Article

Comparison between Deep Learning and Tree-Based Machine Learning Approaches for Landslide Susceptibility Mapping

Sunil Saha [1], Jagabandhu Roy [1], Tusar Kanti Hembram [2], Biswajeet Pradhan [3,4,*], Abhirup Dikshit [3], Khairul Nizam Abdul Maulud [4,5] and Abdullah M. Alamri [6]

1   Department of Geography, University of Gour Banga, Malda 732103, India; sunilgeo.88@gmail.com (S.S.); jagabandhuroy1991@gmail.com (J.R.)
2   Department of Geography, Nistarini College, Purulia 723101, India; tusarpurulia1991@gmail.com
3   Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney 2007, Australia; abhirup.dikshit@student.uts.edu.au
4   Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Malaysia; knam@ukm.edu.my
5   Department of Civil Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
6   Department of Geology & Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 145111, Saudi Arabia; alamri.geo@gmail.com
*   Correspondence: Biswajeet.Pradhan@uts.edu.au or Biswajeet24@gmail.com

Abstract: The efficiency of deep learning and tree-based machine learning approaches has gained immense popularity in various fields. One deep learning model viz. convolution neural network (CNN), artificial neural network (ANN) and four tree-based machine learning models, namely, alternative decision tree (ADTree), classification and regression tree (CART), functional tree and logistic model tree (LMT), were used for landslide susceptibility mapping in the East Sikkim Himalaya region of India, and the results were compared. Landslide areas were delimited and mapped as landslide inventory (LIM) after gathering information from historical records and periodic field investigations. In LIM, 91 landslides were plotted and classified into training (64 landslides) and testing (27 landslides) subsets randomly to train and validate the models. A total of 21 landslides conditioning factors (LCFs) were considered as model inputs, and the results of each model were categorised under five susceptibility classes. The receiver operating characteristics curve and 21 statistical measures were used to evaluate and prioritise the models. The CNN deep learning model achieved the priority rank 1 with area under the curve of 0.918 and 0.933 by using the training and testing data, quantifying 23.02% and 14.40% area as very high and highly susceptible followed by ANN, ADtree, CART, FTree and LMT models. This research might be useful in landslide studies, especially in locations with comparable geophysical and climatological characteristics, to aid in decision making for land use planning.

Keywords: landslide; deep learning model; tree-based machine learning model; ridge regression; GIS; East Sikkim Himalaya

1. Introduction

In mountainous regions, landslides are regarded as one of the reoccurring natural hazards affecting human property and lives. Landside susceptibility (LS) indicates the spatial probability of landslides in an area [1]. Landslides are often triggered by earthquakes or heavy precipitations within certain geomorphological, geological and hydrological settings. However, other primary factors controlling landslide failure mechanisms, including in situ stresses, weathering and heave, play vital roles. In mountainous regions, the effect of landslides can change the topographic characteristic, forest, soil properties (consistence, structure, density, temperature, etc.), road and farming land, depending on the magnitude

of landslides [2]. Landslides in Indian Himalayas account for over 14% of the global landslides as per Froude and Petley [3] database, and one of the understudied regions is Sikkim, despite it having a huge landslide problem. Dikshit et al. [4] revealed that only 10% of landslide studies are conducted in Sikkim, of total studies across the Indian Himalayan states. The study area has witnessed destructive landslide events between 2007 and 2015 (https://data.nasa.gov/Earth-Science/Global-Landslide-Catalog/h9d8-neg4, accessed on 20 December 2020).

A landslide susceptibility map (LSM) delineates the spatial occurrence of landslides that might occur in the future, based on present and past landslide occurrences. LSMs are used as an important input for hazard and risk assessment and can be used for land use management, civil constructions such as roads and railways, and decision making [5,6]. For the LSM assessment, qualitative and quantitative approaches have been utilised [7]. The qualitative methods include knowledge-driven and inventory-based techniques. The statistical models used in LSM are frequency ratio [8], logistic regression [8], information value [9], bivariate statistics [10], multivariate regression [10], multivariate adaptive regression splines [11], weights-of-evidence [12], weighted linear combinations [13] and generalised additive [14].

Recently, data-driven models, which are considered quantitative methods, have been proved as an extremely effective tool for temporal and spatial mapping of landslides. Machine learning (ML) and soft computing methods are commonly used to assess the LSM [15]. Many hybrid ML methods, including support vector machine (SVM), Naïve Bayes (NB), boosted regression tree (BRT), artificial neural network (ANN), decision tree (DT), neuro-fuzzy and random forest (RF), have been widely used in LSM [16]. ML methods are consistent in removing the overfitting and noise problems in modelling, to increase the accuracy of a model [17].

Among the advanced artificial intelligence approaches, neural network (NN) models are found to be capable and reliable for big data processing and modelling [18]. Effective applications of ANNs have been achieved in various fields of domains [19]. ANN has been acknowledged as an effective and powerful approach in different studies, such as slope stability analysis [20], soil analysis [21] and rock characteristic studies [22], due to its capacity to manage data without relying on the measurement scale or the way data are organised. In previous studies, NN has shown good to very good accuracy in LSM studies [23,24]. ML models need manual intervention, thereby affecting the results [25]. Therefore, more emphasis has been given on deep learning (DL) models to eliminate these concerns. DL is one of the main branches of ML that differs from task-specific methods focused on data representations that are supervised [18], semi supervised [18] and unsupervised [26]. DL is used in several fields and provides superior results than other models [4,26]. DL algorithms are a subset of ML algorithms that use several layers of nonlinear information to model complicated data relationships, most frequently with the use of a multilayer NN [27]. According to various literature, DL models outperformed the benchmark ML models, notably in landslide susceptibility modelling [28]. Convolution neural network (CNN) is one of the most popular DL approaches amongst the various DL approaches for recognition of different classification problems [18,29]. In particular, CNN can use convolution and pooling layers to identify patterns with extreme variability, representing the translation-invariant nature of most images [30]. CNN has been used widely to identify the vulnerability areas in the present scenario by using remote sensing data [31]. Tree-based ML provides good results and has been used for modelling in different fields of geoscience, engineering, and medical sciences. Taherdangkoo et al. [32] used regression tree, BRT, least square SVM and Gaussian process regression-based four DT models for methane solubility prediction, and obtained highly accurate results. In another work, Gokgoz et al. [33] used CART, C4.5 and RF for classifying electromyogram signal and obtained an excellent result with the area under the curve (AUC), of more than 90%. Park et al. [34] used DT models in landslide susceptibility modelling. These models acquired more than 82% AUC. DL and tree-based models are used separately in different fields

for modelling, and provide good results. In the Indian Himalayan context, comparison between DL techniques and tree-based models in landslide susceptibility mapping has not been considered in the literature. The accuracy of the same model may differ for landslide susceptibility modelling, and the characteristic selection of the best model is important for managing landslides. The main research questions are: (1) is a DL model (CNN) applicable in landslide susceptibility analysis in Indian Himalayan region, and, (2) can a CNN model provide a better result than tree-based ML models in the Indian Himalayan region. A research gap is found in this respect, as no comparisons have been made between the DL model and tree-based ML models in the Indian Himalayan region. This study encourages Indian researchers who are working in the landslide field. The findings of this study will provide a solid foundation for earth scientists, government officials and other stakeholders to enhance land management and disaster management.

Based on the aforementioned research gap, CNN, ANN, alternative DT (ADTree), CART, functional tree (FTree) and logistic model tree (LMT) DL, and hybrid tree ML, approaches were used for assessing landslide susceptibility in the East Sikkim Himalaya, India, in the present study. Twenty statistical measures and a receiver operating characteristic (ROC) curve were applied to evaluate the performance of the models.

2. Description of the Study Area

The study area of East Sikkim is a mountainous district of the state of Sikkim, in India (Figure 1). The whole district is characterised by hilly terrain and rugged topography [35]. The highest and lowest altitudes of the study area are 4695 m and 264 m, respectively. Geographically, East Sikkim occupies the southeast corner of Sikkim, and the latitudinal/longitudinal extensions are 27°25′ N to 27°8′ N and 88°53′ E to 88°26′ E, respectively, with an area of 964 km^2. The district is considered to be an extremely sensitive area, sharing an international border with China and Bhutan. The state capital, Gangtok, the hub of all administrative activities, is the main town of East Sikkim. In accordance with the census of India, 2011, East Sikkim district has a population of 283,583 with a population density of 295/km^2 (Census of India, 2011). Frequent occurrences of landslides are the main barrier to the development of the tourism industry and socioeconomic growth of this state. The preparation of LSM by sound techniques could be an important method to tackle this problem.
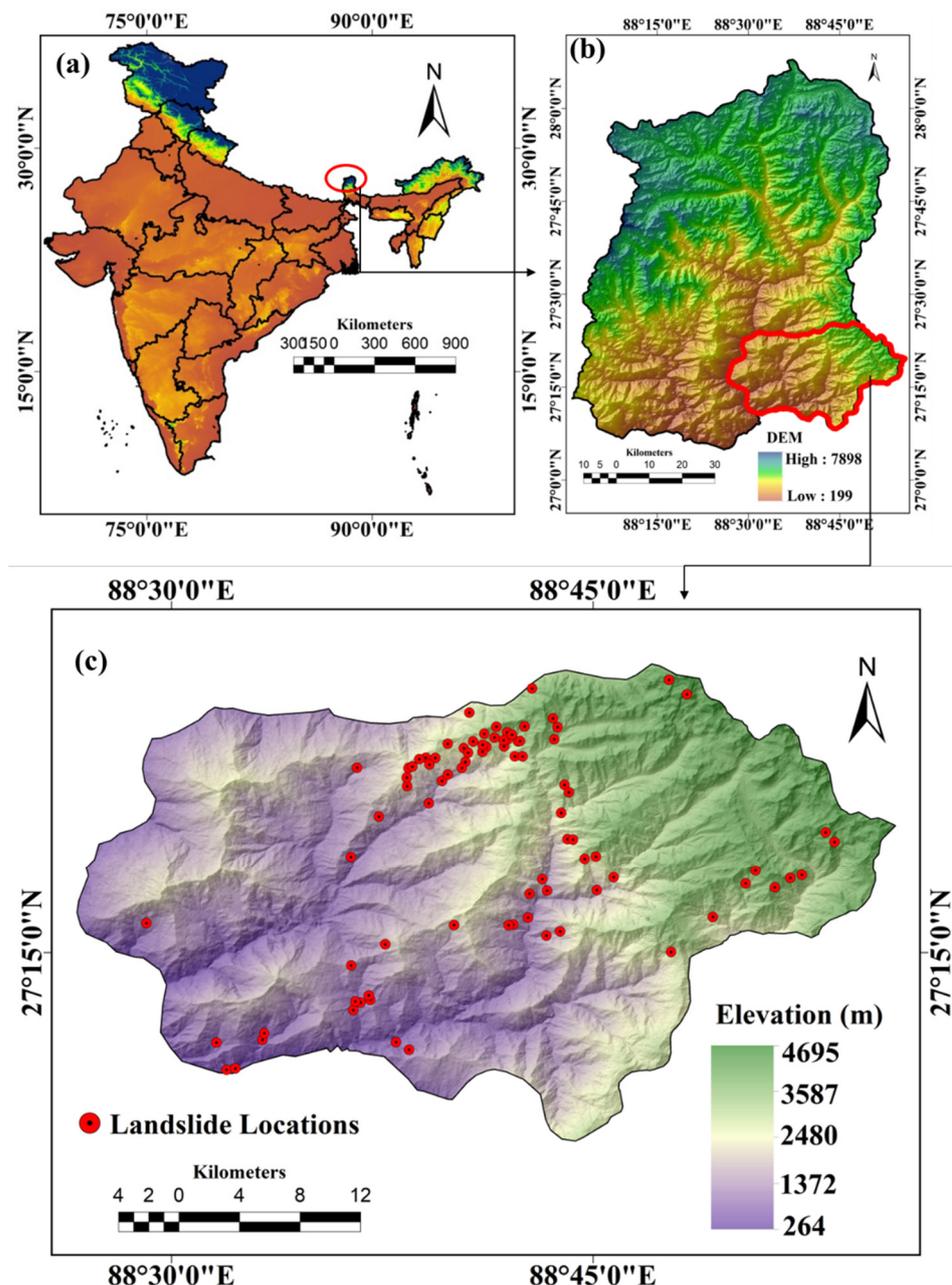
Figure 1. Location map of the study area showing: (**a**) India, (**b**) Sikkim state, (**c**) East Sikkim district.

3. Materials and Methods

The steps and processes involved in the study (Figure 2) are as follows:

(i)   Landslide inventory map (LIM) construction-Landslide locations were firstly identified by using Google Earth images to produce a LIM. Subsequently, a detailed field survey was conducted on November 2019, and historical records were acquired for verifying the location of landslides.

(ii)   After a literature review, and based on the geoenvironmental condition of the study area, landslide conditioning factors (LCFs) were selected, and thematic layers of LCFs were prepared on a geographical information system (GIS) platform.

(iii) After selecting the LCFs based on previous literature and the geoenvironmental condition, a factor selection process was performed by using multicollinearity assessment and chi-square attribute evaluation (CSAE) techniques to choose appropriate

LCFs for landslide susceptibility modelling. A total of 21 LCFs were chosen as appropriate factors.

(iv) LSMs were then generated by using DL (CNN) and hybrid tree-based ML approach (ANN, ADtree, CART, FTree and LMT) models, and their results were compared.

(v) After modelling the landslide susceptibility, ridge regression (RR) was applied to verify the importance of the selected LCFs in producing an LSM.

(vi) The accuracy of each model was assessed by applying the ROC curve and 21 statistical measures. This process was performed to compare the results of the models and select the best amongst them.
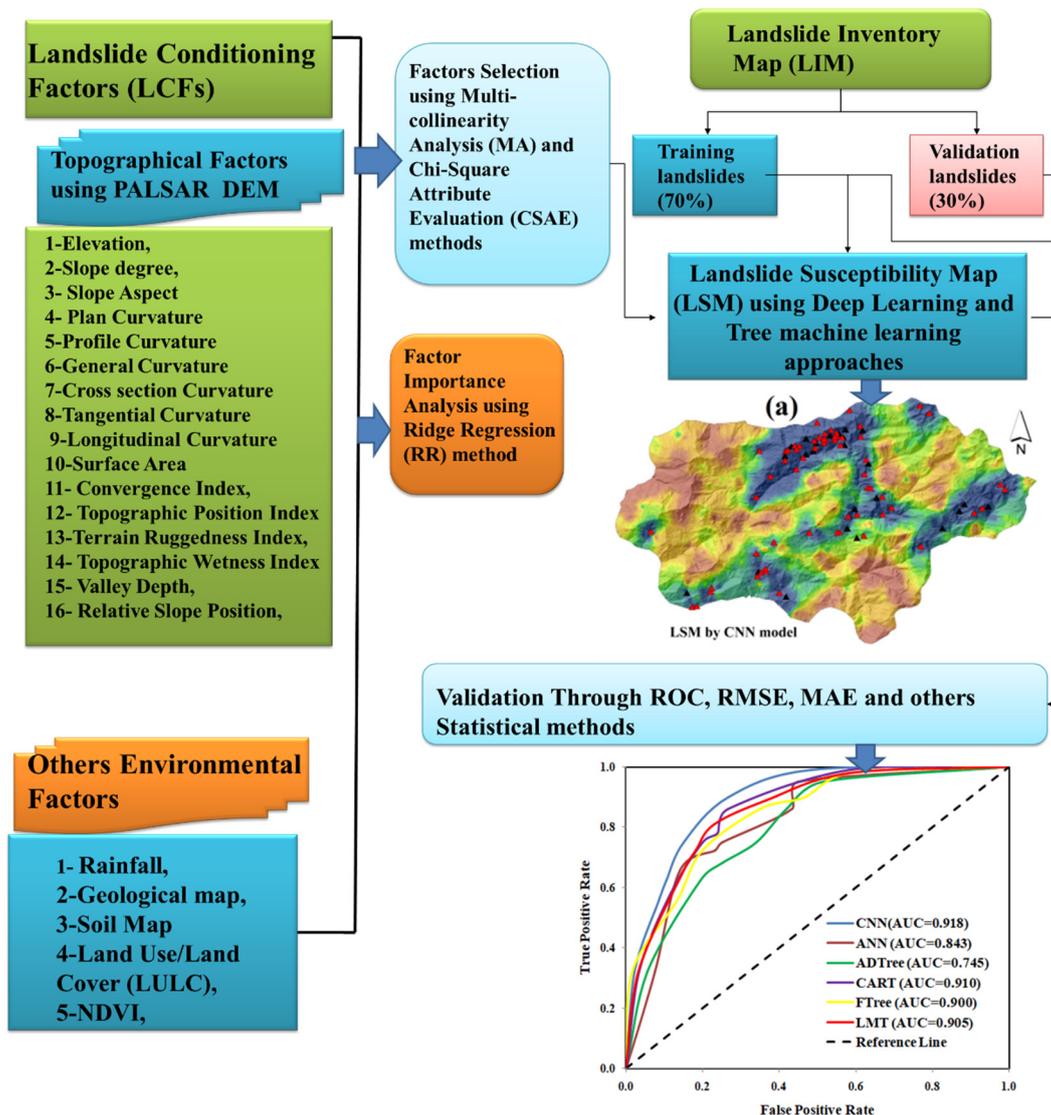
Figure 2. Flowchart representing the methodology.

3.1. Data Used

Primary and secondary datasets were used in this work. The primary data, such as size and location of the landslides, were collected through field observation by using a global positioning system (GPS) device on November 2019. The secondary data of the current study includes PALSAR DEM, geology map, soil map, Landsat 8 OLI/TIRS, topographical map and precipitation data collected from different organisations and government departments. The high-resolution phased array type L-band synthetic aperture radar (PALSAR) DEM with 12.5 m × 12.5 m spatial resolution, dated 2011, was downloaded

from Alaska Satellite Facility (https://asf.alaska.edu/, accessed on 20 December 2020). The geology map with the scale of 1:500,000 was collected from the Geological Survey of India (https://www.gsi.gov.in/, accessed on 20 December 2020). Topographical maps with a scale of 1:50,000 were collected from the Survey of India. Rainfall data were obtained from the Indian Meteorological Department. The National Land Survey and Land Use Planning Bureau provided the soil map. Landsat 8 OLI images were collected from the USGS Earth-Explorer (https://earthexplorer.usgs.gov, accessed on 20 December 2020) with a resolution of 30 m × 30 m. The resolution of the collected data is different. For preparing the LSMs, a resolution of PALSAR DEM (12.5 m × 12.5 m) was selected as the base resolution for the collected data, and other factors having a higher or lower resolution were resampled to 12.5 m × 12.5 m.

3.2. Preparation of LIM

LIM was constructed with the integration of past and present landslide events [36]. Several sources, including primary and secondary combinations, were used to construct the LIM. Field investigation, identification of landslide location and coordinates of landslide using GPS were used as primary sources to create LIM. The field survey was conducted on November 2019. The secondary data sources, including historical landslide information (Administrative Department of Sikkim), were used to build the LIM map. High-resolution Google Earth images can easily detect landside locations in remote and inaccessible areas, and a total of 91 landslides were identified and mapped to produce the LIM layer. For training and testing the models, the same number of non-landslide locations were randomly selected. A total of 182 landslide and non-landslide samples were taken for modelling and validation purposes. Landslide sample points were divided into two subsets, namely, training and validation datasets. Various ratios were used to divide the inventory dataset. However, most studies followed the 70:30 ratio to prepare training and validation datasets [37], and a similar approach was used to divide the dataset. Dao et al. [38] and Nhu et al. [39] used DL NNs for landslide susceptibility assessment and prepared the LIM by considering 217 and 193 samples. These samples were divided into 70:30 ratio for training and testing the LSMs. Measurement of the landslides disclosed that the smallest was 153 m^2, and the largest was 841 m^2. The average maximum and minimum widths of the landslides were 404 m^2 and 94 m^2, respectively. Four types of landslides were found, namely, rock falls (37.1%), debris slides (31.08%), rotational slides (20.28%) and complex and compound (combination in time and/or space of two or more principal types of movement) slides (11.54%). Some field photographs are shown in Figure 3.

3.3. Preparation of LCFs

Consideration of relevant LCFs in the modelling process is extremely essential to model the landslide susceptibility of a given area. Twenty-one LCFs were used for landslide modelling. Sixteen out of 21 factors, such as elevation, slope, aspect, plan curvature, profile curvature, general curvature, tangential curvature, terrain ruggedness index (TRI), topographic wetness index (TWI), cross sectional curvature, longitudinal curvature, topographic position index (TPI), valley depth (VD) and relative slope position (RSP), are topographical factors that were extracted from the PALSAR DEM. The other five factors, namely, rainfall, geology, soil, land use/land cover (LULC) and normalised difference vegetation index (NDVI), were collected from different sources. These factors were categorised as topographical factors, and other environmental factors are detailed below.

3.3.1. Topographical Factors

Elevation is regarded as an important factor because higher altitudes have been found to be a major cause of landslide occurrence in mountain regions [40]. The LCFs, such as slope, aspect, topographic position index, convergence index, plan curvature, terrain ruggedness index, profile curvature, general curvature, tangential curvature, topographic wetness index, longitudinal curvature, cross section curvature, surface area, VD, and RSP,

were extracted from ALOS PALSAR DEM by using the SAGA GIS tool [38]. Slope is considered to be an important factor because it influences the occurrence of landslides and speed of sliding materials. The slope value in the study area varies from 0 (flat) to 80.16°, indicating an extremely steep gradient (Figure 4b) and the possibility of instability. Slope aspect (Figure 4c) indicates the direction of the slope, and has a strong relationship with landslide. It has an indirect influence because it determines the amount, duration, intensity of insolation intake, intensity of rainfall received, vegetation cover, and soil structure and texture [40]. Plan curvature is a horizontal plane and is considered the hypothetical line that crosses the contour of a given cell and regulates the stability of a landmass (Figure 4d). The surface curvature in the direction of the steepest slope is known as the profile curvature. The flow velocity of surface water, soil erosion and deposition are influenced by the profile curvature. The erosion prevails in the convex surface region, and the concave curvature is the area to be deposited [41]. The spatial distribution of the profile curvature ranges from −0.095 to 0.140 (Figure 4e). The general curvature, which was proposed by Wood [42], is defined as the total curvature that has been created by the intersection of the surface with a plan. It can be classified into three categories: convex, concave, and flat surface, corresponding to the peaks, valley, and plain areas, respectively. The spatial value of general curvature varies from −0.819 to 0.934 (Figure 4k). Wilson and Gallant [43] suggested a tangential curvature, which is curvature to a steep gradient along the orthogonal line. The value of tangential curvature in the present study ranges from −0.115 to 0.167 (Figure 4l). A similar approach of tangential curvature is found in longitudinal curvature [42]. The curvature of the line between the surface and the plane is conceptually identical and defined by the direction of slope and aspect. It can be viewed in the same way as curvature of the profile, indicating how a liquid substance is going to accelerate or decelerate over a point. The spatial distribution of the longitudinal curvature ranges from −0.585 to 0.627 (Figure 4n).

Figure 3. Field photographs of some selected landslide locations of the study area; (**a**) Gangtok Forest Block (27°22′20.07″ N and 88°40′32.60″ E); (**b**) Jawarharlal Nehru Marg Road (27°22′16.77″ N and 88°40′36.21″ E); (**c**) Tadong area (27°18′31.60″ N and 88°36′24.91″ E); and (**d**) Jawarharlal Nehru Marg Road (27°21′57.71″ N and 88°39′14.47″ E).
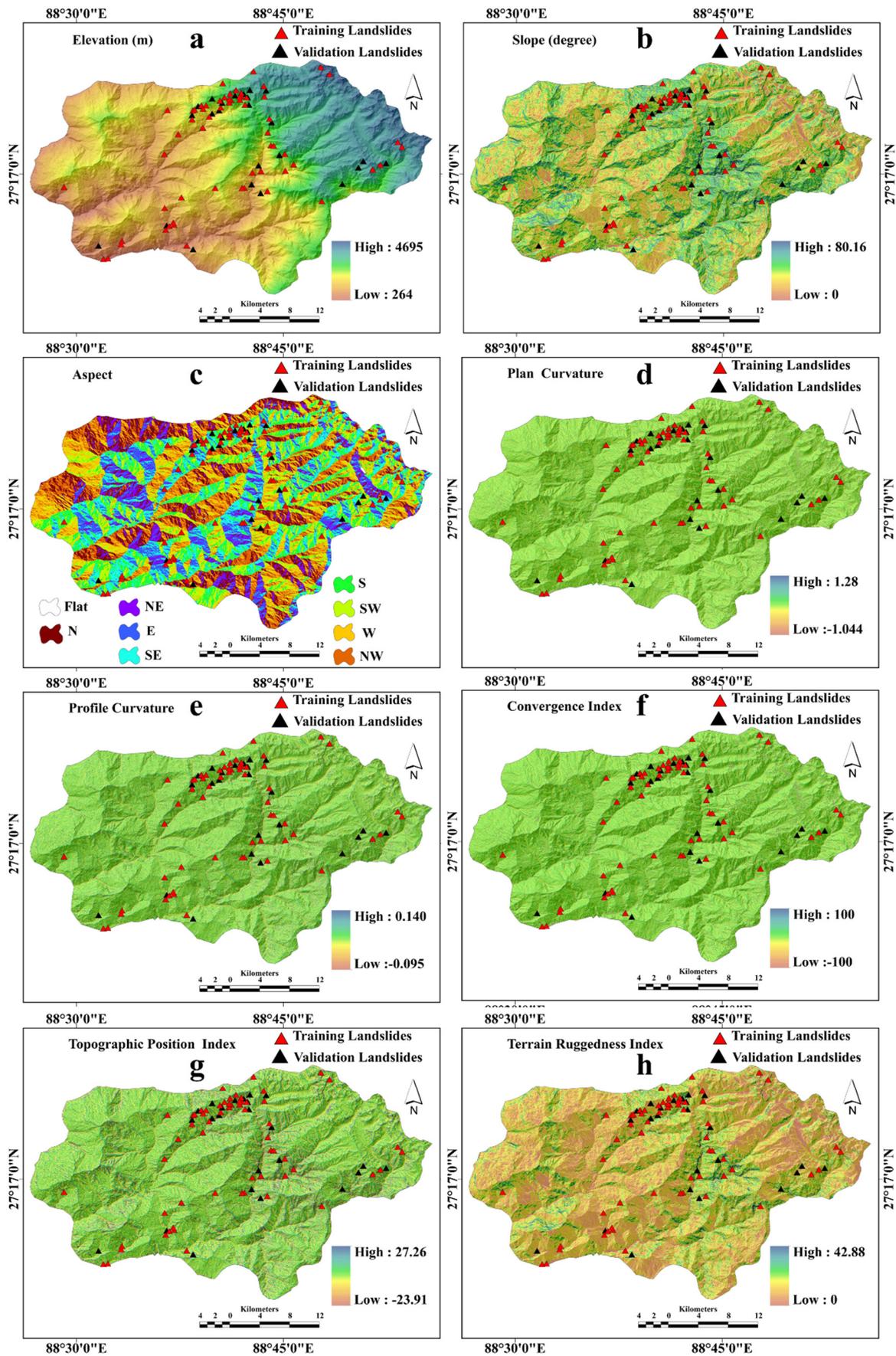
Figure 4. *Cont*.
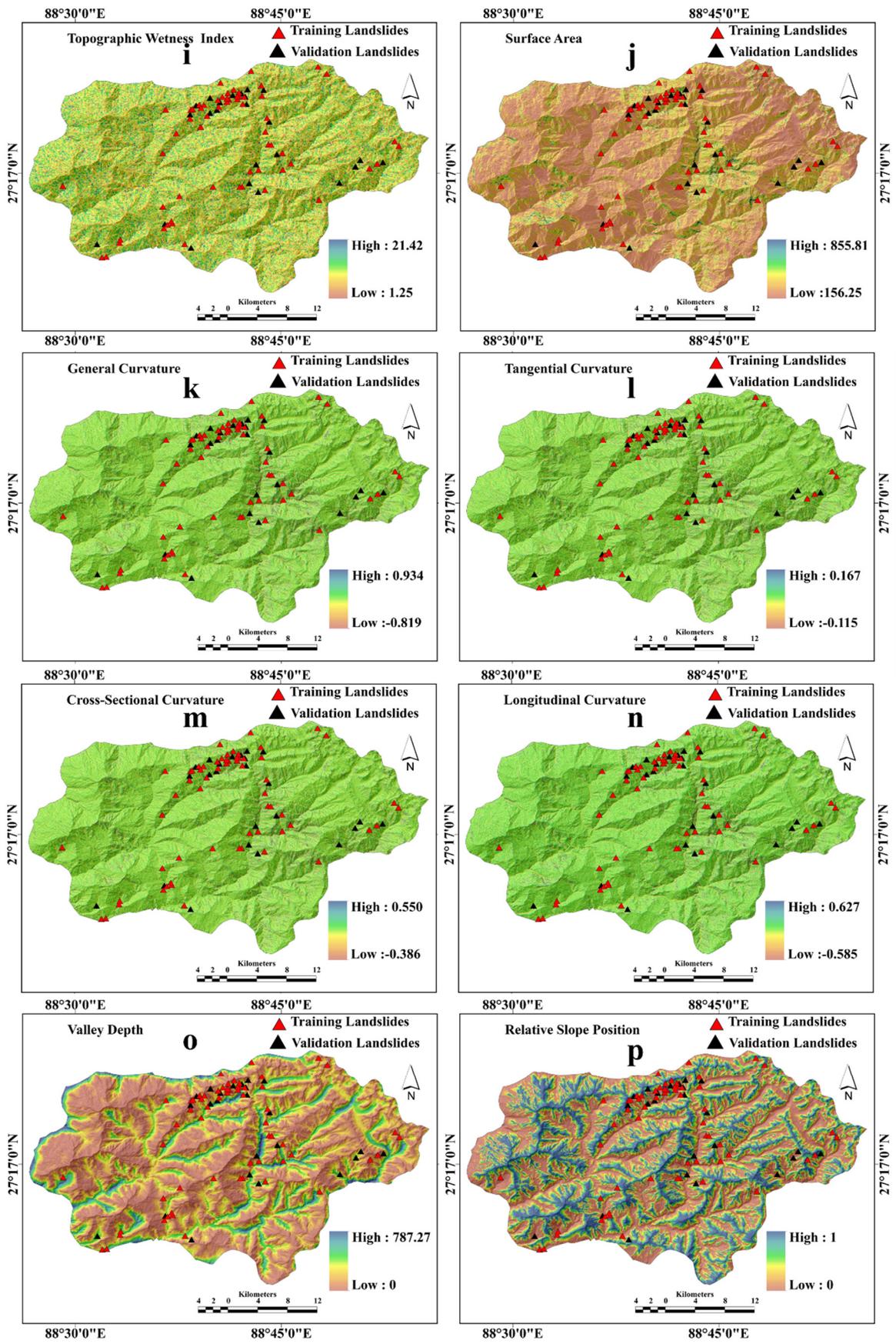
Figure 4. *Cont.*
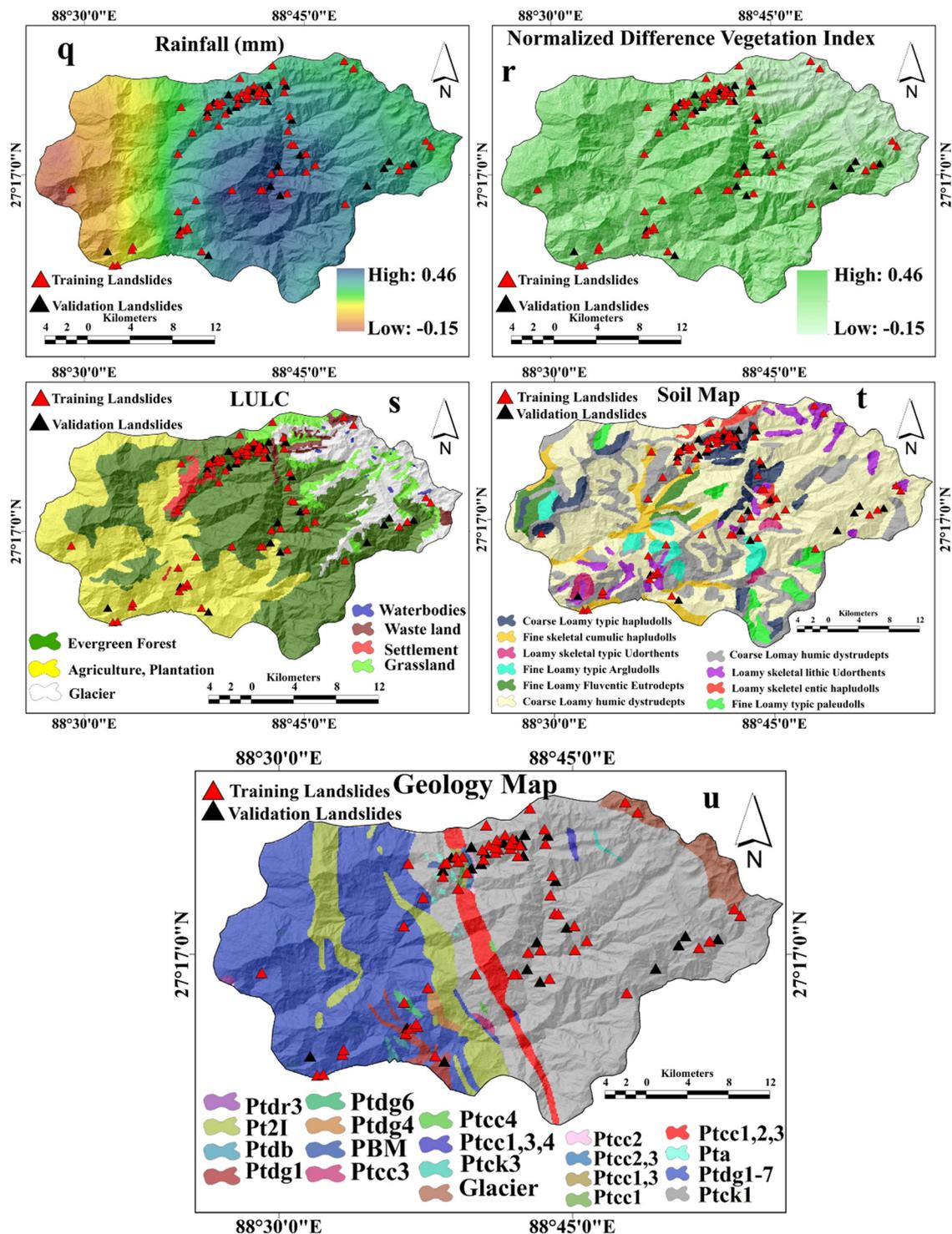
Figure 4. LCFs used in modelling landslide susceptibility: (**a**) elevation, (**b**) slope degree, (**c**) slope aspect, (**d**) plan curvature, (**e**) profile curvature, (**f**) convergence index, (**g**) topographic position index, (**h**) terrain ruggedness index, (**i**) topographic wetness index, (**j**) surface area, (**k**) general curvature, (**l**) tangential curvature, (**m**) cross sectional curvature (**n**) longitudinal curvature, (**o**) valley depth, (**p**) relative slope position, (**q**) rainfall, (**r**) NDVI, (**s**) LULC, (**t**) soil map and (**u**) geology map.

Cross section curvature was also introduced by Wood [42]. The spatial value of cross section curvature ranges from −0.386 to 0.550 (Figure 4m). The calculation of the surface area was conducted by the slope and slope aspects of a specific cell applying Berry's approach [44]. The surface area is the same as the planimetric area. The surface topography inside the cell is defined by its value. The value of the surface area ranges from 156.25 to

855.81 in the study area (Figure 4j). The convergence index is a terrain parameter showing the relief structure as a collection of convergent (channel) and divergent (ridge) areas (Equation (1)).

$$CI = \left(\frac{1}{8} \sum_{i=1}^{8} \theta_i \right) - 90° \tag{1}$$

where θ denotes the average angle between the aspect of adjacent cells and the direction to the central cell. In this field of analysis, the *CI* spatial value varies from -100 to 100 (Figure 4f). TWI represents the degree of wetness of the surface (Equation (2)). The TWI value ranges from 1.25 to 21.42 in the study area (4i).

$$\text{TWI} = \ln\left(\frac{As}{\tan a} \right) \tag{2}$$

where As is the upstream contributing area, and a is the slope gradient (in degrees). The TPI is the certain gap between the cell elevation value (Z_0) and the average surrounding cell elevation ($\overline{Z}$) (Equation (3)).

$$\text{TPI} = Z_0 - \overline{Z} \tag{3}$$

The TPI varies from -23.91 to 26.23 (Figure 4g), with positive values indicating that the cell is higher than its surrounding area, whereas negative values indicate that the cell is lower. TRI was defined by Riley et al. [45] to show the amount of elevation difference between adjacent cells in the digital elevation model (Equation (4)).

$$\text{TRI} = \sqrt{|X|(\max^2 - \min^2)} \tag{4}$$

where x is the elevation of each neighbour cell to a specific cell, and max and min are the largest and smallest elevations amongst the nine neighbouring pixels. VD is measured as the vertical distance to the base level of the channel network. The algorithm consists of two key steps: the interpolation of the base level of the network channel, and the subtraction of the base level from the original elevations. The value of VD ranges from 0 to 787.27 (Figure 4o). RSP is another important factor for determining the stability–instability of a land part (Figure 4p).

3.3.2. Other Environmental Factors

Rainfall is the triggering factor for the majority of landslides. The empirical relationship between landslides and rainfall type has been identified in many previous studies [46]. Five-year annual average rainfall data were used for the spatial mapping of rainfall by using the kriging method in GIS. The annual average rainfall is 2264 mm in this region (Figure 4q). The spatial map of geology (Figure 4u) was prepared with the digitisation process. Detailed geological descriptions are presented in Table 1. Ten soil categories were mapped for this study area (Figure 4t). Amongst them, coarse loamy humic dystrudepts associated with coarse loamy typic udorthents, and coarse loamy humicpachic dystrudepts associated with fine loamy type udorthents, are the dominant types of soil (Table 2). The LULC map was prepared by using Sentinel-2 satellite images through supervised classification determined by a maximum likelihood classification algorithm (Figure 4s). Water bodies (0.27%), wasteland (2.00%), settlement and built-up area (1.65%), grassland (5.47%), evergreen forest (47.18%), agriculture (31.92%) and plantation vegetation glacier (11.51%) are the major land use categories of the study area. The NDVI value ranges between -0.15 and 0.46 in the study area (Figure 4r) and indicates the concentration of vegetative cover. The NDVI of the study area was calculated by using Equation (5).

$$\text{NDVI} = \frac{Band5 - Band4}{Band5 + Band4} \tag{5}$$

3.4. Multicollinearity Assessment

A collinearity test must be performed between the LCFs because a linear association reduces the predictive precision of the model [47]. The multicollinearity test is an important step to investigate if a strong interrelationship may be found amongst the conditioning factors in multiple regression. In this work, a multicollinearity test was performed by applying the variance inflation factor (VIF) and tolerance (TOL) criteria to obtain the relevant LCFs [48]. The thresholds of the TOL and VIF values are ≤ 2 and ≥ 5, and values that exceed these thresholds suggest the existence of collinearity between or amongst the LCFs [48].

Table 1. Lithological units, characteristics, and their areal extension of East Sikkim.

Sl.No	Geo Unit	Description	Formation	Age	Area (km^2)	% of Area
1	Ptdr3	Verlegated cherty phyllite	Reyong Formation, Daling Group	Proterozoic	0.81	0.08
2	Pt2I	Granite gneiss (mylonitic)	Lingtse Gneiss	Meso-Proerozoic	78.08	8.10
3	Ptdb	Dolostone, orthoquartzite, purple phyllite/slat, chert	Boxa Formation, Daling Group	Proterozoic	2.08	0.22
4	Ptdg1	Interbanded chlorite-sericite schist/phyllite and quartzite	Gorubathan Formation, Daling Group	Proterozoic	10.02	1.04
5	Ptdg6	Mica schist with garnet with staurollite	Gorubathan Formation, Daling Group	Proterozoic	2.28	0.24
6	Ptdg4	Biotite phyllite/mica schist	Gorubathan Formation, Daling Group	Proterozoic	3.95	0.41
7	Pollymetallic base metal				3.10	0.32
8	Ptcc3	Calc-granullite (locally gneissic) with intercalculation of quartzite	Chungthang Formation, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	1.51	0.16
9	Ptcc4	Graphitic schist	Chungthang Formation, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	0.80	0.08
10	Ptcc1,3,4	1. Quartzite, 3. Calc-granullite (locally gneissic) with intercalculation of quartzite, 4. Graphitic schist	Chungthang Formation, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	2.20	0.23
11	Ptck3	Sillimanite granite gneiss	Kanchenjunga gneiss, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	1.09	0.11
12	Glacier				25.75	2.67
13	Ptcc2	Garnet kyanite sillimaite biotite schist/ gametiferous mica schist	Chungthang Formation, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	1.10	0.11
14	Ptcc2,3	2. Garnet kyanite sillimaite biotite schist/ gametiferous mica schist, 3.3. Calc-granullite (locally gneissic) with intercalculation of quartzite,	Chungthang Formation, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	2.02	0.21
15	Ptcc1,3	1. Quartzite, 3. Calc-granullite (locally gneissic) with intercalculation of quartzite,	Chungthang Formation, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	1.47	0.15
16	Ptcc1	1. Quartzite	Chungthang Formation, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	1.49	0.15
17	Ptcc1,2,3	1. Quartzite, 2. Garnet kyanite sillimaite biotite schist/gametiferous mica schist, 3. Calc-granullite (locally gneissic) with intercalculation of quartzite,	Chungthang Formation, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	29.81	3.09
18	Pta	amphibole schist/amphibolite	Basic Intrusive	Proterozoic	2.27	0.24
19	Ptdg1-7	1. Interbanded chlorite-sericite schist/phyllite and quartzite, 2. Metagreywacke (quartzo-feldspathic greywacke), 3. Pyritiferious black slate, 4. Biotite phyllite/schist, 5. Biotite quartzite, 6. Mica schist with garnet with/without staurolite, 7. Chlorite quartzite	Gorubathan Formation, Daling Group	Proterozoic	298.90	31.01
20	Ptck1	banded/ streaky migmatite, augen bearing (garnet) biotite gneiss with/without kynite silimanite with palaeosomes or staurolite, kynite, mica schist	Kanchenjunga gneiss, Central Crystalline Gneissic Complex (CCGC)	Proterozoic	495.27	51.38

Table 2. Characteristics of soil types and soil textural classes of East Sikkim.

Code	Soil Texture	Soil Types	Area (km^2)	% of Area
1	Coarse loamy humic pachic dystrudepts associated with fine loamy type udorthents	Inceptisols	202.42	21.00
2	Loamy skeletal lithic udorthents associated with rocks	Entisols	44.05	4.57
3	Loamy skeletal entic hapludolls associated with loamy skeletal type udorthents	Mollisols	11.69	1.21
4	Fine loamy typic paleudolls associated with fine loamy typic hapludools	Mollisols	27.77	2.88
5	Coarse loamy typic hapludols associated with coarse loamy entic hapludols	Mollisols	57.14	5.93
6	Fine skeletal cumulic hapludolls associated with coarse loamy typic Udorthents	Mollisols	49.91	5.18
7	Loamy skeletal typic udorthents associated with coarse loamy lithic dystrudepts	Entisols	10.75	1.12
8	Fine loamy typic argludolls associated with fine loamy cumic hapludolls	Mollisols	30.25	3.14
9	Fine loamy fluventic eutrodepts associated with loamy lithic haploudolls	Inceptisols	23.34	2.42
10	Coarse loamy humic dystrudepts associated with coarse loamy typic udorthents	Inceptisols	506.68	52.56

3.5. CSAE

CSAE method has been used to select the most predictive factors for LSM modelling. CSAE has been used in different artificial intelligence approaches to choose a small number of training datasets, to minimise the time and cost of the modelling process [49]. In this method, zero average merit (AM) values of conditioning factors denote the significance of less contribution to modelling. On the contrary, higher than zero average merit (AM) values indicate the most important and suitable conditioning factors for LSM modelling. The CSEA is calculated by using Equation (6).

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{6}$$

where E and O are the expected and observed values. The higher the value of CSAE of a given LCF indicates greater significance for occurrences of landslides [43].

3.6. Evaluation of Factor Importance by Using RR

RR was firstly developed by Hoerl and Kennard [50]. It belongs to a class of regression that uses L2 regularisation to reduce the issue of overfitting. RR was developed to avoid the extensive problems of instability and collinearity caused by the least square estimator [51]. RR predictions tend to be robust in the sense that minor variations in the data, in which the fitted regression is found, are minimally influenced. At times, the ridge calculated regression function can have strong estimates of the mean responses or forecasts of new observations for independent variable thresholds beyond the area of the observations where the regression function is built. The cost function is changed in RR by adding a penalty equal to the square of the coefficients' (w) size.

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \lambda \sum_{j=0}^{p} w_j^2 \tag{7}$$

Similarly, the cost function in Equation (8) should be minimised under the following conditions.

$$c > 0, \sum_{j=0}^{P} w_j^2 < c \tag{8}$$

Thus, RR constrains the coefficients (w). The penalty term regularises the coefficients so that the optimisation function is punished if the coefficients take high values. In this research, the R programming language 'caret' package (https://cran.r-project.org/web/packages/caret/caret.pdf, accessed on 20 December 2020) was used to determine the importance of LCFs by using RR.

3.7. DL and Tree-Based ML Models
3.7.1. CNN

DL approaches are computational frameworks that are based on NNs, inspired by human brains [18]. CNNs are widely applied in visual imagery processing [52]. They

are recognised as shift invariant or space invariant ANNs due to their shared-weight architecture and translation invariance functions [53]. CNN was first developed in 1980. To date, the CNN model has been widely used in classification and prediction applications in several branches, including the earth science discipline [54]. A CNN is made up of input data, an output layer, and numerous hidden layers. The hidden layers of a CNN usually consist of a sequence of convolutional layers (CLs) converging with a multiplication. The activation function is normally a rectified linear unit (ReLU) layer, and additional convolutions, such as pooling layers, fully connected layers and normalisation layers, are subsequently adopted. These layers are referred to as hidden layers because the activation function and final convolution mask their inputs and outputs. The input data of CNN are images and can be interpreted as images. CLs train the convolutions and give the best performance for data categorisation. PLs provide stable conversion, reduce overfitting and increase computational performance by reducing the number of structures resulting in the convolutions [55]. Using ReLU activation function, a ReLU increases the nonlinear properties of the network. Various techniques based on the form, image and purpose of data have been used by various researchers. The detailed description of these layers, primary parameters and how the CNN handles training data can be found in the literature of LeCun et al. [18]. One-dimensional image data are CNN input data used for optimal initialisation performance. A 1D input data, which contain various characteristics, can be converted into 2D images. The CNN's linearity is decreased by using a ReLU for all linked layers. With the aid of Equation (9), the loss value of the parameters is reduced by utilising the loss function.

$$Loss = -\frac{1}{m}\sum_{i=1}^{m}[x_i \log(z_i) + (1 - x_i)\log(1 - z_i)] \tag{9}$$

where x_i and z_i represent the predicted and true levels of the ith sample, respectively, and m denotes the total number of landslide data. The variables are updated regularly until the loss value converges. In the current study, this approach is used for susceptibility mapping of landslides because images are larger, no data are classified, and analysis is continuous. The CNN structures of the present study are shown in Figure 5. In this study, 3 convolution kernels, 2 pooling kernels, ReLU activation, 600 epochs, AdaGrad and a 0.001 learning rate are used for modelling the landslide by using the CNN DL method.

3.7.2. ANN

ANNs have been extensively used in classification operation [56]. The three main structures of ANN are input layers, hidden layers and output layers (Figure 6). The input layers are LCFs, the outcome layers reflect the outcomes of model predictions as a landslide or non-landslide, and the hidden layers are the classifier layers that process and convert the data from input to output. In ANN, no presumptions are required for the training datasets. Determining the relative value of the different input measures is not needed, and most input measurements are chosen on the basis of weight change during the training process [57]. ANN is constructed with two main steps [58].

(i)   The inputs are extended ahead through the hidden layers to estimate the difference and to generate the output values. The output values are compared with the pre values.

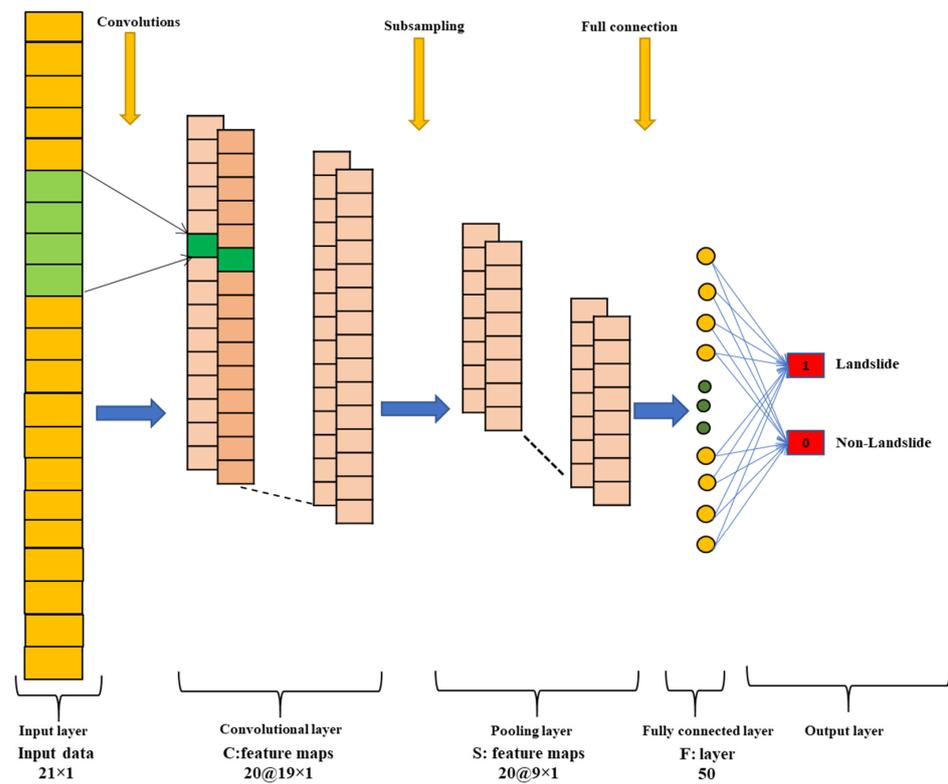(ii)   The connection weights are adjusted to optimise the best results for the least variation.
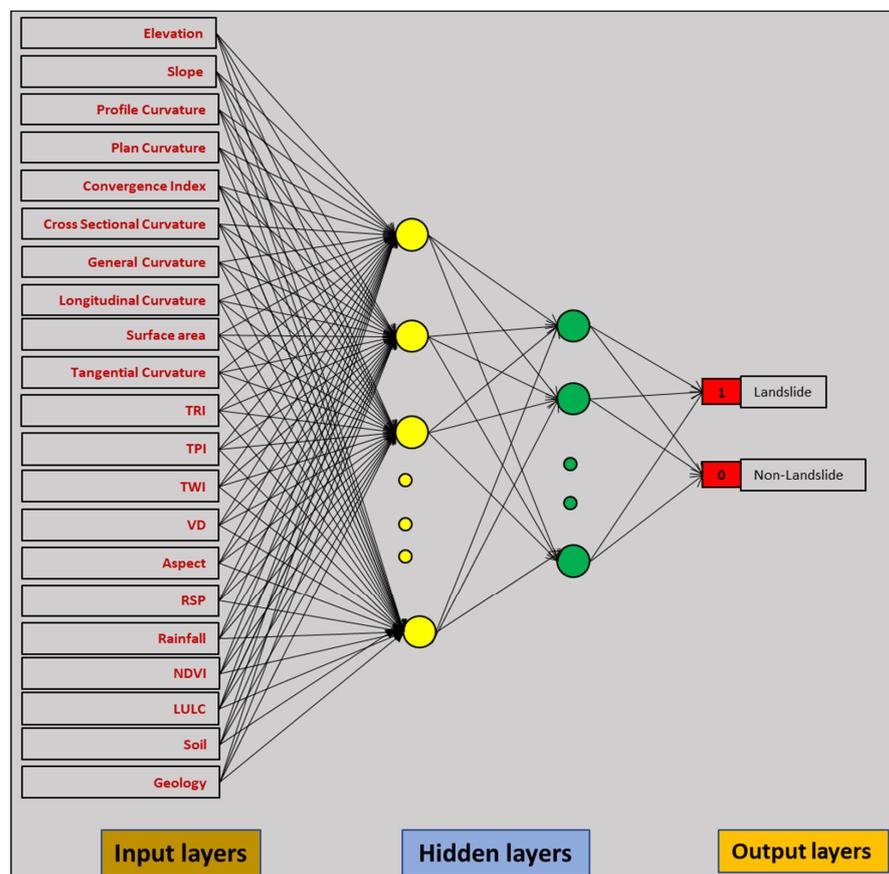
Figure 5. Structure of the CNN model.

Figure 6. Structure of the ANN model for the present study.

Let $x = x_i, i = 1, 2, ...21$ the vector of the 21 LCFs, $y = 1$ or 0 indicates the landslide and non-landside class. A nonlinear sigmoid feature is frequently used to the weighted sum of input data until the data are passed to the next step. ANN function for classification is calculated by using Equation (10).

$$y = f(x) \tag{10}$$

where *f(x)* WW shows the hidden functionality that is enhanced during the modelling by the adjustable weights to produce the ANN network structure. The sum of square difference between the actual and expected result of neuron, that is, known as error, is used to select the number of hidden layers and neuron in ANN. In each neuron, weight is adjusted to reduce the error during the training process. In the present analysis, the trial process was used to eliminate the overfitting and create the ANN model with 2 hidden layers, 500 epochs and 20 number validation thresholds.

3.7.3. ADTree

An ADTree is a classification ML strategy that generalises DTs and links to boosting. Freund and Mason [59] introduced this technique. An ADTree is an alternation of decision nodes describing a predicate position, and prediction nodes containing a single number. An instance is categorised by an ADTree by following all ways that all decision nodes are valid and by combining any estimation nodes. This varies from binary classification trees, such as CART or C4.5, in which an instance takes only one direction through the tree. The ADTree model is more accurate in the event of a classification task than other tree models [60]. Splitter node and prediction nodes are the two main types of ADTree model nodes. The function of the splitter node is that it classifies data in accordance with the selected attributes. The prediction node is the number score used to forecast [60]. The basic principle that maps from instances to actual numbers is a prediction d_1, a base situation d_2, with the result being p or q which are two real numbers. The prediction is p when $d_1 \cap d_2$ or q when $d_1 \cap^- d_2$. The values of p and q are calculated by using Equations (11) and (12).

$$p = \frac{1}{2} In \frac{W_+(d_1 \cap d_2)}{W_-(d_1 \cap d_2)} \tag{11}$$

$$q = \frac{1}{2} In \frac{W_+(d_1 \cap^- d_2)}{W_-(d_1 \cap^- d_2)} \tag{12}$$

where the best d_1 and d_2 are selected by minimising $Z_t(d_1, d_2)$ and is defined as Equation (13).

$$Z_t(d_1, d_2) = 2\sqrt{W_+(d_1 \cap d_2)W_-(d_1 \cap d_2)} + \sqrt{W_+(d_1 \cap^- d_2)W_-(d_1 \cap^- d_2)} + W(d_2) \tag{13}$$

Assuming M is the base rule, then a new rule can be defined as $M_t + 1 = M_t + r_t$. $r_t(x)$ which shows the two prediction values (p and q) at each layer of the tree, and x is a set of instances. The grouping can be used as an indication that accumulates prediction values in $M_t + 1$ such as:

$$Class(x) = sign\left(\sum_{t=1}^{T} r_t(x)\right) \tag{14}$$

A set of routes is constructed for each pixel in the training dataset. When a path hits a decision node, it will continue with the offspring that corresponds to the decision outcome, but if the path encounters a prediction node, it will continue with all of the node's offspring [59]. A pixel's susceptibility index is calculated by adding all the values from any prediction nodes found as the pixel filters down through all applicable branches. A parameter that must be determined is the number of boosting iterations. Larger trees and overfitting may arise from a large number of boosting iterations, whereas a minimal number of boosting iterations may result in a small tree with poor performance. The

classification accuracies of the training and validation datasets were estimated by changing from 1 to 14. The model was prepared by using the RWeka package in R environment.

3.7.4. CART

CART is a rule-based algorithm that constructs a binary tree by binary recursive partitioning. Binary recursive partitioning is a method that partitions a node into a yes/no response. The heterogeneity within each resultant subset is reduced on the basis of a single factor and the rule generated for each phase, which divides them depending on the various relationships of each division. Landslide susceptibility mapping using the CART technique has been used in several studies [61]. A "terminal" node's expected value is considered the average of the answer values in that node [62]. The predictor variables are extremely simple and can be comprised of different types: numeric, binary and categorical types. The model's results are not affected by monotonous transformations and different measurement scales between predictors. In regression trees, independent variables are insensitive to outliers and use surrogates to manage missing data [62]. The hierarchical structure of a regression tree indicates that the response to one input vector relies on higher input variables in the tree to model relationships between predictors automatically. Regression trees typically lead to an overcomplex decision tree where only the most relevant knowledge, that is, the nodes that illustrate the largest amount of deviance, needs to be 'pruned' to communicate [61]. CART, similar to other DT algorithms, does not need the identification of independent variables in advance because the most relevant variables are discovered during the selection of the optimal splitting characteristic in each node [62]. Thus, CART is appropriate for issues in which the correlation between input and output parameters is unknown in advance, making the CART model's outputs interpretable [61]. Depending on whether the output is qualitative or quantitative, CART can be used to solve classification and regression issues. CART is used in this study as a classifier for landslide susceptibility. "tree" package in R-studio was used for preparing the CART model.

3.7.5. FTree

FTrees were introduced by Gama [63]. FTree integrates a discriminating function and multivariate decision trees through constructive induction [63]. It is regarded to be a generalisation of multivariate trees. For learning classification trees, the FTree model adjoins with attributes at leaf nodes, decision node and leaves [63]. Decision nodes are built as the tree grows, and functional leaves are built as the tree prunes [63]. For prediction evaluation, the FTree can predict the value of dependent factors from the unclassified samples. The sample travels from the root node into a leaf over the tree in which the constructor function built at the node is used for collecting the attributes of the sample. Subsequently, the node's decision test is used to define the path by which the sample will take. With the help of the constructor function built on leaf and leaf-related constant, the sample is classified as leaf [63].

Classification tree nodes are created by comparing the values of specific input characteristics to a constant. Three important parts of the FTree are: the regression model (RM), which is utilised in FTree for internal nodes and leaves; for inner nodes, inner FTree uses RMs; and for leaves, FTree leaves utilise RM. FTree leaves were used in this study. FTree applied the gain ratio as a criterion for splitting to determine the input attribute to separate. To avoid overfitting, C4.5 pruning was used, and logit boost (iterative reweighting) was applied to resemble the leaves with the least squares for each class in accordance with the logistic regression functions [63].

3.7.6. LMT

LMT is a classification model in computer science that integrates logistic regression with DT learning, with the related supervised training algorithm [64]. The earlier concept of a model tree is used in logistic LMT. A DT on its leaves uses linear regression (LR) models where a piecewise constant model is generated by ordinary DTs with constants on their

leaves [64]. This process is performed to obtain a piecewise linear regression model. The LogitBoost algorithm is used in the logistic version to generate a logistic regression model at each tree node. The model uses cross-validation for searching the multiple LogitBoost iterations to control overfitting of training data. For each Mi class, the LogitBoost model utilises least-squares fitting additive logistic regression, and the later likelihood of leaf nodes is measured by LR [65].

$$L_M(x) = \sum_{i=1}^{n} \beta_i x_i + \beta_0 \tag{15}$$

where β_i is the coefficient of the ith element of vector x, n is the total factors, and D is the total classes.

In the LMT model, the posterior probabilities of leaf nodes were computed by using the linear logistic regression technique [64].

$$P(M|x) = \frac{\exp(L_M(x))}{\sum_{M=1}^{D} \exp(L'_M(x))} \tag{16}$$

3.8. Validation Methods

LSMs produced using CNN, ANN, CART, ADTree, FTree and LMT models were evaluated through 21 statistical measures mentioned in Table 3. The ROC curve is constructed with the true positive rate (TPR) versus the false positive rate (FPR). The TPR consists of landslide cells, which indicate the most susceptibility. The FPR consists of non-landslide cells, which indicate the non-susceptibility. If the area under the ROC curve remains extremely close to 100%, then the model reflects higher goodness-of-fit and excellent accuracy [66]. The AUC is indicative of the consistency of a calculated model [67]. The ROC curve consists of four elements, being, true positive (A), false positive (B), true negative (C) and false negative (D) [68] (Equations (17) and (18)).

$$TPR = A/P \tag{17}$$

$$FPR = B/N \tag{18}$$

where *FPR* and *TPR* are the false positive rate and true positive rate. The AUC is computed by using Equation (19).

$$AUC = \left(\sum A + \sum C \right) / (P + N) \tag{19}$$

where P and N are the total number of landslides and non-landslides, respectively.

Amongst the statistical measures, lower values of false negative rate (miss rate) and misclassification rate designate a higher model accuracy. By contrast, higher values of the remaining measures denote high model accuracy. The computation methodology of these parameters is presented in Table 3. These various measures are computed to omit the statistical error of using a single dimension. Based on these measures, prioritisation of models in terms of accuracy is conducted by using the compound factor (*CF*) method to determine the hierarchy of best-fit for this study. In the *CF* method, consecutive rank is given in accordance with the relevance of the factors targeting the goal. The mean values of the assigned ranks of each factor represent their relative priority [69]. *CF* method can be formulated as:

$$CF = \sum_{i=1}^{Fn} (R) \tag{20}$$

where rank is represented by R of the factor, and F_n is the number of factors.

Table 3. Formulas for the computation of statistical accuracy measures.

| Measures	Formula	References
TPR or sensitivity	$\frac{A}{P} = \frac{A}{A+D}$	[66]
FPR or fall-out or 1-specificity	$\frac{B}{N} = \frac{B}{B+C} = 1 - \frac{C}{C+B}$	[66]
TNR or specificity	$\frac{C}{N} = \frac{C}{C+B}$	[67]
Miss rate	$\frac{D}{P} = \frac{D}{D+A} = 1 - TPR$	[70]
Accuracy	$\frac{A+C}{T}$	[67]
Misclassification rate	$\frac{B+D}{T}$	[70]
PPV or precision	$\frac{A}{A+B}$	[71]
False discovery rate (FDR)	$\frac{B}{B+A} = 1 - PPV$	[71]
Negative predictive value (NPV)	$\frac{C}{C+D}$	[71]
False omission rate (FOR)	$1 - NPV = \frac{D}{D+C}$	[71]
F-score	$\frac{2A}{2A+B+D}$	[71]
Matthews correlation coefficient (MCC)	$\frac{(A \times C) - (B \times D)}{\sqrt{(A+B)(A+D)(C+B)(C+D)}}$	[71]
Bookmaker informedness (BM)	$TPR + TNR - 1$	[71]
Markedness (MK)	$PPV + NPV - 1$	[71]
Threat score (TS)	$\frac{A}{A+D+B}$	[71]
Equitable threat score	$\frac{A - A_{random}}{A+D+B-A_{random}}$ where, $TP_{random} = \frac{(A+D)(A+B)}{T}$	[71]
True skill statistics (TSS)	$\frac{A}{A+B} - \frac{B}{B+C} = Sensitivity + Specificity - 1$	[71]
Heidke's skill score	$\frac{A+C-E}{T-E}$ where, $E = \frac{1}{T}[(A+D)(A+B) + (C+D)(C+B)]$	[71]
Odd ratio skill score (Yule's Q)	$\frac{(A \times C) - (B \times D)}{(A \times C) + (B \times D)}$	[71]
Cohen's kappa	$\frac{(A+C) - [(A+D)(A+B) + (D+C)(B+C)]/T}{T - [(A \times C))(A+B) + (D+C)(B+C]/T}$	[71]

4. Results

4.1. Analysis of Multicollinearity

TOL and the VIF were used in this study to test the problem of collinearity amongst the LCFs. The derived tolerance values were >0.2, and the VIF values were less than <5, thereby confirming the nonexistence of linearity between the factors. The outcome of the linearity test is shown in Table 4. A total of 21 LCFs were selected after the analysis (Table 4). The highest TOI value and lowest VIF were derived from NDVI (Table 4).

4.2. Results of CSAE

The detailed results of CSAE are shown in Figure 7. Amongst the factors, rainfall (AM = 94.52) was identified as the most important predictive factor, followed by the slope aspect, elevation, NDVI, LULC, VD, soil map, TRI, slope, surface area, TPI, geology, RSP, cross sectional curvature, general curvature, longitudinal curvature, CI, tangential curvature, profile curvature, plan curvature and TWI (Figure 7).

Table 4. Collinearity statistics of LCFs.

Factors	Collinearity Statistics	
	Tolerance	VIF
Elevation	0.54	1.85
Slope	0.31	4.57
Profile curvature	0.21	4.74
Plan curvature	0.40	2.48
Convergence index	0.26	3.73
Cross sectional curvature	0.21	4.76
General curvature	0.23	4.35
Longitudinal curvature	0.23	2.19
Surface area	0.31	3.28
Tangential curvature	0.22	4.55
TRI	0.92	1.09
TPI	0.31	3.19
TWI	0.86	1.17
Valley depth (VD)	0.34	2.97
Aspect	0.81	1.23
Relative slope position (RSP)	0.59	1.69
Rainfall	0.84	1.19
NDVI	0.93	1.07
LULC	0.91	1.10
Soil map	0.83	1.17
Geology	0.71	1.90
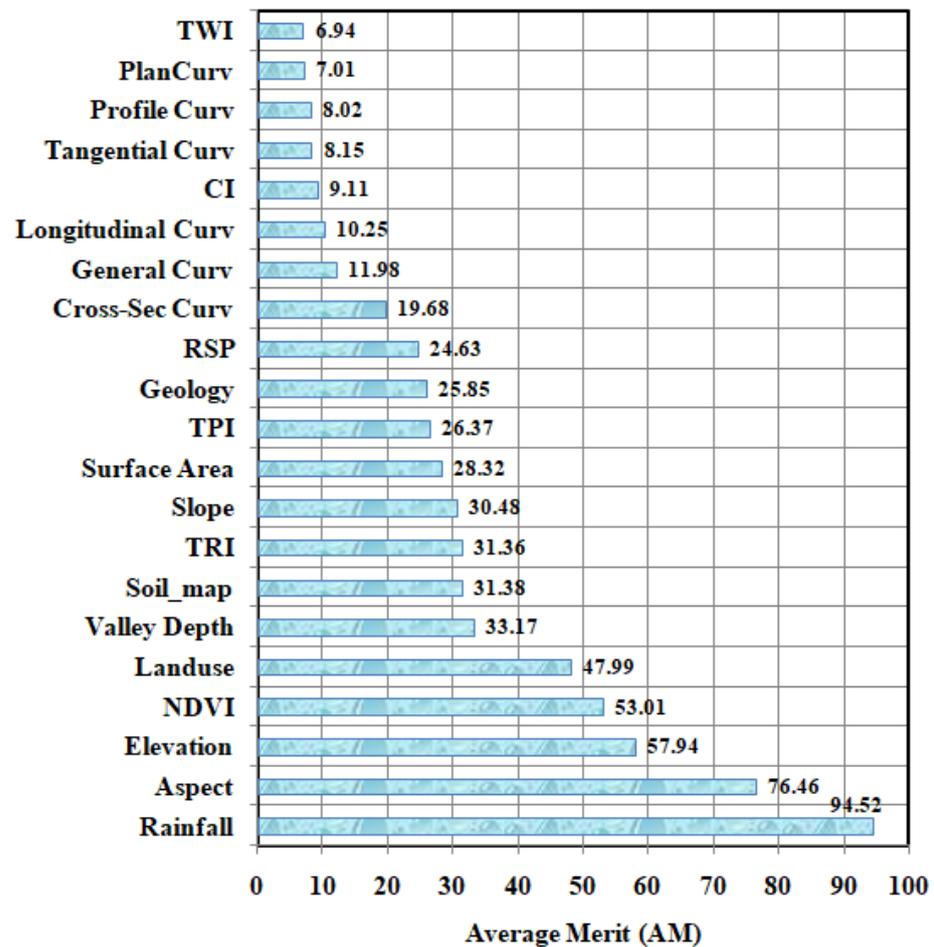
Figure 7. Average merit of the LCFs derived through CSAE method.

4.3. LSMs

LSMs were prepared by using DL and novel tree ML approaches, such as CNN, ANN, CART, ADTree, CART, FTree and LMT tree models. The LSMs are shown in Figure 8. The LSMs were divided into five susceptibility classes by using the natural break method of Jenk, as depicted in Figure 9. The LSM of CNN model assumed that the very high susceptibility class occupied an area of 23.02 and 14.40% under high, 23.07% under moderate, 24.34% under low, and 15.17% under very low susceptibility level of the study area (Figure 9). In the case of the ANN model, the very high, high, medium, low, and very low susceptibility zones encompassed the area of 13.36%, 13.22%, 19.55%, 31.53% and 22.34%, respectively (Figure 9). In the case of the tree ML models, the results were relatively different. For instance, the LSM of ADTree model assumed only 11.80% of the district under very high susceptibility zone and 20.86% under very low, 28.33% under low, 23.83% under moderate, and 15.18% under high susceptibility classes (Figure 9). The percentage distributions of the susceptibility classes of LSM of the CART model for very high, high, moderate, low, and very low classes were 18.59, 28.38, 31.63, 18.17 and 3.23%, respectively (Figure 9). The FTree model predicted 4.77% of the total area as very low, 12.18% as low, 20.70% as moderate, 27.44% as high, and 34.90% as very high susceptibility zones of landslides (Figure 9). The LSM of the LMT model categorised 6.54, 16.71, 26.05, 26.65 and 24.05% of the district under very low, low, moderate, high and very high susceptibility zones (Figure 9).
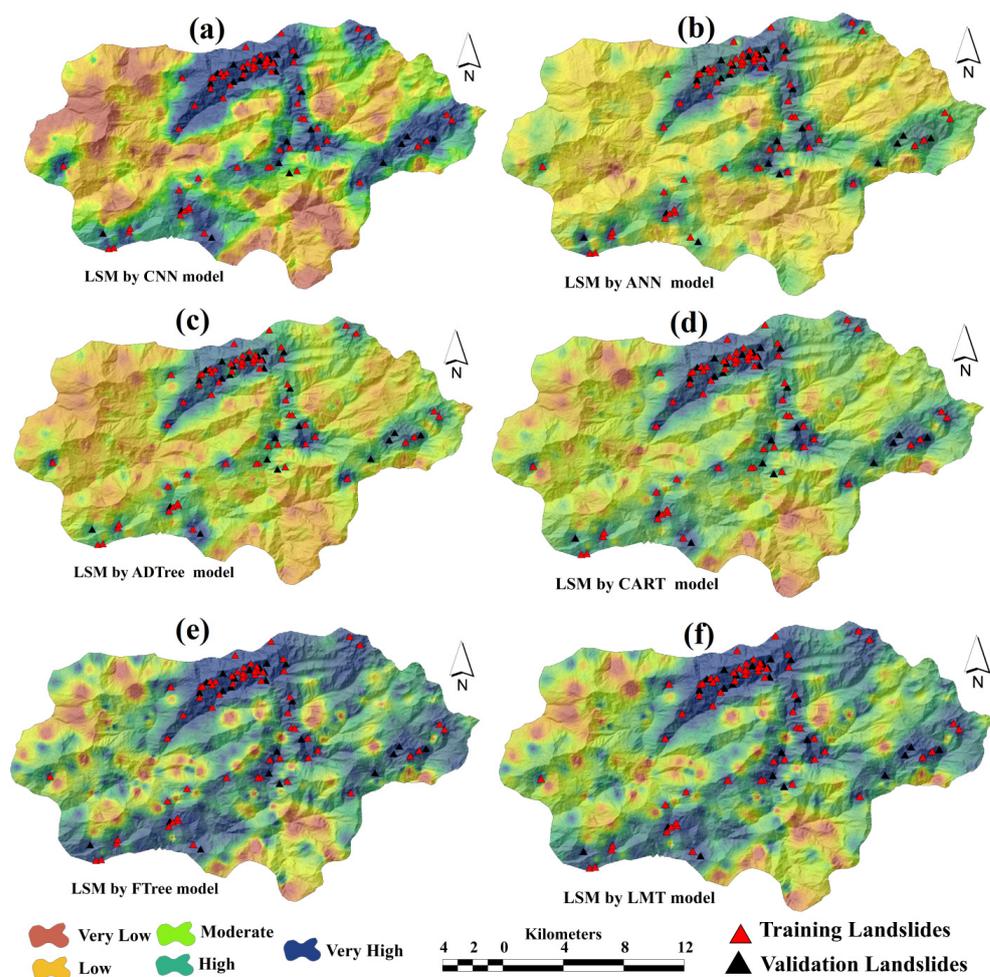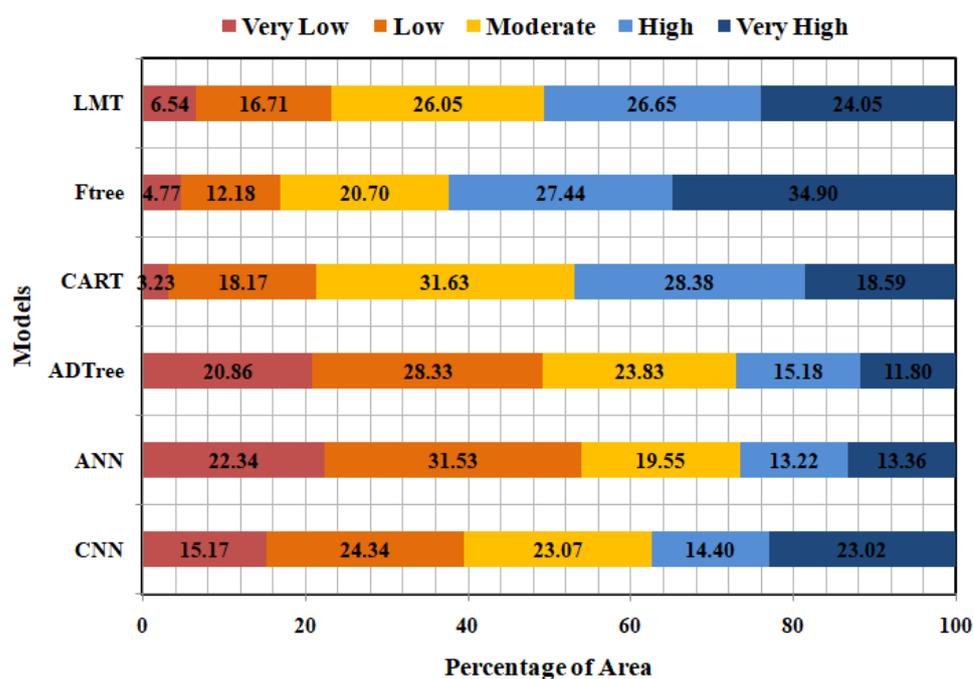

Figure 8. LSMs produced by the ML models: (**a**) CNN, (**b**) ANN, (**c**) ADTree, (**d**) CART, (**e**) FTree and (**f**) LMT.

Figure 9. Distribution of susceptible areas (in %) produced in different LSM models.

4.4. Importance Analysis of LCFs by RR

A total of 21 LCFs were tested through RR to analyse their importance for landslide modelling. The outcome of RR revealed that rainfall (RR = 0.377) had the highest predictive capability in this study. Comparatively, other factors, such as elevation (RR = 0.077), slope (RR = 0.153), aspect (RR = 0.161), plan curvature (RR = 0.021), profile curvature (RR = 0.131), general curvature (RR = 0.089), tangential curvature (RR = 0.135), longitudinal curvature (RR = 0.156), cross-sectional curvature (RR = 0.004), surface area (RR = 0.256), topographic position index (RR=0.016), topographic wetness index (RR = 0.030), terrain ruggedness index (RR = 0.247), valley depth (RR = 0.176), relative slope position (RR = 0.057), geology (RR = 0.081), soil (RR = 0.140), LULC (RR = 0.214) and NDVI (RR = 0.282) played positive and significant roles for the modelling of landslide susceptibility in this research (Figure 10).

4.5. Validation and Accuracy Assessment

The predictability of the models was validated by using the AUC of ROC and several other statistical measures (Table 5). The success rate curve (using the training dataset) was drawn for each model. The result showed that CNN was the best fit with an AUC of 0.918, followed by CART (AUC = 0.910), LMT (AUC = 0.905), FTree (AUC = 0.900), ANN (AUC = 0.843) and ADTree (AUC = 0.745). The predictive capability of the models was assessed by using the prediction rate curve, which provided a similar result to the success rate curve. The AUC for the CNN model under the prediction rate curve was 0.933 and 0.925 for CART, 0.920 for LMT, 0.910 for FTree, 0.889 for ANN and 0.785 for ADTree model (Figure 11). Table 5 contains the summary statistics of other measures. The outcome of CF-based prioritisation provided the relative priority ranking of models by considering all the exactness metrics calculated using the training and validation data sets. Using the training dataset, most priority was assigned to the CNN model because it ranked 1, followed by CART (rank 2), LMT (rank 3), ANN model (rank 4), FTree (rank 5) and ADTree (rank 6) model. In the validation dataset, the result was the same as the training set (Table 5).
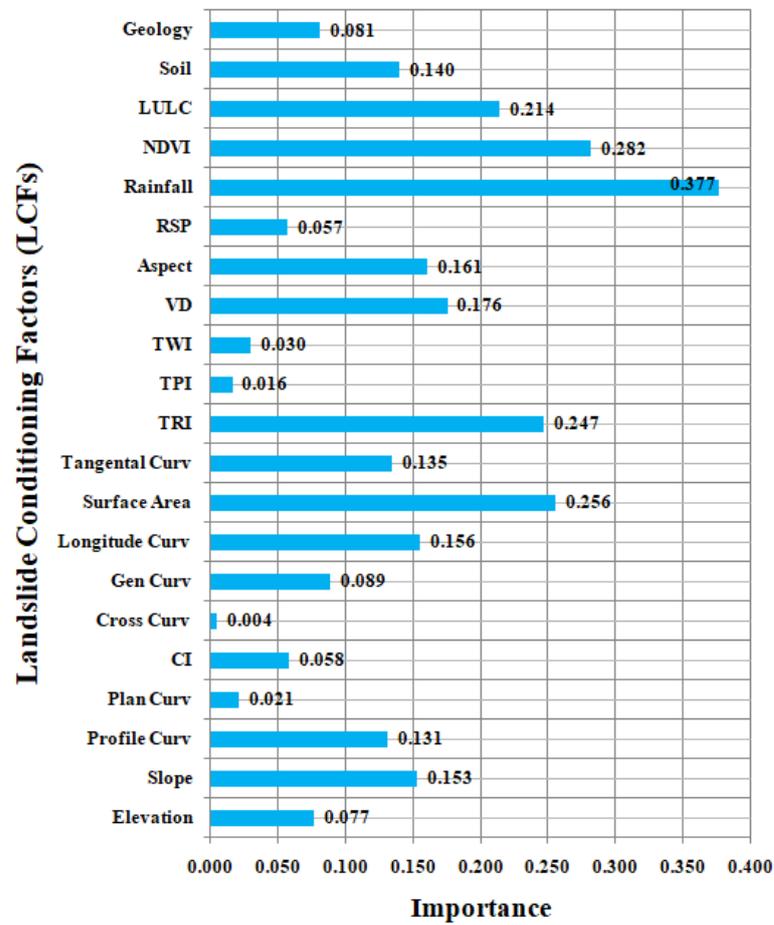
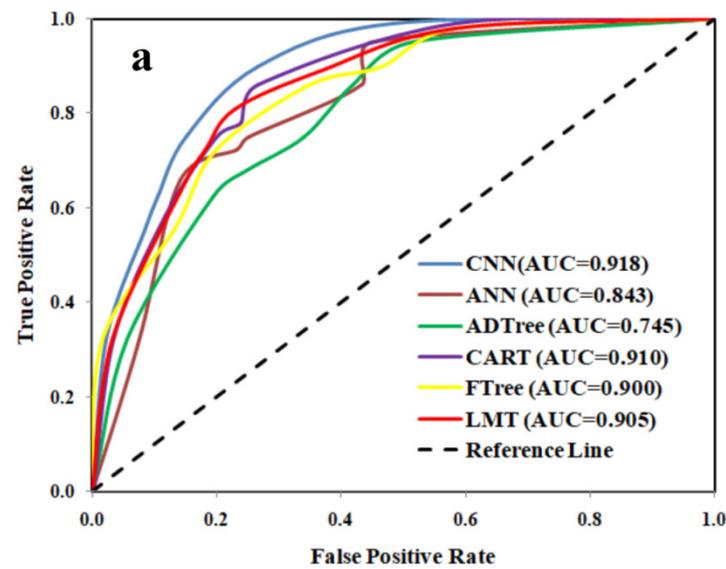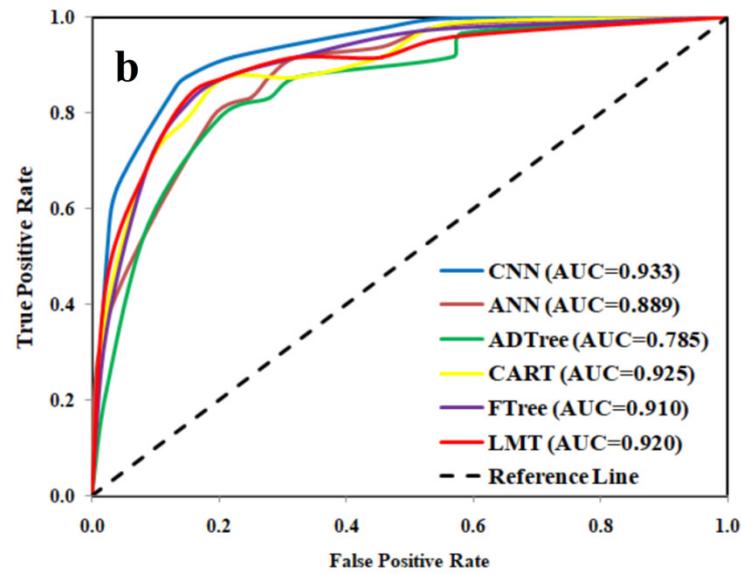**Figure 10.** Coefficient values of RR showing the importance of LCFs in LSM.

Figure 11. *Cont.*

Figure 11. Area under the success rate curve and prediction rate curve for different models: (**a**) training and (**b**) testing.

Table 5. Accuracy statistics and prioritisation of models.

	Using Training Dataset											
Criteria	**Results**						**Rank**					
	CNN	**ANN**	**ADTree**	**CART**	**FTree**	**LMT**	**CNN**	**ANN**	**ADTree**	**CART**	**FTree**	**LMT**
TPR	0.9	0.75	0.71	0.84	0.76	0.79	1	5	6	2	4	3
FPR	0.07	0.13	0.23	0.14	0.15	0.12	1	3	6	4	5	2
efficiency	0.92	0.8	0.73	0.85	0.8	0.83	1	4	5	2	4	3
TSS	0.83	0.63	0.48	0.7	0.61	0.68	1	4	6	2	5	3
TNR	0.93	0.88	0.77	0.86	0.85	0.88	1	2	5	3	4	2
FNR	0.1	0.25	0.29	0.16	0.24	0.21	1	5	6	2	4	3
Misclassification rate	0.08	0.2	0.27	0.15	0.2	0.17	1	4	5	2	4	3
PPV	0.93	0.9	0.8	0.87	0.87	0.9	1	2	4	3	3	2
FDR	0.07	0.1	0.2	0.13	0.13	0.1	1	2	4	3	3	2
NPV	0.9	0.7	0.67	0.83	0.73	0.77	1	5	6	2	4	3
F-score	0.92	0.82	0.75	0.85	0.81	0.84	1	4	6	2	5	3
Matthews correlation coefficient (MCC)	0.83	0.61	0.47	0.7	0.61	0.67	1	4	5	2	4	3
BM	0.83	0.63	0.48	0.7	0.61	0.68	1	4	6	2	5	3
MK	0.83	0.6	0.47	0.7	0.6	0.67	1	4	5	2	4	3
Threat score (TS)	0.85	0.69	0.6	0.74	0.68	0.73	1	4	6	2	5	3
Odd ratio skill score	0.98	0.91	0.78	0.94	0.89	0.93	1	4	6	2	5	3
Equitable threat score	0.71	0.44	0.3	0.54	0.43	0.5	1	5	6	2	4	3
Heidke's skill score	0.21	0.18	0.16	0.19	0.18	0.19	1	3	4	2	3	2
Cohen's Kappa	0.83	0.6	0.47	0.7	0.61	0.67	1	5	6	2	4	3
AUC	0.918	0.843	0.745	0.91	0.9	0.905	1	5	6	2	4	3
	Rank total						20	78	109	45	83	55
	CF						1.00	3.93	5.45	2.25	4.15	2.75
	Priority rank						1	4	6	2	5	3
	Using Validation Dataset											
TPR	0.86	0.71	0.64	0.79	0.71	0.79	1	3	5	2	4	2
FPR	0.08	0.25	0.33	0.17	0.31	0.17	1	3	5	2	4	2
efficiency	0.38	0.32	0.28	0.35	0.32	0.35	1	3	4	2	3	2
TSS	0.77	0.46	0.31	0.62	0.41	0.62	1	3	5	2	4	2
TNR	0.92	0.75	0.67	0.83	0.69	0.83	1	3	5	2	4	2
FNR	0.14	0.29	0.36	0.21	0.29	0.21	1	3	4	2	3	2
Misclassification rate	0.12	0.27	0.35	0.19	0.3	0.19	1	3	5	2	4	2
PPV	0.92	0.77	0.69	0.85	0.71	0.85	1	3	5	2	4	2
FDR	0.08	0.23	0.31	0.15	0.29	0.15	1	3	5	2	4	2
NPV	0.85	0.69	0.62	0.77	0.69	0.77	1	3	4	2	3	2
F-score	0.89	0.74	0.67	0.81	0.71	0.81	1	3	5	2	4	2
MCC	0.77	0.46	0.31	0.62	0.41	0.62	1	3	5	2	4	2
BM	0.77	0.46	0.31	0.62	0.41	0.62	1	3	5	2	4	2
MK	0.77	0.46	0.31	0.62	0.41	0.62	1	3	5	2	4	2
Threat score (TS)	0.8	0.59	0.5	0.69	0.56	0.69	1	3	5	2	4	2
Odd ratio skill score	0.97	0.76	0.57	0.9	0.7	0.9	1	3	5	2	4	2
Equitable threat score	0.63	0.3	0.18	0.44	0.26	0.44	1	3	5	2	4	2
Heidke's skill score	0.09	0.08	0.07	0.08	0.07	0.08	1	2	3	2	3	2
Cohen's Kappa	0.77	0.46	0.31	0.62	0.41	0.62	1	3	5	2	4	2
AUC	0.933	0.889	0.785	0.925	0.91	0.92	1	5	6	2	4	3
	Rank total						20	61	96	40	76	41
	CF						1.00	3.05	4.80	2.00	3.80	2.05
	Priority rank						1	4	6	2	5	3

5. Discussion

In this research, DL and tree ML models were chosen for the spatial susceptibility assessment of landslides in the East Sikkim region. A comparative analysis was performed with other state-of-the-art DL models and tree ML models. The DL model, namely, CNN, and ML models, such as ANN, ADTree, CART, FTree and LMT, were used to produce LSMs of the East Sikkim region. The results showed that the CNN model had the highest goodness-of-fit and excellent predictive capability, followed by ANN, ADTree, CART, FTree and LMT models. Several studies related to landsides have emphasised the different modelling approaches, such as ML, DL and ensemble ML. The DL and ML approaches have been considered as robust and efficient tools and have been used in different fields of geographical research, geotechnical application, and natural hazards, including LS mapping [72,73]. Wang et al. [74] applied DL and ML models, such as logistic regression, SVM and RF models, for LS assessment. Their research proved that CNN had the highest peak performance of predictive modelling, followed by the ML models. Moosavi et al. [75] compared the ANN and SVM by using pixel-based methods to produce the predictive model of landslides. CNN and texture shift recognition with pre and post landslide optical images for automated landslide detection were introduced by Ding et al. [54]. Ghorbanzadeh et al. [76] applied DL and ML approaches, such as ANN, SVM, RF and CNN models, to predict the most landslide-affected areas in Rasuwa District, Nepal. CNN performs several convolution and pooling operations to retrieve the characteristics. These features become more sophisticated and more abstract with the growth of convolutions and pooling. The degree of susceptibility to landslides, which was the deciding factor in assessing the susceptibility to landslides, is represented in these abstract features. CNN decreases the number of weights that need to be trained and the numerical complexity of the network [77]. The main advantages of the CNN model are that it considers all neighbourhood information and can determine manifold stages of representations from input data [77]. In agreement with the results of these prior studies, the present work confirms the CNN model as achieving the relatively highest adaptability, as demonstrated by the produced validation and accuracy measure results.

Using the mechanisms of factor selection is relevant to maximising the efficiency of landslide models by eliminating unwanted or trivial variables before training them [78]. Multicollinearity analysis (MA) and CSAE methods were applied, thereby confirming 21 LCFs, amongst the selected factors, were suitable and appropriate for this study. Effective application of the two methods can be found in various literature, such as [79]. Following Jenk's algorithm of natural breaks classification, LS maps were divided into five groups of susceptibility classes. In LSMs, these classification techniques have been widely used in the literature [80]. This strategy of clustering data helps to reduce the mean variance of each class from the mean within the class range and to increase the discrepancy between each class from the means of the other classes [81]. These studies have suggested the effective use of these classification techniques for this research. In accordance with the natural break classification method, 23.02 and 14.40% of areas were covered under very high to high susceptibility classes by the best fitting CNN model. The area covered by the high and very high susceptible classes of CNN model was greater than the ANN and ADTree models (Figure 9).

An accurate LSM is a vital and reliable tool that helps to map the present, and predict the future, landslides of an area. In this research, LSMs are justified by using the ROC curve and several other accuracy measures. The ROC curve is a widely accepted validation technique in various fields of research [82]. Consideration of various techniques can be more effective and more justified in resolving the issue than a single approach of validation. Therefore, a prioritisation approach using CF was performed amongst the models by considering the ROC and 23 other measures to determine the best result. CF has been widely accepted for prioritisation studies, especially in watershed analysis [69]. Many researchers have emphasised the need to examine and compare susceptibility models for landslides with different approaches and techniques, because a small percentage of

improvement in precision can control the resulting susceptibility areas of landslide [83,84]. Compared with the use of traditional models, the implementation of a CNN model may seem a more challenging task, but its slightly higher predictive efficiency is of considerable significance [84]. The accuracy of a CNN model with regard to AUC is greater than conventional ML models. In our study, CF-based prioritisation provided the relative priority amongst the models in which CNN model ranked first (Table 5), designating the best fit of this DL model, similar to previous studies.

Another significant part of this study is the factor contribution analysis in producing LSM. All factors in a region are not equally responsible for causing landslides. Several researchers have used different methods for the analysis of the importance of conditioning factors [85–89]. These methods include Relief F test, CSAE, information gain ratio and RR. All these methods have been applied successfully to determine the degree of association of the factors with the goal. In this study, RR was adopted, confirming factors such as rainfall, NDVI, LULC, slope aspect, elevation, and TRI, as the most important driving factors for landslides occurring in the study area. The outcome showed that the most vulnerable zones of the landslide are found in the northern portion of the district where soil condition, weak geology, torrent runoff, high altitude, steep sloping, rugged topography and heavy rainfall are the chief reasons for the occurrence of landslides. The East Sikkim district is covered with high hill-slope mountains. Therefore, the topographical ruggedness and steep sloping with certain geological structure, bears instability in landmasses in different parts of the study area. During the monsoon period, heavy precipitations and thunderstorms work as accelerating factors of landslides. Human activities, such as road construction, building construction, tea plantation and other agricultural activities are disturbing the balance of the slope gradient, thereby causing man-made landslides. The role of these factors, especially in the hillslope regions, were found to be significant in the previous literature [89,90]. However, every research work has a certain limitation. The limitation of the present study is the absence of some geological properties, such as joint, foliation and bedding. Only surface geology data were used. However, despite this limitation, the current study has good scope to accurately demarcate the landslide-susceptible area for future planning and management.

6. Conclusions

Every year, landslides cause tremendous damage to humans and their properties. Therefore, a comprehensive strategy is urgently needed to resolve this phenomenon. Identifying the places where landslides have previously occurred, and regions that are prone to landslides, is necessary to avoid future landslides. Different approaches and procedures may be used to geographically forecast landslide-prone regions, with some being more accurate than others. In this study, DL and hybrid tree ML models, such as CNN, ANN, ADTree, CART, FTree and LMT models, were used to generate LSMs. The results showed that DL models outperformed tree ML models in this analysis. These methods involved identifying possible landslide sites and implementing preventative and management measures. The CNN DL model had the best prediction performance amongst the models in this study, followed by ANN, ADTree, CART, FTree and LMT. Rainfall, NDVI and LULC were found to have greatly contributed to making an area prone to landslide based on RR. Therefore, future damage in mountainous areas or areas with comparable physiographic characteristics may be prevented by generating more precise LSMs using DL techniques, such as the CNN model, and selecting appropriate LCFs for generalisation. The major benefit of the three DL and DT models is that they automate the task of searching numerous databases for useful information. However, these methods have some drawbacks, including that pre-processing might take a long time due to the numerous operations involved. The LSMs developed in this study can be used for landslide mitigation strategies in the East Sikkim district region. LSMs can assist decision makers in making development and planning decisions. These maps can be used as a preliminary step for landslide risk management research. Landslide forecasting is relatively accurate because DL and tree-based

modelling are paired with RS and GIS spatial data. We only considered the region of East Sikkim in the Indian Himalaya for our study. The occurrence of landslides in other regions of India should be investigated in future research to give data that are more relevant. If the proper variables for measuring landslide incidence are identified, further study will enable comprehensive management and control of landslides in vulnerable regions.

Author Contributions: Conceptualization, S.S. and J.R.; methodology, S.S.; software, J.R.; validation, J.R., T.K.H. and S.S.; formal analysis, S.S. and J.R.; investigation, J.R. and T.K.H.; data curation, J.R. and S.S.; writing—original draft preparation, J.R., S.S. and T.K.H.; writing—review and editing, S.S., B.P., A.D. and T.K.H.; visualization, J.R.; supervision, S.S.; funding acquisition, B.P., K.N.A.M. and A.M.A.; B.P. has effectively edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw data were generated at University of Gour Banga, Malda, India. Derived data supporting the findings of this study are available from the corresponding author [B.P.] on request.

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

References

1. Reichenbach, P.; Rossi, M.; Malamud, B.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
2. Cao, J.; Zhang, Z.; Wang, C.; Liu, J.; Zhang, L. Susceptibility assessment of landslides triggered by earthquakes in the Western Sichuan Plateau. *CATENA* **2018**, *175*, 63–76. [CrossRef]
3. Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [CrossRef]
4. Dikshit, A.; Pradhan, B.; Alamri, A.M. Pathways and challenges of the application of artificial intelligence to geohazards modelling. *Gondwana Res.* **2020**. [CrossRef]
5. Bordoni, M.; Meisina, C.; Valentino, R.; Lu, N.; Bittelli, M.; Chersich, S. Hydrological factors affecting rainfall-induced shallow landslides: From the field monitoring to a simplified slope stability analysis. *Eng. Geol.* **2015**, *193*, 19–37. [CrossRef]
6. Corominas, J.; Van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.-P.; Fotopoulou, S.; Catani, F.; Eeckhaut, M.V.D.; Mavrouli, O.; Agliardi, F.; et al. Recommendations for the quantitative analysis of landslide risk. *Bull. Int. Assoc. Eng. Geol.* **2013**, *73*, 209–263. [CrossRef]
7. Depicker, A.; Jacobs, L.; Delvaux, D.; Havenith, H.-B.; Mateso, J.-C.M.; Govers, G.; Dewitte, O. The added value of a regional landslide susceptibility assessment: The western branch of the East African Rift. *Geomorphology* **2019**, *353*, 106886. [CrossRef]
8. 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]
9. DU, G.; Zhang, Y.-S.; Iqbal, J.; Yang, Z.-H.; Yao, X. Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China. *J. Mt. Sci.* **2017**, *14*, 249–268. [CrossRef]
10. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. An assessment of multivariate and bivariate approaches in landslide susceptibility mapping: A case study of Duzkoy district. *Nat. Hazards* **2014**, *76*, 471–496. [CrossRef]
11. Wang, L.-J.; Guo, M.; Sawada, K.; Lin, J.; Zhang, J. Landslide susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models. *CATENA* **2015**, *135*, 271–282. [CrossRef]
12. Poli, S.; Sterlacchini, S. Landslide Representation Strategies in Susceptibility Studies using Weights-of-Evidence Modeling Technique. *Nat. Resour. Res.* **2007**, *16*, 121–134. [CrossRef]
13. Armas, I. Weights of evidence method for landslide susceptibility mapping. Prahova Subcarpathians, Romania. *Nat. Hazards* **2011**, *60*, 937–950. [CrossRef]
14. Meena, S.R.; Ghorbanzadeh, O.; Blaschke, T. A Comparative Study of Statistics-Based Landslide Susceptibility Models: A Case Study of the Region Affected by the Gorkha Earthquake in Nepal. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 94. [CrossRef]

15. 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]

16. Merghadi, A.; Yunus, A.P.; Dou, J.; Whiteley, J.; ThaiPham, B.; Bui, D.T.; Avtar, R.; Abderrahmane, B. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Sci. Rev.* **2020**, *207*, 103225. [CrossRef]

17. Schmidt, J.; Marques, M.R.; Botti, S.; Marques, M.A. Recent advances and applications of machine learning in solid-state materials science. *NPJ Comput. Mater.* **2019**, *5*, 83. [CrossRef]

18. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

19. Kardani, N.; Bardhan, A.; Samui, P.; Nazem, M.; Zhou, A.; Armaghani, D.J. A novel technique based on the improved firefly algorithm coupled with extreme learning machine (ELM-IFF) for predicting the thermal conductivity of soil. *Eng. Comput.* **2021**, 1–20. [CrossRef]

20. Kardani, N.; Zhou, A.; Shen, S.-L.; Nazem, M. Estimating unconfined compressive strength of unsaturated cemented soils using alternative evolutionary approaches. *Transp. Geotech.* **2021**, *29*, 100591. [CrossRef]

21. Kardani, N.; Zhou, A.; Nazem, M.; Shen, S.-L. Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. *J. Rock Mech. Geotech. Eng.* **2020**, *13*, 188–201. [CrossRef]

22. Asteris, P.G.; Mamou, A.; Hajihassani, M.; Hasanipanah, M.; Koopialipoor, M.; Le, T.-T.; Kardani, N.; Armaghani, D.J. Soft computing based closed form equations correlating L and N-type Schmidt hammer rebound numbers of rocks. *Transp. Geotech.* **2021**, *29*, 100588. [CrossRef]

23. Wang, Y.; Fang, Z.; Wang, M.; Peng, L.; Hong, H. Comparative study of landslide susceptibility mapping with different recurrent neural networks. *Comput. Geosci.* **2020**, *138*, 104445. [CrossRef]

24. Caniani, D.; Pascale, S.; Sdao, F.; Sole, A. Neural networks and landslide susceptibility: A case study of the urban area of Potenza. *Nat. Hazards* **2007**, *45*, 55–72. [CrossRef]

25. Andrieu, C.; De Freitas, N.; Doucet, A.; Jordan, M.I. An Introduction to MCMC for Machine Learning. *Mach. Learn.* **2003**, *50*, 5–43. [CrossRef]

26. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Prabhat Deep learning and process understanding for data-driven Earth system science. *Nat. Cell Biol.* **2019**, *566*, 195–204. [CrossRef]

27. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]

28. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

29. Deng, L.; Yu, D. Deep learning: Methods and applications found. *Trends Signal Process.* **2014**, *7*, 197–387. [CrossRef]

30. Bui, D.T.; Tsangaratos, P.; Nguyen, V.-T.; Van Liem, N.; Trinh, P.T. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *CATENA* **2020**, *188*, 104426. [CrossRef]

31. Fang, Z.; Wang, Y.; Peng, L.; Hong, H. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Comput. Geosci.* **2020**, *139*, 104470. [CrossRef]

32. Taherdangkoo, R.; Liu, Q.; Xing, Y.; Yang, H.; Cao, V.; Sauter, M.; Butscher, C. Predicting methane solubility in water and seawater by machine learning algorithms: Application to methane transport modeling. *J. Contam. Hydrol.* **2021**, *242*, 103844. [CrossRef] [PubMed]

33. Gokgoz, E.; Subasi, A. Comparison of decision tree algorithms for EMG signal classification using DWT. *Biomed. Signal Process. Control.* **2015**, *18*, 138–144. [CrossRef]

34. Park, S.; Hamm, S.Y.; Kim, J. Performance evaluation of the GIS-based data-mining techniques decision tree, random forest, and rotation forest for landslide susceptibility modeling. *Sustainability* **2020**, *11*, 5659. [CrossRef]

35. Saha, S.; Roy, J.; Pradhan, B.; Hembram, T.K. Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at East Sikkim Himalayan, India. *Adv. Space Res.* **2021**, *68*, 2819–2840. [CrossRef]

36. Yilmaz, C.; Topal, T.; Süzen, M.L. GIS-based landslide susceptibility mapping using bivariate statistical analysis in Devrek (Zonguldak-Turkey). *Environ. Earth Sci.* **2011**, *65*, 2161–2178. [CrossRef]

37. Saha, A.; Saha, S. Application of statistical probabilistic methods in landslide susceptibility assessment in Kurseong and its surrounding area of Darjeeling Himalayan, India: RS-GIS approach. *Environ. Dev. Sustain.* **2020**, *23*, 4453–4483. [CrossRef]

38. Van Dao, D.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayedi, H.; Van Phong, T.; Ly, H.B.; Le, T.T.; Trinh, P.T.; et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *CATENA* **2020**, *1*, 104451.

39. Nhu, V.-H.; Hoang, N.-D.; Nguyen, H.; Ngo, P.T.T.; Bui, T.T.; Hoa, P.V.; Samui, P.; Bui, D.T. Effectiveness assessment of Keras based deep learning with different robust optimization algorithms for shallow landslide susceptibility mapping at tropical area. *CATENA* **2020**, *188*, 104458. [CrossRef]

40. Tian, H.; Nan, H.; Yang, Z. Select landslide susceptibility main affecting factors by multi-objective optimization algorithm. In Proceedings of the 2010 6th International Conference on Natural Computation, Yantai, China, 10–12 August 2010; Volume 4, pp. 1830–1833. [CrossRef]

41. Alkhasawneh, M.; Ngah, U.K.; Tay, L.T.; Isa, N.A.M.; Al-Batah, M.S. Determination of Important Topographic Factors for Landslide Mapping Analysis Using MLP Network. *Sci. World J.* **2013**, *2013*, 415023. [CrossRef]

42. Wood, J. Geomorphometry in landserf. *Dev. Soil Sci.* **2009**, *33*, 333–349. [CrossRef]

43. Wilson, J.P.; Gallant, J.C. *Terrain Analysis: Principles and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2000.

44. Berry, J.K. Use surface area for realistic calculations. *Geo. World* **2002**, *15*, 20–21.

45. Riley, S.J.; De Gloria, S.D.; Elliot, R. Index that quantifies topographic heterogeneity. *Intermt. J. Sci.* **1999**, *5*, 23–27.

46. Chen, C.-W.; Saito, H.; Oguchi, T. Rainfall intensity–duration conditions for mass movements in Taiwan. *Prog. Earth Planet. Sci.* **2015**, *2*, 14. [CrossRef]

47. 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]

48. Cama, M.; Lombardo, L.; Conoscenti, C.; Rotigliano, E. Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). *Geomorphology* **2017**, *288*, 52–65. [CrossRef]

49. Vafaie, H.; Imam, I.F. Feature selection methods: Genetic algorithms vs. greedy-like search. *Proc. Int. Conf. Fuzzy Intell. Control Syst.* **1994**, *51*, 28.

50. Hoerl, A.E.; Kennard, R.W. Ridge regression: Applications to nonorthogonal problems. *Technometrics* **1970**, *12*, 69–82. [CrossRef]

51. Tikhonov, A.N.; Goncharsky, A.V.; Stepanov, V.V.; Yagola, A.G. Regularization methods. *Numer. Methods Solut. Ill-Posed Probl.* **1995**, *328*, 7–63. [CrossRef]

52. Valueva, M.; Nagornov, N.; Lyakhov, P.; Valuev, G.; Chervyakov, N. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Math. Comput. Simul.* **2020**, *177*, 232–243. [CrossRef]

53. Gupta, T.K.; Raza, K. Optimization of ANN architecture: A review on nature-inspired techniques. Machine learning in bio-signal analysis and diagnostic imaging. In *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 1, pp. 159–182.

54. Ding, A.; Zhang, Q.; Zhou, X.; Dai, B. Automatic recognition of landslide based on CNN and texture change detection. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 11–13 November 2016; pp. 444–448.

55. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the ICML, Haifa, Israel, 21–24 June 2010; Available online: https://openreview.net/forum?id=rkb15iZdZB (accessed on 20 December 2020).

56. Haykin, S. *Neural Networks and Learning Machines.[sl] Pearson Upper Saddle River*; Publisher-Pearson: Upper Saddle River, NJ, USA, 2009; Volume 3.

57. Gardner, M.; Dorling, S. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [CrossRef]

58. Moayedi, H.; Mehrabi, M.; Mosallanezhad, M.; Rashid, A.S.A.; Pradhan, B. Modification of landslide susceptibility mapping using optimized PSO-ANN technique. *Eng. Comput.* **2018**, *35*, 967–984. [CrossRef]

59. Freund, Y.; Mason, L. The alternating decision tree learning algorithm. *ICML* **1999**, *99*, 124–133.

60. Sok, H.K.; Ooi, M.; Kuang, Y.C.; Demidenko, S. Multivariate alternating decision trees. *Pattern Recognit.* **2016**, *50*, 195–209. [CrossRef]

61. Nefeslioglu, H.A.; Sezer, E.A.; Gokceoglu, C.; Bozkir, A.S.; Duman, T.Y. Assessment of Landslide Susceptibility by Decision Trees in the Metropolitan Area of Istanbul, Turkey. *Math. Probl. Eng.* **2010**, *2010*, 901095. [CrossRef]

62. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Wadsworth International Group: Belmont, CA, USA, 1984.

63. Gama, J. Functional trees. *Mach. Learn.* **2004**, *55*, 219–250. [CrossRef]

64. Landwehr, N.; Hall, M.; Frank, E. Logistic model trees. *Mach. Learn.* **2005**, *59*, 161–205. [CrossRef]

65. Wang, L.-J.; Guo, M.; Sawada, K.; Lin, J.; Zhang, J. A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. *Geosci. J.* **2015**, *20*, 117–136. [CrossRef]

66. Akgün, A.; Türk, N. Mapping erosion susceptibility by a multivariate statistical method: A case study from the Ayvalık region, NW Turkey. *Comput. Geosci.* **2010**, *37*, 1515–1524. [CrossRef]

67. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*; John Wiley & Sons. Inc.: New York, NY, USA, 2001.

68. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2005**, *27*, 861–874. [CrossRef]

69. Altaf, S.; Meraj, G.; Romshoo, S.A. Morphometry and land cover based multi-criteria analysis for assessing the soil erosion susceptibility of the western Himalayan watershed. *Environ. Monit. Assess.* **2014**, *186*, 8391–8412. [CrossRef]

70. Rahmati, O.; Kornejady, A.; Samadi, M.; Deo, R.C.; Conoscenti, C.; Lombardo, L.; Dayal, K.; Taghizadeh-Mehrjardi, R.; Pourghasemi, H.R.; Kumar, S.; et al. PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. *Sci. Total Environ.* **2019**, *664*, 296–311. [CrossRef]

71. Fukuda, S.; De Baets, B.; Waegeman, W.; Verwaeren, J.; Mouton, A.M. Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. *Environ. Model. Softw.* **2013**, *47*, 1–6. [CrossRef]

72. Arabameri, A.; Saha, S.; Roy, J.; Chen, W.; Blaschke, T.; Bui, D.T. Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in The Gallicash River Watershed, Iran. *Remote. Sens.* **2020**, *12*, 475. [CrossRef]

73. Roy, J.; Saha, S.; Arabameri, A.; Blaschke, T.; Bui, D.T. A Novel Ensemble Approach for Landslide Susceptibility Mapping (LSM) in Darjeeling and Kalimpong Districts, West Bengal, India. *Remote. Sens.* **2019**, *11*, 2866. [CrossRef]

74. Wang, Y.; Fang, Z.; Hong, H. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. *Sci. Total Environ.* **2019**, *666*, 975–993. [CrossRef] [PubMed]

75. Moosavi, V.; Talebi, A.; Shirmohammadi, B. Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. *Geomorphology* **2014**, *204*, 646–656. [CrossRef]

76. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote. Sens.* **2019**, *11*, 196. [CrossRef]

77. Zhang, G.; Wang, M.; Liu, K. Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. *Int. J. Disaster Risk Sci.* **2019**, *10*, 386–403. [CrossRef]

78. Pham, B.T.; Prakash, I.; Singh, S.; Shirzadi, A.; Shahabi, H.; Tran, T.-T.; Bui, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *CATENA* **2018**, *175*, 203–218. [CrossRef]

79. Sahin, E.K.; Ipbuker, C.; Kavzoglu, T. Investigation of automatic feature weighting methods (Fisher, Chi-square and Relief-F) for landslide susceptibility mapping. *Geocarto Int.* **2016**, *32*, 956–977. [CrossRef]

80. Roy, J.; Saha, S. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenviron. Disasters* **2019**, *6*, 1–18. [CrossRef]

81. Jenks, G.F. The data model concept in statistical mapping. *Int. Yearb. Cartogr.* **1967**, *7*, 186–190.

82. Mahato, S.; Pal, S. Groundwater Potential Mapping in a Rural River Basin by Union (OR) and Intersection (AND) of Four Multi-criteria Decision-Making Models. *Nat. Resour. Res.* **2018**, *28*, 523–545. [CrossRef]

83. Hong, H.; Liu, J.; Bui, D.T.; Pradhan, B.; Acharya, T.D.; Pham, B.T.; Zhu, A.-X.; Chen, W.; Bin Ahmad, B. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *CATENA* **2018**, *163*, 399–413. [CrossRef]

84. Yi, Y.; Zhang, Z.; Zhang, W.; Jia, H.; Zhang, J. Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. *CATENA* **2020**, *195*, 104851. [CrossRef]

85. Pourghasemi, H.R.; Pouyan, S.; Farajzadeh, Z.; Sadhasivam, N.; Heidari, B.; Babaei, S.; Tiefenbacher, J.P. Assessment of the outbreak risk, mapping and infection behavior of COVID-19: Application of the autoregressive integrated-moving average (ARIMA) and polynomial models. *PLoS ONE* **2020**, *15*, e0236238. [CrossRef] [PubMed]

86. Roy, J.; Saha, S. Integration of artificial intelligence with meta classifiers for the gully erosion susceptibility assessment in Hinglo river basin, Eastern India. *Adv. Space Res.* **2020**, *67*, 316–333. [CrossRef]

87. Hosseinalizadeh, M.; Kariminejad, N.; Chen, W.; Pourghasemi, H.R.; Alinejad, M.; Behbahani, A.M.; Tiefenbacher, J.P. Spatial modelling of gully headcuts using UAV data and four best-first decision classifier ensembles (BFTree, Bag-BFTree, RS-BFTree, and RF-BFTree). *Geomorphology* **2019**, *329*, 184–193. [CrossRef]

88. 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]

89. Saha, S.; Saha, A.; Hembram, T.K.; Pradhan, B.; Alamri, A.M. Evaluating the performance of individual and novel ensemble of machine learning and statistical models for landslide susceptibility assessment at Rudraprayag District of Garhwal Himalaya. *Appl. Sci.* **2020**, *10*, 3772. [CrossRef]

90. Ayalew, L.; Yamagishi, H.; Ugawa, N. Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. *Landslides* **2004**, *1*, 73–81. [CrossRef]