in Figure 2. The flowchart is divided into four main steps, as follows. Step 1: Data used: here, the landslide inventory map (LIM) and landslide conditioning factors (LCFs) data layers were prepared. Step 2: Multicollinearity analysis of landslide conditioning factors was carried out. Step 3: New ensembles of weight-of-evidence (WofE) and SVM models were applied to prepare the landslide susceptibility maps (LSMs). Step 4: LSMs were validated using the receiver operating characteristics (ROC) and quality sum (Qs) methods to measure the capability of the models and identify the best suitable model.
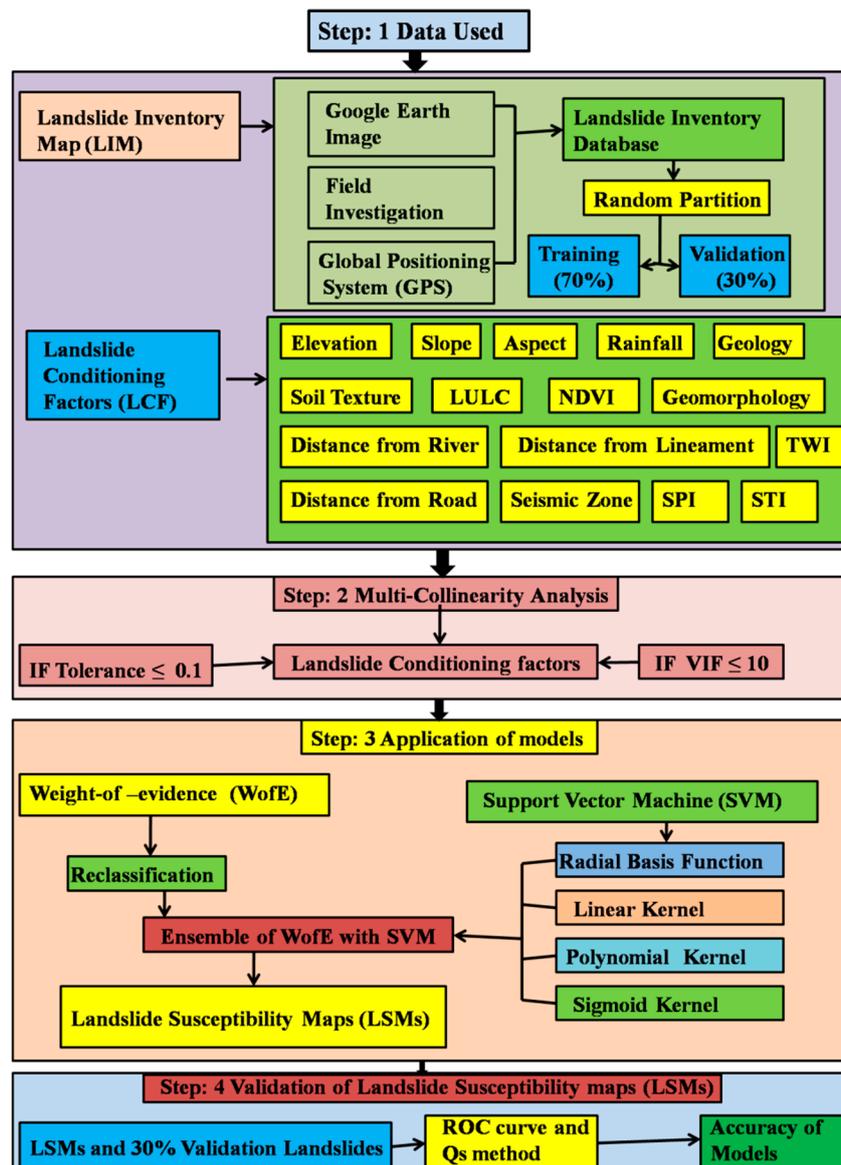
Figure 2. Methodological flowchart of the present work.

2.3. Data Preparation

2.3.1. Landslide Inventory Dataset

It is vital to analyze the landslide distribution and landslide conditioning factors to determine which areas are most at risk of landslide occurrence. The landslide inventory map (LIM) is an important part of the evaluation and assessment of landslide hazards and risks. Some researchers have used landslide inventory datasets for landslide susceptibility mapping [8–17,23]. In the present study, a total of 326 landslides were identified through extensive field investigations using a global positing system (GPS) and Google Earth imagery. Out of 326 landslides, 228 (70%) landslides were chosen randomly for landslide modeling purposes, and 98 (30%) landslides were used to validate the prepared landslide susceptibility maps. The landslide inventory map (LIM) was prepared in a GIS environment and is shown in Figure 1. Field photographs of some landslides in the study area are shown in Figure 3.

Figure 3. Field photographs of some landslides in the study area. (**a**) Sikkim-Kalimpong road (27°03′20″N, 88°26′03″E) (**b**) Sevokekalimandir (26°54′01″N, 88°28′18″E). (**c**) Lish catchment (26°57′N, 88°30′17″E). (**d**) Darjeeling road (26°54′33″N, 88°17′10″E). (**e**) Pagla Jhora (26°52′47.70″N, 88°18′11.24″E). (**f**) Sevoke Road (26°54′33″N, 88°28′04″E).

2.3.2. Preparing Effective Factors

Landslides are processes of mass movement under the influence of different effective factors. Accordingly, it is essential to analyze the conditions of the selected factors to assess landslide susceptibility. The topographic (altitude, slope, aspect), climatic (rainfall), lithological (geology, distance from lineament), hydro-morphological (geomorphology, distance from river, sediment transportation index, stream power index, topographic wetness index), land use, vegetation index, soil texture physical properties, and earthquake intensity are the major effective factors responsible for landslides in general. Previous studies, including Yilmaz [8], Abedini et al. [9], Regmi et al. [10], Chawla et al. [11], Shahabi et al. [12], Roy and Saha [13], Pradhan [14], Pourghasemi et al. [15], Pham et al. [16], and Goetz et al. [17] used these effective factors for landslide susceptibility mapping. In the present study, rainfall (Figure 4d), elevation (Figure 4a), slope (Figure 4b), aspect (Figure 4c),

geology (Figure 4e), soil texture (Figure 4f), distance from river (Figure 4 g) distance from lineament (Figure 4h), distance from road (Figure 4k), geomorphology (Figure 4o), land use/land cover (Figure 4i), normalized differential vegetation index (Figure 4j), topographic wetness index (Figure 4l), sediment transportation index (Figure 4n), stream power index (Figure 4m), and seismic zone (Figure 4p) maps were used to delineate the landslide susceptibility area. Different techniques, which are mentioned in Table 2, were used to prepare the thematic layers of these effective factors. A DEM with a spatial resolution of 30m* 30m was selected to prepare the landslide susceptibility maps, and all of the parameters with scales greater or lesser than the DEM were resampled into 30m*30m resolution.

Slope is one of the main landslide conditioning factors. The spatial distribution of slope ranges from 0 to 89 degrees (Figure 4b). The aspect (Figure 4c) was classified into ten categories, i.e. flat (−1), north (0–22.5; 337.5–360), northeast (22.5–67.5), east (67.5–112.5), southeast (112.5–157.5), south (157.5–202.5), southwest (202.5–247.5), west (247.5–292.5), north-west (292.25–337.5). The altitude of the study area ranges from 15 m to 3602 m above mean sea level (Figure 4a). The spatial distribution of average rainfall ranges from 1877 mm to 2333 mm (Figure 4d). The geological map was obtained from the geological survey of India. The river buffer map was classified into five classes, based on the distance from the river, using the natural breaks classification method. The maximum distance from the river in this study area is 4.33 km (Figure 4g). Similarly, the maximum distances from the road and lineament are 16.4 km (Figure 4k) and 10 km (Figure 4h), respectively. The land use of this study area was classified into five categories; namely, water bodies, settlement, vegetation, tea gardens, fallow land, and agricultural land (Figure 4i). The NDVI values range from −0.072 to 0.432 (Figure 4j). The topographic wetness index (TWI) value of the study area ranges from 1.95 to 18.41 (Figure 4l). Geomorphologically, the research area consists of active flood plain, alluvial plain, folded ridge, highly dissected hill slope, inter mountain valley, and piedmont fan plain (Figure 4o). The seismic map of the study area was classified into two categories; namely, moderate and high seismic zones. The values of the moderate risk zone range from 3 to 5 on the Richter scale, while values above 5 on the Richter scale characterize the high seismic prone areas (Figure 4p). The spatial value of STI ranges from 0 to 203 (Figure 4n). The value of SPI in the study area ranges from −11.0 to 7.81 (Figure 4m).

The elevation, slope, aspect, rainfall, normalized differential vegetation index (NDVI), topographic wetness index (TWI), stream power index (SPI), and sediment transportation index (STI) factors were classified into five sub-layers using the natural breaks classification method in a GIS environment (Figure 4). The land use/land cover (LULC) was determined by the maximum likelihood classification method (Figure 4). The geology, soil texture, geomorphology, and seismic zone maps were categorized into different sub-layers using a general classification technique in a GIS environment (Figure 4).
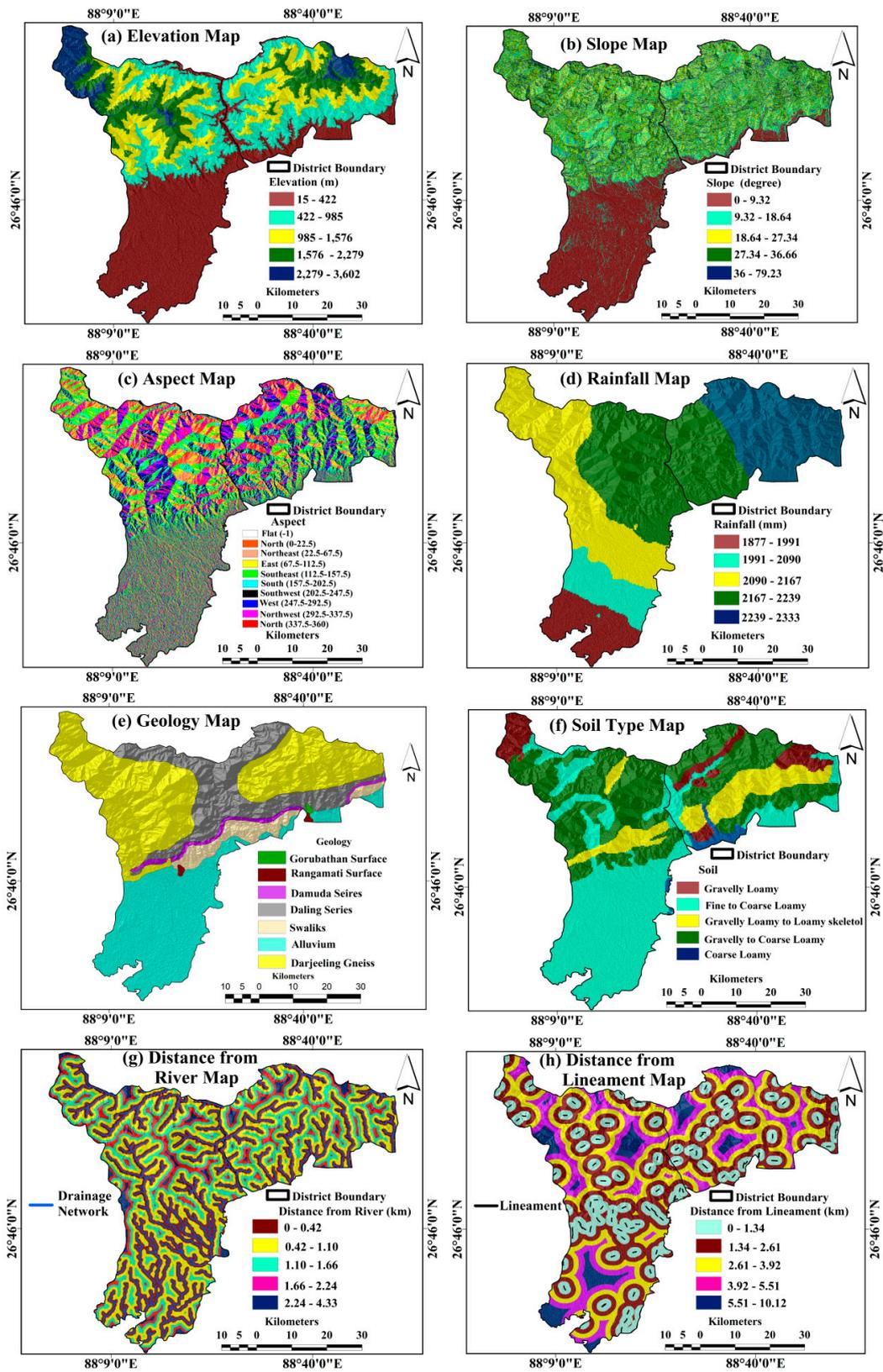
Figure 4. *Cont.*
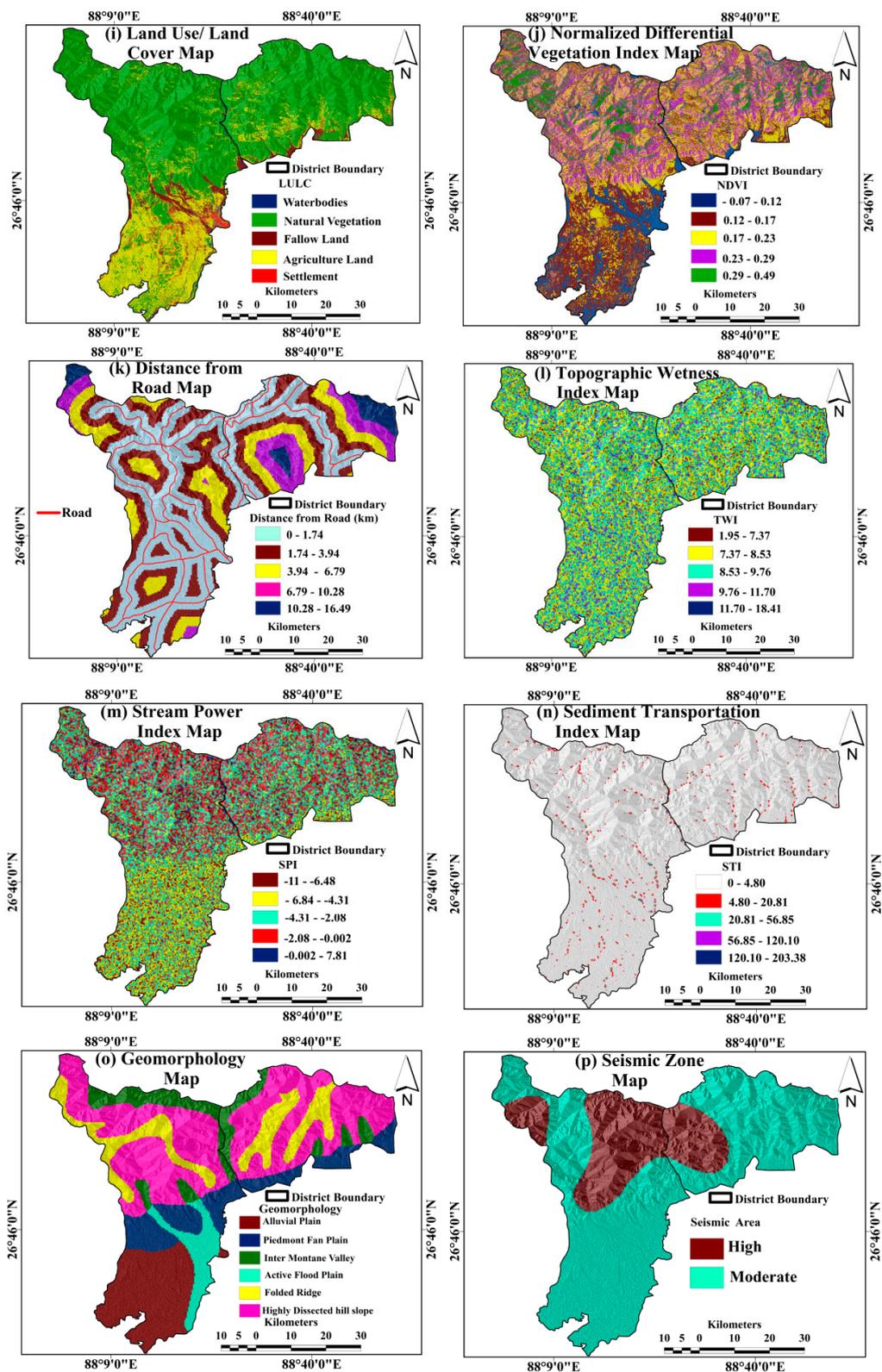
Figure 4. Landslide conditioning factors - **a**. elevation, **b**. slope, **c**. aspect, **d**. rainfall, **e**. geology, **f**. soil texture, **g**. distance from river, **h**. distance from lineament, **i**. land use/land cover (LULC), **j**. normalized differential vegetation index (NDVI), **k**. distance from road, **l**. topographic wetness index (TWI), **m**. stream power index (SPI), **n**. sediment transportation index (STI), **o**. Geomorphology, **p**. Seismic map.

Table 2. Production techniques used for the various thematic data layers.

Sl. No.	Parameters	Data Used & Scale	Sources of Data Types	Techniques	References
1	Elevation	DEM 30 m × 30	U.S Geological Survey	30 m × 30 m digital elevation model	[24]
2	Slope	DEM 30 m × 30	U.S Geological Survey	$\text{Tan}\theta = \frac{N \times i}{636.6}$ N=No. of Contour Cutting; i=Contour Interval	[25]
3	Aspect	DEM 30 m × 30	U.S Geological Survey	$Aspect = 57.29 \times \alpha tan2(\left[\frac{dz}{dy}\right] - [dz/dx]$ Where, dz/dx= ((c+2f+i)−(a+2d+g))/8 dz/dy=((g+2h+i)−(a+2b+c))/8 Here, a to i indicates the cell value of 3*3 window.	[26]
4	Rainfall	Annual average rainfall data of different stations in the last 5 years	Indian Metrological Department (IMD)	Kriging Interpolation method	[27]
5	Geology	Reference geological map 1: 50,000	Geological Survey of India	Digitization process	[28]
6	Soil	Reference district soil map 1: 50,000	National Bureau of Soil Survey and Land Use Planning	Digitization process	[28]
7	Distance from River	Reference Topomap 1: 50,000	Survey of India	Euclidian Distance Buffering	[29]
8	Distance from Lineament	Reference sheet of Lineament 30 m × 30	"https://bhuvan-vec2.nrsc.gov. in/bhuvan/wms"	Euclidian Distance Buffering	[29]
9	Land use/land cover (LULC)	Landsat 8 OLI/TIRS 30 m × 30	U.S Geological Survey	Maximum likelihood Classification	[30]

Table 2. *Cont.*

Sl. No.	Parameters	Data Used & Scale	Sources of Data Types	Techniques	References
10	Normalized differential vegetation index (NDVI)	Landsat 8 OLI/TIRS 30 m × 30	U.S Geological Survey	$NDVI = \frac{NIR-IR}{NIR+IR}$ Where NIR is the near infrared band or band 4 and IR is the infrared band or band 3.	[31]
11	Distance from road	Reference Topomap 1: 50,000	Survey of India	Euclidian Distance Buffering	[29]
12	Topographic wetness index (TWI)	DEM 30 m×30 1: 50,000	U.S Geological Survey	$TWI = In(As/tan\theta)$ Where α is the cumulative upslope area draining through a point (per unit contour length), and β is the slope gradient (in degree).	[32].
13	Stream power index (SPI)	DEM 30 m × 30 1: 50,000	U.S Geological Survey	$SPI = As * tan\beta$ Where A_S is the upstream contributing area and β is the slope gradient (in degrees)	[32].
14	Sediment transportation index (STI)	DEM 30 m × 30	U.S Geological Survey	$STI = (m+1) \times (As/22.13)^m \times \sin(B/0.0896)^n$ Where, As, is the specific catchment area; 'B' is the local slope gradient in degrees; m is usually set to 0.4, 'n', is usually set to 0.0896	[33]
15	Geomorphology	Reference sheet 1: 50,000	"https://bhuvan-vec2.nrsc.gov.in/bhuvan/wms"	Digitization process	[27]
16	Seismic zone map	Last 200 years point data of earthquake 30 m × 30	National Centre for Seismology, New Delhi, India	Gridding and Interpolation (Inverse distance weight method)	[11]

2.4. Multicollinearity Analysis

Multicollinearity analysis is a vital way of identifying and selecting appropriate landslide conditioning factors [13]. In this study, multicollinearity was evaluated through the tolerance value and variance inflation factor (VIF). In normal conditions, tolerance values under 0.10 or VIF values of 10 and above indicate multicollinearity [31–33]. In the present study, the multicollinearity test of landslide conditioning factors was done using SPSS software.

2.5. Models

2.5.1. Weight-of-Evidence (WofE) Model

The present study demonstrates the application of the ensemble WofE and SVM model (a Bayesian probability model) for the assessment of landslide susceptibility in the GIS environment. Two types of data were incorporated in the weight-of-evidence model; namely, the landslide inventory data and landslide conditioning factors. The weights were assigned to each landslide conditioning factor by the weight-of-evidence (WofE) model. This model may be compared to the other statistical methods such as the data-driven model that is generally used for the Bayesian probability model [29,34–40]. Mohammady et al. [38] and Regmi et al. [10] emphasized the value of using the weight-of-evidence model for the evaluation of landslide hazard zones.

The positive weight (W^+) and negative weight (W^-) were calculated to complete the weight-of-evidence function. This calculation was the basis for assigning the weights to the landslide conditioning factors (B) based on the presence and absence of landslides within the area [35] using the following equations (1, 2).

$$W_i^+ = \text{In} \frac{P\{B/A\}}{P\{B/\overline{A}\}} \tag{1}$$

$$W_i^- = \text{In} \frac{P\{\overline{B}/A\}}{P\{\overline{B}/\overline{A}\}} \tag{2}$$

Here, P is the probability and ln is the natural log. Similarly, B and $\overline{B}$ indicate the presence and absence of landslide predictive factors. A and $\overline{A}$ indicate the presence and absence of landslides. A positive weight (W^+) indicates the presence of landslides in a sub-category of landslide conditioning factors and the magnitude of this weight is an indication of the positive correlation between landslide conditioning factors and landslide occurrence. A negative weight (W^-) indicates the absence of landslides in a sub-category of landslide conditioning factors. A negative weight also indicates a negative correlation between the landslide conditioning factors and the occurrence of landslides [36]. For modeling purposes, the weight contrast C ($C = W^+ - W^-$) measures the spatial association between landslide conditioning factors and landslide occurrences. A positive C value indicates a positive spatial association and a negative C value indicates a negative spatial association [37].

The standard deviation of W is calculated using Equation (3):

$$S(C) = \sqrt{S^2W^+ + S^2W^-} \tag{3}$$

where $S(W^+)$ indicates the variance of the positive weights and $S(W^-)$ indicates the variance of the negative weights. The variance of the weights was calculated using the following equation:

$$S^2W^+ = \frac{1}{N\{B \cap A\}} + \frac{1}{B \cap \overline{A}} \tag{4}$$

$$S^2W^- = \frac{1}{N\{\overline{B} \cap A\}} + \frac{1}{\overline{B} \cap \overline{A}} \tag{5}$$

The studentized contrast is the final weight. It is a measure of confidence and is defined as the ratio of the contrast divided by its standard deviation. The studentized contrast serves as an informal test of whether C is significantly different from zero or if the contrast is likely to be "real" [35]. After applying the WofE model, the factor weights calculated by this model were ensembled with the SVM model.

2.5.2. Support Vector Machine (SVM) Model

Among the different machine learning algorithms, SVM is an important supervised learning binary classifier that is based on the structural risk minimization principle [41–44]. This method separates the hyperplane formation from the training dataset. The separating hyperplane is prepared in the original space of n coordinates (xi parameter in vector x) between the points of two distinct classes [43]. The maximum margin of separation between the classes is discovered by SVM and, therefore, builds a classification hyperplane in the center of the maximum margin [14,44]. If a point is located over the hyperplane, it will be classified as +1 and, if not, will be classified as −1. The training points adjoined to the optimal hyperplane are called support vectors. Once the decision surface is acquired, new data can be classified [45] considering a training data set of instance label pairs $(X_i Y_i)$ with $X_i \in R^n$, $Y_i \in \{+1, -1\}$ and $i = 1......, m$. To delineate the landslide susceptibility zones, X represents the vector space that includes rainfall, slope, aspect, elevation, geology, soil texture, land use/land cover, normalized differential vegetation index, distance from river, distance from lineament, distance from road, topographic wetness index, sediment transportation index, stream power index, geomorphology, and the seismic zone map. Meanwhile, the +1 class indicates landslide pixels, whereas the −1 class indicates non-landslide pixels.

The aim of SVM is to find the optimal separating hyperplane that can separate the training dataset into the two classes of landslides and non-landslides {+1, −1}. The separating hyperplane separates data using the following equations:

$$Y_i = (W.X_i + b) \geq 1 - \xi_i \tag{6}$$

where w is a coefficient vector that defines the orientation of the hyperplane in the feature space, b is the offset value of the hyperplane from the origin, and ξ_i represents the weak positive variables [46]. The problem of optimization will be solved through the determination of an optimal hyperplane using Lagrangian multipliers [47].

$$Minimize \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j Y_i Y_j (X_i X_j) \tag{7}$$

$$\text{Subject to } \sum_{i=1}^{n} \alpha_i Y_i = O, \ 0 \leq \alpha_i \leq C, \tag{8}$$

where ai represents the Lagrange multipliers, C is the penalty value, and the slack variables ni allow for penalized constraint violation. The decision function, which will be used for the classification of new data, can then be written as:

$$g(X) = sign(\sum_{i=1}^{n} Y_i \alpha_i X_i + b) \tag{9}$$

If the hyperplane cannot be separated by the linear kernel function, the original input data may be shifted into a high-dimensional feature space through some nonlinear kernel functions. The classification decision function is presented in Equation (10):

$$g(X) = sign(\sum_{i=1}^{n} Y_i \alpha_i K(X_{i,} X_j) + b) \tag{10}$$

where *K(Xi, Xj)* is the kernel function.

Linear kernel (LN), polynomial kernel (PL), radial basis function kernel (RBF), and sigmoid kernel (SIG) are the most popular kernel types for SVM analysis [14]. PL and RBF are called Gaussian kernels, and they are the most commonly used kernels in the literature [43]. To prepare the landslide susceptibility map using SVM, we used the remote sensing (RS) software ENVI 4.7, which is an environment for visualizing images. The ENVI 4.7 SVM classifier has four types of kernels; namely, radial basis function (RBF), linear kernel, polynomial kernel, and sigmoid kernel. The mathematical calculation was carried out as shown in Table 3.

Table 3. SVM kernel types and their equations.

Kernel Types	Equations	Kernel Parameters
Radial Basis Function (RBF)	$K(X_i, X_j) = \exp(-\gamma X_i - X_j^2)$	γ
Linear kernel	$K(X_i X_j) = X_i^T X_j$	—
Polynomial kernel	$K(X_i, X_j) = (-\gamma X_i^T X + 1)^d$	γ, d
Sigmoid kernel	$K(X_i, X_j) = Tanh(-\gamma X_i^T X + 1)^d$	γ

(Source: Tien Bui et al. [46], Yao et al. [43]).

3. Results

3.1. Considering the Multicollinearity Analysis of the Effective Factors

The landslide conditioning factors were tested for multicollinearity. The results show that the lowest tolerance value of landslide conditioning factors is 0.446 for rainfall and the highest tolerance value is 0.824 for slope (Table 4). The highest variance inflation factor (VIF) value is 2.241, and the lowest VIF value is 1.213 (Table 4). However, the tolerance values of landslide conditioning factors are greater than 0.1, and VIF values are less than 0.1 and 10, suggesting that there are no collinearity problems among these factors. Therefore, the selected 16 landslide conditioning factors are suitable and accurate for modeling landslide susceptibility.

Table 4. Multicollinearity analysis of landslide conditioning factors.

Landslide Conditioning Factors	Collinearity Statistics	
	Tolerance	VIF
Rainfall	0.446	2.241
Elevation	0.520	1.924
Slope	0.824	1.213
Aspect	0.672	1.488
Geology	0.688	1.453
Soil	0.756	1.323
Distance from River	0.570	1.753
Distance from lineament	0.773	1.294
Distance from Road	0.499	2.003
Land use/land cover (LULC)	0.754	1.326
Normalized differential vegetation index (NDVI)	0.757	1.320
Topographic wetness index (TWI)	0.677	1.477
Stream power index (SPI)	0.684	1.461
Sediment transportation index (STI)	0.768	1.302
Geomorphology	0.789	1.268
Seismic zone	0.618	1.618

3.2. Relationship Between Landslide Location and Effective Factors

The WofE values of each class of explanatory variables stand for the degree of landslide occurrence (Table 5). The topographic factors of elevation, slope, and aspect are vital factors which determine the landslide susceptibility of an area. Areas in high elevations are more susceptible to landslides compared to lower altitude areas. In the present study, the altitude level between 422 m to 985 m has the highest WofE value, which indicates a high susceptibility to landslides. The other sub-layers of elevation are comparatively less susceptible to landslides. Slope plays a vital role in landslide hazard assessment. When slope stability becomes weak, the tendency of landslide occurrence is very high. Therefore, high slope values indicate a high probability of landslide occurrence. In our study area, the slope sub-class of 36 to 79 degrees is more prone to landslides compared to the other sub-layers of slope because this sub-class of slope has attained the maximum value of the WofE model. Aspect is also correlated with the probability of landslide occurrence. Aspect is the direction that a slope faces. In this study, south facing slopes obtained the maximum WofE value, indicating a high susceptibility to landslides. Heavy rainfall detaches the soil and rock easily, leading to an increased probability of landslide occurrence. The study area is highly influenced by the monsoon rainfall from June to November, during which the tendency of landslide occurrence is very high. The rainfall sub-layer of 2167 mm to 2239 mm attained the highest WofE values and, therefore, has a higher risk of landslides compared to the other sub-layers of rainfall. Regarding the geology, Darjeeling gneiss, daling series, and swaliks geological segments attained the highest WofE values, suggesting the highest risk of landslides. The soil texture is strongly associated with the probability of landslide occurrence. Gravelly-loamy, gravelly-loamy to loamy-skeletal, and coarse-loamy soil texture classes, with WofE values of 23.44, 21.01 and 19.05, respectively, indicate a higher risk of landslide occurrence compared to the other soil texture classes. River proximity also increases the chances of landslide occurrence. Areas nearest to rivers have a higher landslide risk compared to areas in further distance classes. Here, areas in the class of 0 to 1.66 km distance from rivers have a high probability of landslide occurrence with a WofE value of 14.78. Similarly, areas closest roads and lineaments have a high probability of landslide occurrence based on the WofE values. In recent times, land use has had a strong influence on the occurrence of landslides. Our study area is categorized into five land use types; namely, water bodies, settlement, vegetation, fallow land, and agricultural land. The outcome of the WofE model indicates that the fallow land has a higher risk of landslides compared to vegetation and other land uses. High normalized differential vegetation index areas are less prone to landslide occurrence and vise-versa. Here, the −0.07 to 0.12 NDVI sub-class with a WofE value of 33.27 is the most critical zone for landslide occurrence. The other sub-layers of NDVI indicate lower probabilities of landslide occurrence. For the factors of TWI, STI, and SPI, the maximum values have the highest probability of landslide occurrence. Geomorphologically, the folded ridge and highly dissected mountain regions have the highest potentiality of landslide occurrence, with WofE values of 15 and 33, respectively. Comparatively, the hilly and mountainous regions are more prone to landslides than the plain and plateau regions. Seismologically, the high seismic zone is more susceptible to landslide occurrence than the low seismic zone.

All sub-layers of the different landslide conditioning factors were assigned a weight by the WofE model in the GIS environment. The weighted layers were then converted to a raster layer to prepare the landslide susceptibility map. Before the landslide susceptibility mapping, the weighted (by WofE) layers were reclassified as the input data layers of the support vector machine (SVM) for ensembling with WofE.

Table 5. Spatial relationship between landslide conditioning factors and landslide occurrence extracted by the Weight-of-evidence (WofE) model.

Rainfall (mm)	Pixels	% of Pixels	Landslide Pixels	% of Pixels	W^+	W^-	C	$S2W^+$	$S2W^-$	S©	W
1877.38–1991.97	322590	8.784	0	0.000	0.000	0.000	0.092	0.000	0.000	0.000	0.000
1991.97–2090.54	289906	7.894	0	0.000	0.000	0.000	0.082	0.000	0.000	0.000	0.000
2090.45–2167.44	944320	25.712	393	7.895	−1.182	0.215	−1.397	0.003	0.000	0.053	−26.580
2167.44–2239.06	1333493	36.309	3670	73.684	0.709	−0.885	1.594	0.000	0.001	0.032	49.504
2239.06–2333.96	782357	21.302	918	18.421	−0.145	0.036	−0.182	0.001	0.000	0.037	−4.963
Slope (Degree)											
0–9.32	1175818	32.015	92	1.847	−2.854	0.368	−3.222	0.011	0.000	0.105	−30.614
9.32–18.64	665098	18.109	571	11.464	−0.458	0.078	−0.536	0.002	0.000	0.044	−12.044
18.44–27.34	813896	22.161	1172	23.529	0.060	−0.018	0.078	0.001	0.000	0.033	2.326
27.34–36.66	694449	18.909	1579	31.700	0.518	−0.172	0.690	0.001	0.000	0.030	22.622
36.66–79.23	323404	8.806	1567	31.460	1.277	−0.286	1.563	0.001	0.000	0.031	51.122
Altitude(m)											
15–422	1351511	36.799	417	8.373	−1.482	0.372	−1.854	0.002	0.000	0.051	−36.226
422 – 985	837224	22.796	2491	50.000	0.787	−0.435	1.222	0.000	0.000	0.028	43.079
985 –1576	738499	20.108	1005	20.173	0.003	−0.001	0.004	0.001	0.000	0.035	0.115
1576 – 2279	518669	14.122	839	16.844	0.176	−0.032	0.209	0.001	0.000	0.038	5.509
2279 – 3602	226762	6.174	230	4.610	−0.293	0.017	−0.309	0.004	0.000	0.068	−4.572
Aspect											
Flat(−1)	1905	0.052	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
north	236967	6.452	39	0.788	−2.104	0.059	−2.163	0.025	0.000	0.160	−13.495
northeast	462023	12.580	363	7.289	−0.546	0.059	−0.605	0.003	0.000	0.055	−11.098
east	454970	12.388	651	13.061	0.053	−0.008	0.061	0.002	0.000	0.042	1.443
southeast	522200	14.219	1098	22.045	0.439	−0.096	0.535	0.001	0.000	0.034	15.640
south	525807	14.317	1292	25.946	0.596	−0.146	0.742	0.001	0.000	0.032	22.922
southwest	457236	12.450	890	17.868	0.362	−0.064	0.426	0.001	0.000	0.037	11.505
west	378362	10.302	462	9.279	−0.105	0.011	−0.116	0.002	0.000	0.049	−2.376
northwest	419573	11.424	154	3.093	−1.308	0.090	−1.398	0.006	0.000	0.082	−17.074
north	213621	5.817	31	0.630	−2.223	0.054	−2.277	0.032	0.000	0.179	−12.718

Table 5. *Cont.*

Rainfall (mm)	Pixels	% of Pixels	Landslide Pixels	% of Pixels	W$^+$	W$^-$	C	S2W$^+$	S2W$^-$	S©	W
Geology											
Swaliks	1936266	52.721	3182	63.889	0.192	−0.270	0.462	0.000	0.001	0.030	15.659
Darjeeling Gneiss	270526	7.366	692	13.889	0.635	−0.073	0.709	0.001	0.000	0.041	17.273
Daling series	131471	3.580	415	8.333	0.847	−0.051	0.897	0.002	0.000	0.051	17.480
Alluvium	678512	18.475	0	0.000	0.000	0.205	0.000	0.000	0.000	0.000	0.000
Damuda series	655890	17.859	692	13.889	−0.252	0.047	−0.299	0.001	0.000	0.041	−7.293
Soil											
Gravelly-loamy	274651	7.478	830	16.667	0.803	−0.105	0.908	0.001	0.000	0.038	23.845
Fine loamy_Coarse Loamy	1477848	40.239	1107	22.222	−0.594	0.264	−0.858	0.001	0.000	0.034	−25.171
Gravelly loamy_LoamySkeletol	450035	12.254	1107	22.222	0.596	−0.121	0.717	0.001	0.000	0.034	21.019
Gravelly Loamy_Coarse Loamy	1404794	38.250	1660	33.333	−0.138	0.077	−0.214	0.001	0.000	0.030	−7.131
Coarse Loamy	65336	1.779	277	5.556	1.142	−0.039	1.181	0.004	0.000	0.062	19.055
Distance from River (km)											
0–0.42	1160959	31.611	1049	21.053	−0.407	0.144	−0.551	0.001	0.000	0.035	−15.837
0.42–1.10	1291696	35.171	1966	39.474	0.116	−0.069	0.184	0.001	0.000	0.029	6.356
1.10–1.66	750401	20.432	1442	28.947	0.349	−0.113	0.462	0.001	0.000	0.031	14.784
1.66–2.26	371677	10.120	393	7.895	−0.249	0.024	−0.273	0.003	0.000	0.053	−5.195
2.26–4.33	97931	2.666	131	2.632	−0.013	0.000	−0.014	0.008	0.000	0.089	−0.153
Distance from Lineament(km)											
0–1.54	763490	20.788	906	18.182	−0.134	0.032	−0.167	0.001	0.000	0.037	−4.531
1.54–2.85	1093457	29.773	1019	20.455	−0.376	0.125	−0.501	0.001	0.000	0.035	−14.243
2.85–4.20	941314	25.630	1472	29.545	0.142	−0.054	0.197	0.001	0.000	0.031	6.323
4.20–5.75	633142	17.239	1245	25.000	0.372	−0.099	0.471	0.001	0.000	0.033	14.378
5.75–10.12	241263	6.569	340	6.818	0.037	−0.003	0.040	0.003	0.000	0.056	0.710

Table 5. *Cont.*

Rainfall (mm)	Pixels	% of Pixels	Landslide Pixels	% of Pixels	W$^+$	W$^-$	C	S2W$^+$	S2W$^-$	S©	W
Distance from Road(km)											
0–1.74	1636028	44.546	792	15.909	−1.031	0.417	−1.448	0.001	0.000	0.039	−37.353
1.74–3.94	988335	26.911	906	18.182	−0.393	0.113	−0.506	0.001	0.000	0.037	−13.754
3.94–6.72	589253	16.044	906	18.182	0.125	−0.026	0.151	0.001	0.000	0.037	4.109
6.72–10.22	316628	8.621	1472	29.545	1.235	−0.260	1.495	0.001	0.000	0.031	48.066
10.22–16.49	142420	3.878	906	18.182	1.550	−0.161	1.711	0.001	0.000	0.037	46.466
Land use/Land cover											
Water bodies	40427	1.101	0	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000
Vegetation	2650294	72.163	1119	22.464	−1.168	1.027	−2.195	0.001	0.000	0.034	−64.607
Fallow land	168382	4.585	1624	32.609	1.970	−0.348	2.318	0.001	0.000	0.030	76.445
Agricultural land	763256	20.782	2238	44.928	0.773	−0.364	1.137	0.000	0.000	0.029	39.858
Settlement	50306	1.370	0	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000
Normalized differential vegetation index (NDVI)											
−0.07–0.12	442450	12.047	1399	28.093	0.849	−0.202	1.050	0.001	0.000	0.032	33.271
0.12–0.17)	972514	26.480	1421	28.523	0.074	−0.028	0.103	0.001	0.000	0.031	3.270
0.17–0.23)	997257	27.154	1312	26.336	−0.031	0.011	−0.042	0.001	0.000	0.032	−1.297
0.23–0.29	816592	22.234	618	12.411	−0.584	0.119	−0.703	0.002	0.000	0.043	−16.346
0.29–0.49	443851	12.085	231	4.636	−0.959	0.081	−1.041	0.004	0.000	0.067	−15.436
Topographic wetness index (TWI)											
1.95–7.37	582990	15.874	918	18.421	0.149	−0.031	0.180	0.001	0.000	0.037	4.916
7.37–8.53	1326854	36.128	2097	42.105	0.153	−0.098	0.252	0.000	0.000	0.029	8.765
8.53–9.76	1088701	29.643	1311	26.316	−0.119	0.046	−0.165	0.001	0.000	0.032	−5.140
9.76–11.70	547267	14.901	655	13.158	−0.125	0.020	−0.145	0.002	0.000	0.042	−3.454
11.70–18.91	126853	3.454	0	0.000	0.000	0.000	0.035	0.000	0.000	0.000	0.000

Table 5. *Cont.*

Rainfall (mm)	Pixels	% of Pixels	Landslide Pixels	% of Pixels	W⁺	W⁻	C	S2W⁺	S2W⁻	S©	W
				Sediment transportation index (STI)							
0–4.80	3576809	97.390	4850	97.368	0.000	0.008	−0.008	0.000	0.008	0.089	−0.096
4.80–20.81	78362	2.134	131	2.632	0.210	−0.005	0.215	0.008	0.000	0.089	2.429
20.81–56.85	13728	0.374	0	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
56.85–120.10	3037	0.083	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
120.10–203.38	729	0.020	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
				Stream power index (SPI)							
−11.16 − −6.84	457701	12.462	427	8.571	−0.375	0.044	−0.418	0.002	0.000	0.051	−8.260
−6.84 − −4.31	670452	18.255	1139	22.857	0.225	−0.058	0.283	0.001	0.000	0.034	8.385
−4.31 − −2.08	994622	27.082	1139	22.857	−0.170	0.056	−0.226	0.001	0.000	0.034	−6.700
−2.08 − −0.002	1003492	27.323	1423	28.571	0.045	−0.017	0.062	0.001	0.000	0.031	1.978
−0.002 − 7.81	546398	14.877	854	17.143	0.142	−0.027	0.169	0.001	0.000	0.038	4.491
				Geomorphology							
Alluvial plain	591694	16.111	0	0.000	0.000	0.176	0.000	0.000	0.000	0.000	0.000
Piedmont fan plain	453016	12.335	119	2.381	−1.646	0.108	−1.754	0.008	0.000	0.093	−18.867
Inter montane valley	383190	10.434	474	9.524	−0.091	0.010	−0.101	0.002	0.000	0.048	−2.101
Active flood plain	205950	5.608	0	0.000	0.000	0.058	0.000	0.000	0.000	0.000	0.000
Folded ridge	499607	13.603	1067	21.429	0.455	−0.095	0.550	0.001	0.000	0.035	15.919
Highly dissected hill slope	1539208	41.910	3321	66.667	0.465	−0.556	1.021	0.000	0.001	0.030	33.948
				Seismic zone map							
High	1000641	27.246	2604	52.273	0.653	−0.422	1.075	0.000	0.000	0.028	37.859
Moderate	2672024	72.754	2377	47.727	−0.422	0.653	−1.075	0.000	0.000	0.028	−37.859

3.3. Landslide Susceptibility Models

The support vector machine is an important machine learning algorithm that is used to assess an area's susceptibility to landslides and other natural hazards. In the present study, the SVM classification was used and ensembled with WofE. The landslide conditioning factors; namely, elevation, slope, aspect, rainfall, geology, soil texture, land use land cover, normalized differential vegetation index (NDVI), distance from river, distance from road, distance from lineament, topographic wetness index (TWI), stream power index (SPI), sediment transportation index (STI), geomorphology, and seismic zone map were used as the input of the SVM classification. The probability values of the SVM classification ranges from 0 to 1. Pixels of images or conditioning factors indicate the landslide susceptibility index with two values, i.e., 0 to 1 where 0 represents stable conditions and 1 value indicates a high chance of landslides occurrence. The SVM classification has four kernel types; namely, radial basis function, linear kernel, polynomial kernel, and sigmoid kernel. These functions were applied in the SVM classification. The output file images created by the SVM classification were integrated and used to prepare the landslide susceptibility maps (LSMs) in the GIS environment.

The four landslide susceptibility maps (LSMs) shown in Figure 5a–d were prepared using the four ensemble models of WofE and SVM; namely, WofE & RBF-SVM, WofE & Linear-SVM, WofE&Polynimial-SVM, and WofE& Sigmoid-SVM. These landslide susceptibility maps (LSMs) were classified into four categories; namely, low, medium, high, and very high susceptibility to landslides, using the natural breaks classification method in the GIS environment. In the WofE& RBF-SVM ensemble map, the four landslide susceptibility classes of low, medium, high, and very high covered 1071 km^2 (34%), 813 km^2 (25.8%), 635 km^2 (20.2%), and 630 km^2(20%) area of the districts, respectively (Table 6 and Figure 6). In the WofE and Linear-SVM model, the low, medium, high, and very high landslide susceptibility classes covered an area of 1128 km^2 (35.8%), 918 km^2 (29.1%), 630 km^2 (20%), and 474 km^2 (15%), respectively (Table 6). In the WofE& Polynomial-SVM model, the low, medium, high, and very high susceptibility classes covered an area of 1095 km^2 (34.8 %), 944 km^2 (30%), 608 km^2 (19.3%) and 501 km^2 (15.9 %), respectively (Table 6). Meanwhile, in the WofE & Sigmoid-SVM ensemble landslide map, the classes of low, medium, high, and very high landslide susceptibility covered 1153 km^2 (36.6%), 893 km^2 (28.3%), 605 km^2 (19.2%) and 398 km^2 (15.8%) of the area, respectively (Table 6).

Table 6. Areal distribution of ensemble model landslide susceptibility maps (LSMs).

| Landslide Susceptibility Classes	WofE& RBF-SVM		WofE&Linear-SVM		WofE& Polynomial-SVM		WofE& Sigmoid-SVM	
	Area in sq.km	% of Area	Area in sq.km	% of Area	Area in sq.km	% of Area	Area in sq.km	% of Area
Low	1071	34.0	1128	35.8	1095	34.8	1153	36.6
Medium	813	25.8	918	29.1	944	30.0	893	28.3
High	635	20.2	630	20.0	608	19.3	605	19.2
Very High	630	20.0	474	15.0	501	15.9	498	15.8
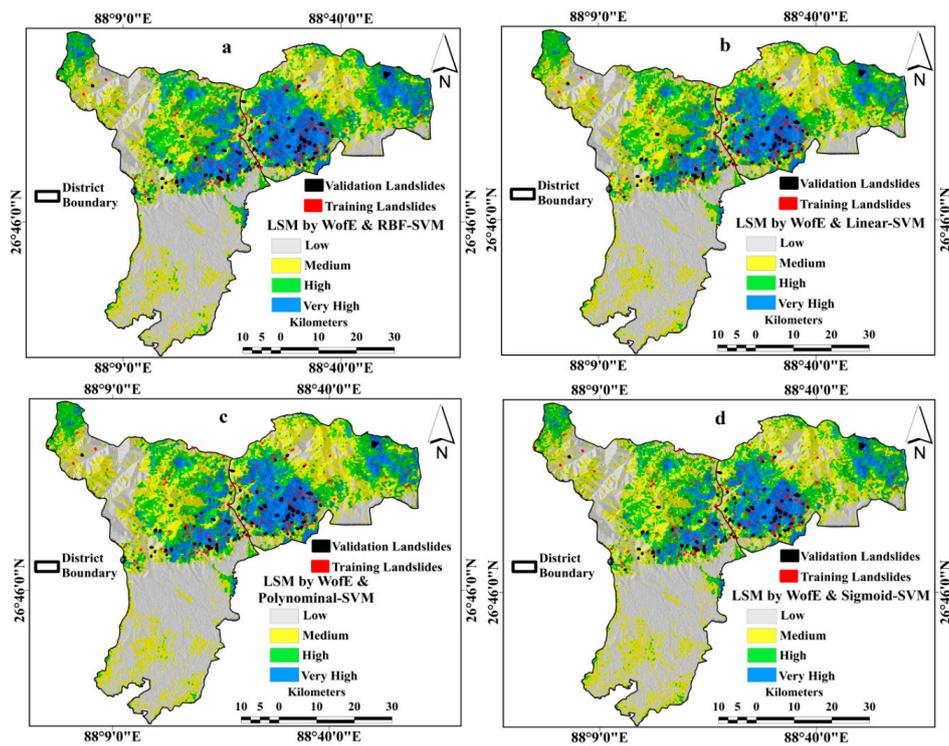
Figure 5. Landslide Susceptibility maps (LSMs) produced by different ensemble models – (**a**). WofE&
RBF-SVM, (**b**). WofE&Linear-SVM, (**c**). WofE& Polynomial-SVM, (**d**). WofE& Sigmoid-SVM models.
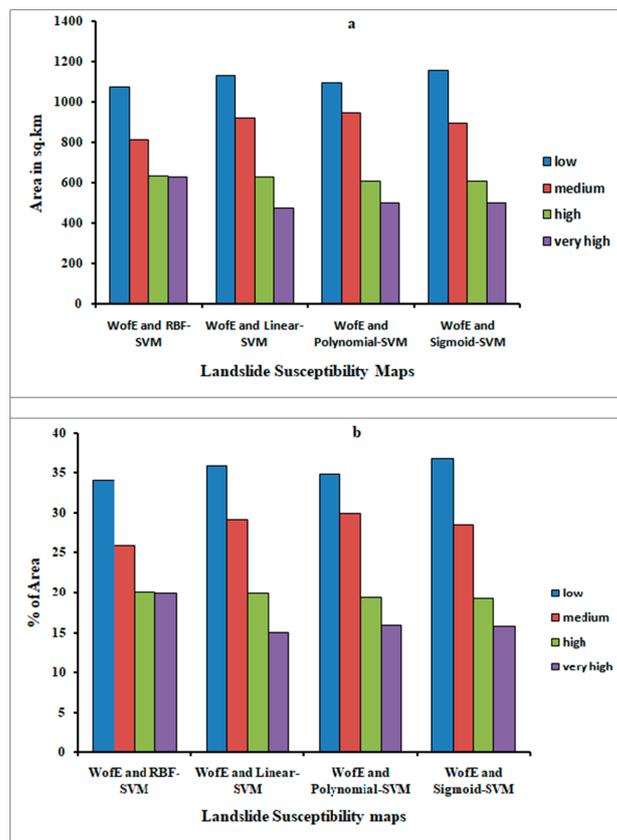

Figure 6. Areal distributions of LSMs by– (**a**). area distribution of LSMs, (**b**). percentage of area
distribution of LSMs.

3.4. Validation and Comparison of Models

The landslide susceptibility maps of Darjeeling and Kalimpong districts were prepared by the ensembles of WofE and SVM. These LSMs were then validated using the receiver operating characteristics (ROC) curve, which justifies and evaluates the accuracy of the models [48–56]. The ROC curve was prepared along the X and Y-axis. The X-axis indicates the false positive rate (1-specificity) and the Y-axis indicates the true positive rate (sensitivity) [57]. ROC curves have been extensively used for the assessment of susceptibility maps [8,12,15,58–66]. In the present study, of the 326 landslides, 98 (30%) landslides were used to validate the landslide susceptibility maps. The area under curve (AUC) values of the ensemble models WofE& RBF-SVM, WofE& Linear-SVM, WofE& Polynomial-SVM, and WofE& Sigmoid-SVM are 87%, 90%, 88%, and 85%, respectively, indicating that they are very good models for the identification of landslide hazard zones (Figure 7a–d). Based on the results of the ROC curves, the WofE& Linear-SVM model is considered more accurate (AUC = 90%) than the other three ensemble models.
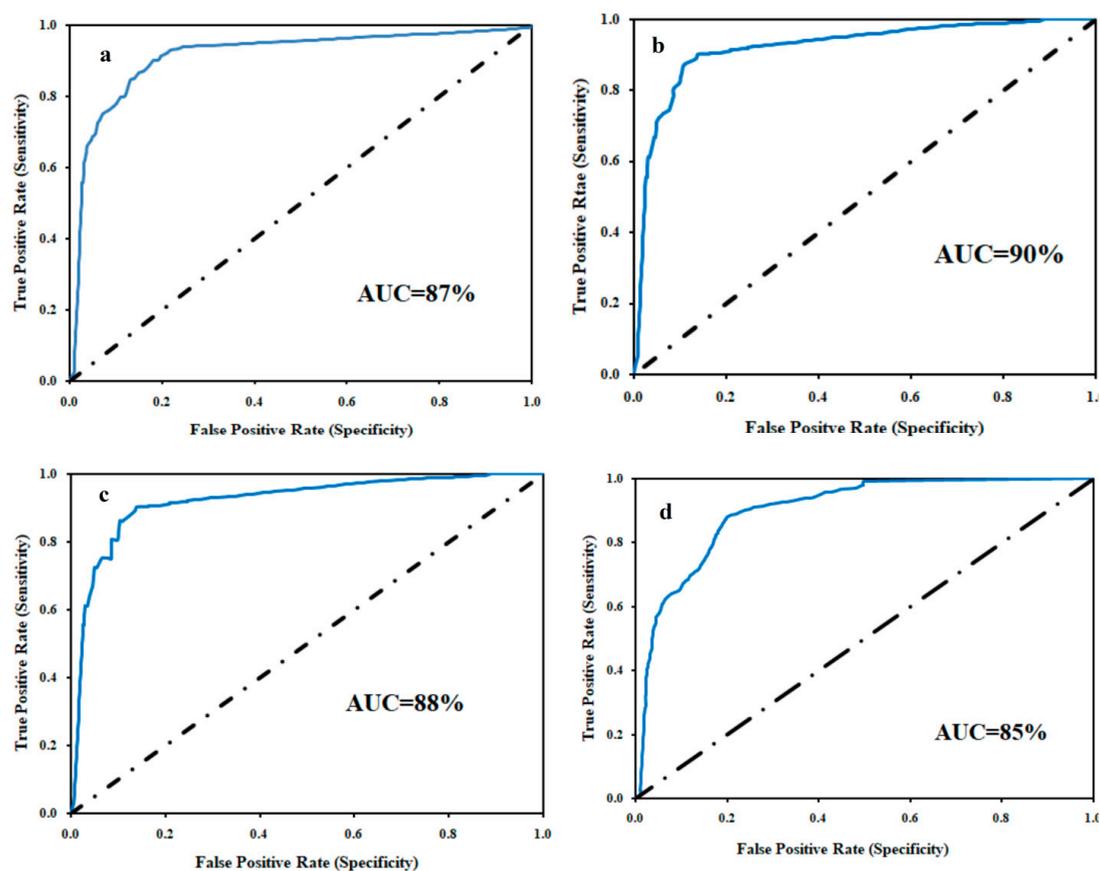
Figure 7. Validation of LSMs using the ROC curve showing the area under curve (AUC) – (**a**). WofE& RBF-SVM, (**b**). WofE&Linear-SVM, (**c**). WofE& Polynomial-SVM, (**d**). WofE& Sigmoid-SVM models.

It is not sufficient to validate the susceptibility models with only one validation method because this can lead to erroneous results if the samples are randomly distributed across the basin. Therefore, it is essential to cross check the validation result using another suitable validation method. In the present study, the quality sum (Qs) index was used as a second method to assess the accuracy and compare the landslide susceptibility models. Abedini and Tulabi [67] used the Qs method for landslide hazard assessment. In the Qs method, greater values indicate a higher accuracy and correctness of

the landslide susceptibility map, whereas lower values indicate lower accuracy [67]. To evaluate this index, the density ratio (Dr) was first calculated using Equation (11).

$$D_r = \frac{\frac{S_i}{A_i}}{\frac{\sum_i^n S_i}{\sum_i^n A_i}} \qquad (11)$$

where *Si* is the sum of the area of the landslides in each risk class, *Ai* is the area in the class of risk, and *n* is the number of risk classes in a zonation map. The Qs index is shown in Equation (12).

$$Q_s = \sum_{i=1}^{n} (D_r - 1)^2 \times S \qquad (12)$$

where *Qs* is the quality sum index, *Dr* is the density ratio, and *S* is the areal ratio of each risk class to the total area. The Qs method is a reliable validation technique which is calculated based on the landslide distribution and landslide hazards map using Equation (20). The four ensemble models in this study obtained the following Qs values: the WofE & RBF-SVM ensemble model scored 2.10; the WofE& Linear-SVM ensemble model scored 2.24; the WofE&Polynominal-SVM ensemble model scored 2.10, and the WofE& Sigmoid-SVM ensemble model scored 2.18 (Table 7). In line with the ROC results, the Qs validation results also indicate that the WofE and Linear-SVM model is more accurate than the other ensemble models.

Table 7. Mathematical Calculation of Qs Method of Ensemble LSMs.

Ensemble Models	Classes	ai (sq.km)	si (sq.km)	DR	s	Qs
WofE& RBF-SVM	Low	1071.23	0.00	0.00	0.34	2.10
	Medium	812.95	0.12	0.10	0.26	
	High	635.02	0.93	1.07	0.20	
	Very High	629.80	3.26	3.78	0.20	
WofE& Linear-SVM	Low	1127.55	0.00	0.00	0.36	2.24
	Medium	917.57	0.34	0.27	0.29	
	High	630.04	1.13	1.32	0.20	
	Very High	473.84	2.84	4.37	0.15	
WofE& Polynomial-SVM	Low	1095.14	0.00	0.00	0.35	2.10
	Medium	944.15	0.34	0.26	0.30	
	High	608.44	1.13	1.36	0.19	
	Very High	501.27	2.84	4.13	0.16	
WofE& Sigmoid-SVM	Low	1153.40	0.00	0.00	0.37	2.18
	Medium	892.57	0.23	0.19	0.28	
	High	604.55	1.25	1.51	0.19	
	Very High	498.48	2.84	4.16	0.16	

4. Discussion

Landslide susceptibility maps play a vital role in stakeholders making suitable decisions in landslide-prone areas. Landslide events not only cost human lives, but also destroy residential areas, roads, and agricultural fields. The assessment of landslide hazards using LSMs performed in this study is an important tool to mitigate landslide hazards, sustain the environment, and help the residents of high risk landslide susceptibility zones. In this study, ensemble models of weight-of-evidence (WofE)

and SVM were used to prepare the landslide susceptibility maps (LSMs). The different statistical, knowledge-driven, probabilistic, and machine learning models were used to recognize which areas are at severe risk of landslide occurrence. Several past studies have produced landslide susceptibility maps using different methods and models, such as landslide numerical risk factor (LNRF), frequency ratio (FR), analytical hierarchical process (AHP), SVM, artificial neural network (ANN), logistic regression (LR), conditional probability (CP), multi-criteria decision approach (MCDA), bivariate statistical, bivariate and multivariate models [8,9,60–65]. These studies determined the critical zones of landslide risk in their respective study regions. However, in the present study, a new ensemble technique was used, which has shown better results than those of previous studies. An ensemble of the two or three models may provide better results than any single model. In the present study, landslide susceptibility maps were prepared using ensemble models of WofE& RBF-SVM, WofE& Linear-SVM, WofE & Polynomial-SVM, and WofE & Sigmoid-SVM. These models are reliable and accurate in this field. The landslide susceptibility maps were created using landslide inventory data (326 landslides) and landslide conditioning factors (16 environmental factors). The landslide susceptibility maps (LSMs) produced by the ensemble models were classified into four susceptibility classes; namely, low, medium, high, and very high susceptibility to landslide occurrence. The high susceptibility landslide probability zones of the WofE & RBF-SVM, WofE& Linear-SVM, WofE & Polynomial-SVM and WofE& Sigmoid-SVM models cover areas of 630 km^2 (20%), 474 km^2 (15%), 501 km^2 (15%), 497 km^2 (15%), respectively.

The landslide susceptibility maps (LSMs) were validated and compared using the receiver operating characteristic (ROC) and quality sum (Qs) validation methods. Based on these validation methods, all models are considered very good to excellent. A high resolution DEM for this area is not freely available, posing the main challenge for the researchers in this study. If high resolution images were used for the extraction of landslide conditioning factors instead of a 30m DEM, these methods could be used to model landslide susceptibility at a micro level and achieve better results [68,69]. Of the four ensemble models, the landslide map produced by the WofE & Linear-SVM model is more suitable and accurate than those produced by other models. The areal distribution of the landslide susceptibility maps is shown in Figure 7. In the present study, these very high susceptibilities landslide probability zones are found in the middle portion of the study area. The areas in these districts closer to roads, such as NH-31 road, Rohini road, Rishi road, Darjeeling road, Sevoke road, and Sikkim-Kalimpong roads, are highly affected by landslides. Teesta River is the major river in these districts. The areas closer to the Teesta River are the most critical zone of landslide susceptibility. The Lish catchment, Mahananda catchment, and Torsha catchment are major catchments which are highly affected by landslides. The other critical landslide areas are Sukhia-Pokhari, Kurseong, Sevoke, Majua tea garden, and Kalimpong. The main factors determining landslide risk in these regions are heavy rainfall, steep slope, elevation, soil texture, geology, distance from road and LULC. During the monsoon season, these areas are strongly affected by landslides due to heavy rainfall. These regions are also affected by high seismic intensity, which is an important cause of landslides. However, the study carefully chalks out the landslides risk zones of Darjeeling and Kalimpong districts. This study will help the government to mitigate the landslides effect and strengthen the public conscious for sustainable development.

5. Conclusions

Landslides are very harmful natural hazards that cost human lives and cause widespread damage to roads, residences, gardens, and agricultural land. In this study, the weight-of-evidence (WofE) and SVM models were ensembled to produce landslide susceptibility maps (LSMs) for the Darjeeling and Kalimpong districts. The ensemble approach is an appropriate method for landslide susceptibility mapping that provides better results than using a single model. The four LSMs produced in this study were classified into four categories; namely, low, medium, high, and very high susceptibility to landslide occurrence. In the various models, the very high susceptibility class covered 20% (WofE& RBF-SVM mode), 15% (WofE& Linear-SVM model), 15.9% (WofE& Polynomial-SVM model), and

15% (WofE& Sigmoid-SVM models) of the study area, respectively. The very high landslide-prone areas are mainly located in the southern and middle parts of Darjeeling and Kalimpong districts. In particular, the Lish catchment area, Teesta catchment area, Sevoke road, and Majua tea garden areas are highly susceptible to landslide occurrences. The results of the ensemble models were validated using the QS index and ROC methods. Both validation methods confirmed the landslide susceptibility maps produced by the WofE& RBF-SVM, WofE& Linear-SVM, WofE& Polynomial-SVM, and WofE& Sigmoid-SVM ensemble methods as being excellent and appropriate. Of the four ensemble models, the WofE & Linear-SVM model was found to be more accurate than other ensemble models. This work helps to increase awareness of the public and government and aims to reduce the impact of landslides by providing steps and suitable strategies of hazard mitigation. Some necessary steps and techniques are essential in the very high landslide risk zones of the study area. Identification of faults, weak geological regions, proper drainage management, and afforestation programs in landslide-prone areas may reduce the landslide risks. The results obtained from this study can provide proper and significant information to the decision-makers and policy planners in the landslide-prone areas of these districts.

Author Contributions: Methodology, J.R., S.S., and A.A.; formal analysis, J.R., and S.S.; investigation, J.R., S.S., and A.A.; writing—original draft preparation, J.R., S.S., and A.A.; writing—review and editing, J.R., S.S., A.A., T.B., and D.T.B.

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

References

1. Bui, D.T.; 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* **2015**, *13*, 361–378. [CrossRef]

2. Cruden, D.M.; Varnes, D.J. Landslide Types and Processes, Transportation Research Board, U.S. National Academy of Sciences. *Spec. Rep.* **1996**, *247*, 36–75.

3. Gerrard, J. The landslide hazard in the Himalayas: Geological control and human action. *Geomorphology* **1994**, *10*, 221–230. [CrossRef]

4. Bhandari, R.K. Landslide hazard zonation: Some thoughts. In *Coping with Natural Hazards: Indian Context*; Valdiya, K.S., Ed.; Orient Longman: Hyderabad, India, 2004; pp. 134–152.

5. Panikkar, S.V.; Subramanyan, V.A. geomorphic evaluation of the landslides around Dehradun and Mussoorie, Uttar Pradesh, India. *Geomorphology* **1996**, *15*, 169–181. [CrossRef]

6. Sarkar, S. Landslides in Darjiling Himalayas. *Trans. Jpn. Geomorphol. Union* **1999**, *20*, 299–315.

7. Fan, X.; Scaringi, G.; Domènech, G.; Yang, F.; Guo, X.; Dai, L.; Huang, R. Two multi-temporal datasets that track the enhanced landsliding after the 2008 Wenchuan earthquake. *Earth Syst. Sci. Data* **2019**, *11*, 35–55. [CrossRef]

8. Yilmaz, I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: Conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ. Earth Sci.* **2009**, *61*, 821–836. [CrossRef]

9. Abedini, M.; Ghasemyan, B.; Mogaddam, M.H.R. Landslide susceptibility mapping in Bijar city, Kurdistan Province, Iran: A comparative study by logistic regression and AHP models. *Environ. Earth Sci.* **2017**, *76*, 308. [CrossRef]

10. Regmi, A.D.; Devkota, K.C.; Yoshida, K.; Pradhan, B.; Pourghasemi, H.R.; Kumamoto, T.; Akgun, A. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arab. J. Geosci.* **2013**, *7*, 725–742. [CrossRef]

11. Chawla, A.; Pasupuleti, S.; Chawla, S.; Rao, A.C.S.; Sarkar, K.; Dwivedi, R. Landslide Susceptibility Zonation Mapping: A Case Study from Darjeeling District, Eastern Himalayas, India. *J. Indian Soc. Remote Sens.* **2019**, *47*, 497–511. [CrossRef]

12. Shahabi, H.; Hashim, M. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. *Sci. Rep.* **2015**, *5*, 15. [CrossRef] [PubMed]

13. Roy, J.; Saha, S. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenvironmental Disasters* **2019**, *6*, 11. [CrossRef]

14. 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]

15. Pourghasemi, H.R.; Jirandeh, A.G.; Pradhan, B.; Xu, C.; Gokceoglu, C. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *J. Earth Syst. Sci.* **2013**, *122*, 349–369. [CrossRef]

16. Pham, B.T.; Pradhan, B.; Bui, D.T.; Prakash, I.; Dholakia, M. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [CrossRef]

17. Goetz, J.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [CrossRef]

18. Gravina, T.; Figliozzi, E.; Mari, N.; Schinosa, F.D.L.T. Landslide risk perception in Frosinone (Lazio, Central Italy). *Landslides* **2016**, *14*, 1419–1429. [CrossRef]

19. Pham, B.T.; Bui, D.T.; Prakash, I.; Dholakia, M. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **2017**, *149*, 52–63. [CrossRef]

20. Pawde, M.B.; Saha, S.S. *Geology of Darjeeling Himalaya*; GSI: Kolkata, India, 1982.

21. Pramanik, M.K. Site suitability analysis for agricultural land use of Darjeeling district using AHP and GIS techniques. *Model. Earth Syst. Environ.* **2016**, *2*. [CrossRef]

22. Government of West Bengal. *Bureau of Applied Economics and Statistics*; Department of Statistics & Programme Implementation, District Statistical Handbook, Government of West Bengal: Kolkata, India, 2013.

23. Guzzetti, F.; Reichenbach, P.; Ardizzone, F.; Cardinali, M.; Galli, M. Estimating the quality of landslide susceptibility models. *Geomorphology* **2006**, *81*, 166–184. [CrossRef]

24. Li, Z.; Zhu, Q.; Gold, C. *Digital Terrain Modeling: Principles and Methodology*; CRC Press: Boca Raton, FL, USA, 2005.

25. Wentworth, C.K. A simplified method of determining the average slope of land surfaces. *Am. J. Sci.* **1930**, *117*, 184–194. [CrossRef]

26. Burrough, P.A.; McDonell, R.A. *Principles of Geographical Information Systems*; Oxford University Press: New York, NY, USA, 1998; p. 190.

27. Bayraktar, H.; Turalioglu, S. A Kriging-based approach for locating a sampling site—In the assessment of air quality. *Stoch. Environ. Res. Risk Assess.* **2005**, *19*, 301–305. [CrossRef]

28. Anderson, C.G.; Maxwell, D.C. *Starting a Digitization Center*; Elsevier: Amsterdam, The Netherlands, 2004; ISBN 978-1843340737.

29. Ay, N.; Amari, S.-I. A Novel Approach to Canonical Divergences within Information Geometry. *Entropy* **2015**, *17*, 8111–8129. [CrossRef]

30. Myung, I.J. Tutorial on Maximum Likelihood Estimation. *J. Math. Psychol.* **2003**, *47*, 90–100. [CrossRef]

31. Crippen, R.E. Calculating the vegetation index faster. *Remote Sens. Environ.* **1990**, *34*, 71–73. [CrossRef]

32. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30. [CrossRef]

33. Moore, I.D.; Burch, G.J. Physical Basis of the Length Slope Factor in the Universal Soil Loss Equation. *Soil Sci. Soc. Am.* **1986**, *50*, 1294–1298. [CrossRef]

34. Available online: http://dx.doi.org/10.2136/sssaj1986.03615995005000050042x (accessed on 21 October 2017).

35. O'Brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **2007**, *41*, 673–690. [CrossRef]

36. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Lautenbach, S. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2012**, *36*, 27–46. [CrossRef]

37. Wang, H.; Wang, G.; Wang, F.; Sassa, K.; Chen, Y. Probabilistic modeling of seismically triggered landslides using Monte Carlo simulations. *Landslide* **2008**, *5*, 387–395. [CrossRef]

38. Mohammady, M.; Pourghasemi, H.R.; Pradhan, B. Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster-Shafer, and weights-ofevidence models. *J. Asian Earth Sci.* **2012**, *61*, 221–236. [CrossRef]

39. Bonham-Carter, G.F. Geographic information systems for geoscientists: Modeling with GIS. In *Computer Methods in the Geosciences*; Bonham-Carter, F., Ed.; Pergamon: Oxford, UK, 1994; p. 398.

40. Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Dhakal, S.; Paudyal, P. Predictive modeling of rainfall-induced landslide hazard in the Lesser Himalaya of Nepal based on weights-of evidence. *Geomorphology* **2008**, *102*, 496–510. [CrossRef]

41. Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Masuda, T.; Nishino, K. GIS-based weights-of-evidence modeling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ. Geol.* **2008**, *54*, 314–324. [CrossRef]

42. Wan, S.; Lei, T.C. A knowledge-based decision support system to analyze the debris-flow problems at Chen-Yu-Lan River, Taiwan. *Knowl. Based Syst.* **2009**, *22*, 580–588. [CrossRef]

43. Yao, X.; Tham, L.; Dai, F. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [CrossRef]

44. Marjanovic, M.; Kovacevic, M.; Bajat, B.; Vozenilek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [CrossRef]

45. Tehrany, M.S.; Pradhan, B.; Jebu, M.N. A comparative assessment between object and pixel-based classification approaches for land use/land cover mapping using SPOT 5 imagery. *Geocarto Int.* **2013**, *29*, 1–19. [CrossRef]

46. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naïve Bayes Models. *Math. Probl. Eng.* **2012**, *2012*, 1–26. [CrossRef]

47. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

48. Samui, P. Slope stability analysis: A support vector machine approach. *Environ. Geol.* **2008**, *56*, 255–267. [CrossRef]

49. Arabameri, A.; Pradhan, B.; Rezaei, K. Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS. *J. Environ. Manag.* **2019**, *232*, 928–942. [CrossRef]

50. Arabameri, A.; Pradhan, B.; Rezaei, K. Spatial prediction of gully erosion using ALOS PALSAR data and ensemble bivariate and data mining models. *Geosci. J.* **2019**, *23*, 1–18. [CrossRef]

51. Arabameri, A.; Cerda, A.; Tiefenbacher, J.P. Spatial pattern analysis and prediction of gully erosion using novel hybrid model of entropy-weight of evidence. *Water* **2019**, *11*, 1129. [CrossRef]

52. Arabameri, A.; Pradhan, B.; Rezaei, K.; Conoscenti, C. Gully erosion susceptibility mapping using GISbased multi-criteria decision analysis techniques. *Catena* **2019**, *180*, 282–297. [CrossRef]

53. Arabameri, A.; Rezaei, K.; Cerda, A.; Lombardo, L.; Rodrigo-Comino, J. GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *Sci. Total Environ.* **2019**, *658*, 160–177. [CrossRef]

54. Arabameri, A.; Pourghasemi, H.R.; Yamani, M. Applying different scenarios for landslide spatial modeling using computational intelligence methods. *Environ. Earth Sci.* **2017**, *76*, 832. [CrossRef]

55. Arabameri, A.; Pradhan, B.; Rezaei, K.; Sohrabi, M.; Kalantari, Z. GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. *J. Mt. Sci.* **2019**, *16*, 595–618. [CrossRef]

56. Arabameri, A.; Pradhan, B.; Rezaei, K.; Saro, L.; Sohrabi, M. An ensemble model for landslide susceptibility mapping in a forested area. *Geochem. Int.* **2019**, 1–18. [CrossRef]

57. Chung, C.J.F.; Fabbri, A.G. Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Nat. Hazards* **2003**, *30*, 451–472. [CrossRef]

58. Negnevitsky, M. *Artificial Intelligence—A Guide to Intelligent Systems*; Addison-Wesley Co.: Boston, MA, USA, 2002; p. 394.

59. Mallick, J.; Singh, R.K.; Alawadh, M.A.; Islam, S.; Khan, R.A.; Qureshi, M.N. GIS-based landslide susceptibility evaluation using fuzzy-AHP multi-criteria decision-making techniques in the Abha Watershed, Saudi Arabia. *Environ. Earth Sci.* **2018**, *77*, 276. [CrossRef]

60. Feizizadeh, B.; Roodposhti, M.S.; Jankowski, P.; Blaschke, T. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. *Comput. Geosci.* **2014**, *73*, 208–221. [CrossRef] [PubMed]

61. Park, S.; Choi, C.; Kim, B.; Kim, J. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ. Earth Sci.* **2012**, *68*, 1443–1464. [CrossRef]

62. Bijukchhen, S.M.; Kayastha, P.; Dhital, M.R. A comparative evaluation of heuristic and bivariate statistical modelling for landslide susceptibility mappings in Ghurmi–DhadKhola, east Nepal. *Arab. J. Geosci.* **2012**, *6*, 2727–2743. [CrossRef]

63. Pradhan, B.; Youssef, A.M. Manifestation of remote sensing data and GIS on landslide hazard analysis using spatial-based statistical models. *Arab. J. Geosci.* **2009**, *3*, 319–326. [CrossRef]

64. Ozdemir, A.; Altural, T. A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. *J. Asian Earth Sci.* **2013**, *64*, 180–197. [CrossRef]

65. Arabameri, A.; Yamani, M.; Pradhan, B.; Melesse, A.; Shirani, K.; Bui, D.T. Novel ensembles of COPRAS multi-criteria decision-making with logistic regression, boosted regression tree, and random forest for spatial prediction of gully erosion susceptibility. *Sci. Total Environ.* **2019**, *688*, 903–916. [CrossRef]

66. Tien Bui, D.; Hoang, N.D.; Martínez-Álvarez, F.; Ngo, P.T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Sci. Total Environ.* **2019**. [CrossRef]

67. Abedini, M.; Tulabi, S. Assessing LNRF, FR, and AHP models in landslide susceptibility mapping index: A comparative study of Nojian watershed in Lorestan province, Iran. *Environ. Earth Sci.* **2018**, *77*, 405. [CrossRef]

68. Haneberg, W.C.; Cole, W.F.; Kasali, G. High-resolution lidar-based landslide hazard mapping and modeling, UCSF Parnassus Campus, San Francisco, USA. *Bull. Eng. Geol. Environ.* **2009**, *68*, 263–276. [CrossRef]

69. Nichol, J.E.; Shaker, A.; Wong, M.S. Application of high-resolution stereo satellite images to detailed landslide hazard assessment. *Geomorphology* **2006**, *76*, 68–75. [CrossRef]