Article

A Comparative Assessment of Sampling Ratios Using Artificial Neural Network (ANN) for Landslide Predictive Model in Langat River Basin, Selangor, Malaysia

Siti Norsakinah Selamat [1], Nuriah Abd Majid [1,*] and Aizat Mohd Taib [2]

[1] Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia
[2] Department of Civil Engineering, Faculty of Engineering and Build Environment, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia
* Correspondence: nuriah@ukm.edu.my; Tel.: +60-3-8921-7639

Abstract: Landslides have been classified as the most dangerous threat around the world, causing huge damage to properties and loss of life. Increased human activity in landslide-prone areas has been a major contributor to the risk of landslide occurrences. Therefore, machine learning has been used in landslide studies to develop a landslide predictive model. The main objective of this study is to evaluate the most suitable sampling ratio for the predictive landslide model in the Langat River Basin (LRB) using Artificial Neural Networks (ANNs). The landslide inventory was divided randomly into training and testing datasets using four sampling ratios (50:50, 60:40, 70:30, and 80:20). A total of 12 landslide conditioning factors were considered in this study, including the elevation, slope, aspect, curvature, topography wetness index (TWI), distance to the road, distance to the river, distance to faults, soil, lithology, land use, and rainfall. The evaluation model was performed using certain statistical measures and area under the curve (AUC). Finally, the most suitable predictive model was chosen based on the model validation results using the compound factor (CF) method. Based on the results, the predictive model with an 80:20 ratio indicates a realistic finding and was classified as the first rank among others. The AUC value for the training dataset is 0.931, while the AUC value for the testing dataset is 0.964. These attempts will help a great deal when it comes to choosing the best ratio of training samples to testing samples to create a reliable and complete landslide prediction model for the LRB.

Keywords: landslide; predictive model; sampling ratio; landslide susceptibility; Langat River Basin

1. Introduction

Landslides are dangerous geological phenomena that can cause significant loss of life and property and disrupt social development [1,2]. Generally, the increase in landslide occurrences is due to the growing instability caused by forest destruction, uncontrolled development, and rapid urbanization caused by the growing population [3,4]. Landslides are phenomena that are mainly triggered by climate or geophysical reasons in significant geological motions. Landslides are frequently caused by deforestation activities in the highlands for development purposes such as roads and residence areas [5]. The consequences of climate change and human activities such as deforestation and urbanization are considered primary factors that trigger landslide tragedies around the world [6]. The landslide resulted in environmental and socio-economic damage, such as loss of life, damaged properties, and disruption of communications. However, these phenomena become dangerous when they interact with human activity [7]. These issues will be confronted by regions around the world, especially in developing countries where management, adaptation, and mitigation are difficult to sustain.

The LRB has developed rapidly in the past decade in urbanization, agriculture, and industry. Rapid urbanization in the LRB has been influenced by improvements in infrastructure projects and transport networks. This area has developed rapidly with the transportation system, which increases the distribution of development activities and increases the population rate. The total population in the LRB reached over 1.7 million people in 2015 [8], with an increase of over 2.6 million in the year 2020 [9]. Thus, this area frequently experiences landslides due to the environmental factor of a high and hilly area undergoing rapid development. Moreover, extreme rainfall conditions influence landslides in the LRB. Landslides in this area have been on the increase in recent years. Over several days, heavy rain caused hundreds of landslides and debris flows on small to large scales in various locations throughout the LRB.

The ability to use spatial and temporal data to accurately evaluate landslide susceptibility is fundamental to the management of landslide-prone areas around the world [10]. Landslide occurrences are caused by a variety of factors [11–13]. There are several categories of factors that affect the natural stability of a slope and determine landslide susceptibility factor studies, such as topographical, hydrological, geological, and human activities [13–16]. In addition, identifying the spatial patterns of landslide occurrences under natural geo-environmental conditioning factors throughout a large-scale area with only field surveys is an exceptionally challenging task [10]. Thus, modelling the landslide susceptibility is a desirable substitute for field methods since it can provide analytical frameworks for assessing and understanding the underlying patterns of this phenomenon under different local conditions [17].

The landslide susceptibility map (LSM) is able to predict the probability of landslides that may occur in the future. The prediction is based on historical landslide occurrences and places that have similar environmental characteristics. Identifying landslide-prone areas was an important part of disaster management for development planning. Nowadays, landslide susceptibility can be modelled using a machine learning approach, including an artificial neural network [18,19], a support vector machine [20,21], random forest [22,23], decision trees [24,25], and Naïve Bayes [26,27]. The model is derived from machine learning using training and testing datasets to make a landslide prediction.

The ratio of training and testing datasets is a basic element in the model development process. The accuracy of landslide susceptibility maps can be greatly impacted by both the sampling technique and the size of the training data [28]. The availability of historical landslide records can have an impact on the sample ratio, but the predictive ability of the models is highly dependent on it. Several researchers have used 50:50 [29], 60:40 [22], 70:30 [27], and 80:20 [30] sampling ratios for training and testing data sets during landslide model development. However, the sampling ratio for the training and testing datasets is not clearly defined. The assessment of sampling ratios for training and testing dataset studies is still limited. This study contributes to filling this gap. Therefore, the main objective of this study is to identify the most suitable sampling ratio for the landslide predictive model in the LRB using an Artificial Neural Network.

2. Materials and Methods

2.1. Study Area

The LRB was chosen as the study area, as shown in Figure 1. It includes several districts, such as Hulu Langat, the Federal Territory of Putrajaya, Sepang, Kuala Langat, and part of the Seremban area. The total area of LRB is approximately 2750 km^2. There are mountains in the northern part of the LRB, whereas the southern part is relatively flat. The LRB is the most highly urbanized river basin in Malaysia [31], serves as a catchment area, and supplies water to two-thirds of the state of Selangor. The LRB is one of the largest basins in Selangor. Malaysia is a tropical country, which results in hot, humid, and rainy weather for almost the whole year. The monsoon season in the country starts from May to September and November to March. During this season, the intensity of rainfall increases significantly. The LRB receives an average annual rainfall of between 144. 586 mm and 296.

254 mm. Meanwhile, the mean annual temperature ranges between 33 °C and 24.2 °C [32]. The highest elevation in this area is 1448.25 m, which represents the top of the hilly area.
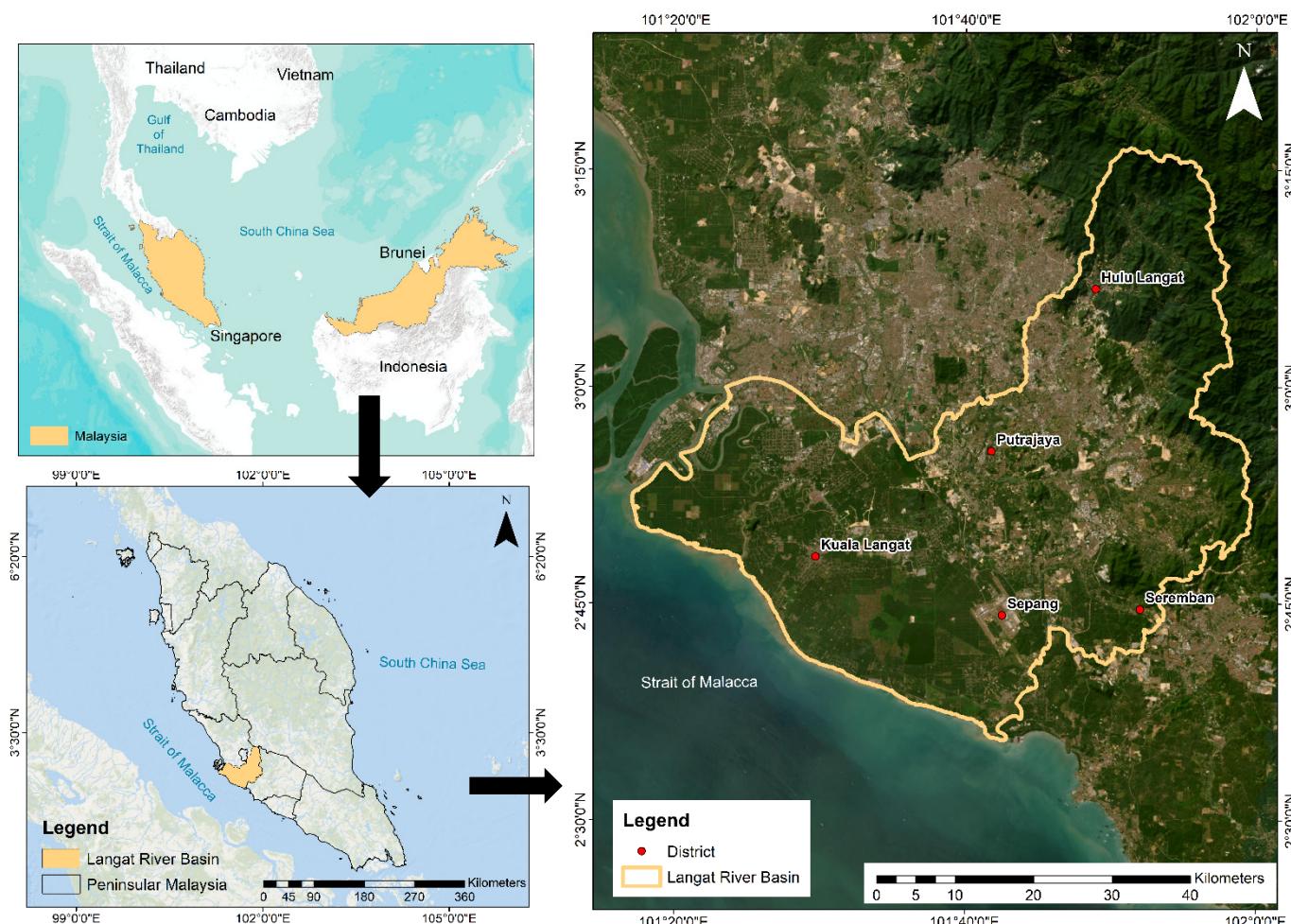

Figure 1. Langat River Basin, Selangor, Malaysia.

2.2. Data Collection and Description

Landslides are caused by the complex interaction between multiple geo-environmental factors. The selection of landslide conditioning factors as independent variables that trigger landslide occurrences is an important task for developing the landslide model. A total of 12 landslide conditioning factors were selected, including the elevation, slope, curvature, aspect, Topographic Wetness Index (TWI), distance to the river, distance to the road, distance to the fault, soil, lithology, land use, and rainfall, as shown in Table 1.

2.3. Methodology

This study was conducted using the following methods, as shown in Figure 2.

2.3.1. Landslide Inventory Data

The preparation of landslide inventory data is an important step in the landslide modelling process. Landslide inventory data are an important dataset for constructing an accurate and efficient landslide prediction model [33]. The landslide inventory data represents historical landslide information, such as landslide location, date, and year. In this study, landslide inventory data were identified by the interpretation of satellite imagery and field observations. A total of 70 landslide locations were recorded between the years 2000 to 2020.

Table 1. Spatial database for landslide conditioning factors.

Conditioning Factor	Type of Data	Scale/Resolution	Sources
Elevation Slope Aspect Curvature TWI	IFSAR	5-m pixel size	Department of Survey and Mapping Malaysia (JUPEM)
Distance to river	River map	1:10,000	Department of Irrigation and Drainage (JPS)
Distance to road	Road map	1:10,000	Open street map
Soil	Soil series map	1:100,000	Department of Agriculture (DOA)
Lithology	Geology map	1:100,000	Department of Mineral and Geoscience (JMG)
Distance to faults	Faults map	1:10,000	
Land use	Land use map	1:100,000	PLAN MALAYSIA
Rainfall	Rainfall station	1:10,000	Department of Irrigation and Drainage (JPS)

Since this study used a binary classification model to establish a predictive model between landslide and non-landslide events, it is important to create random points to represent non-landslide events [34,35]. A total of 70 non-landslide points were randomly generated within the LRB boundary. Therefore, this study used a 140-training dataset to represent landslide and non-landslide locations at LRB as shown in Figure 3. All the landslide data, including landslide and non-landslide points, were divided into training and testing for landslide predictive modelling. The training dataset was used to develop the model, while the testing dataset was used to evaluate the model quality [12]. In this study, various ratios of testing and training datasets, including 50:50, 60:40, 70:30, and 80:20, were considered to investigate the influence of model performance. The landslide predictive models were produced using WEKA software. The WEKA software package is available for free download and offers a set of machine learning algorithms useful for data mining projects. It was a comprehensive software that could run multiple machine learning algorithms on large datasets and compare the results afterward.

2.3.2. Landslide Conditioning Factors

The elevation values in this study were obtained from the digital elevation model (DEM), which was derived from Interferometric Synthetic Aperture Radar (IFSAR) with a pixel size of 5 m × 5 m (Figure 4a). The elevation is extensively utilized to examine landslide studies and is considered an essential factor that can influence the occurrence of landslides. Slope (Figure 4b), curvature (Figure 4c), and aspect (Figure 4d) were extracted from DEM with a 5-m spatial resolution. Slope angle is very frequently used in landslide studies, and researchers have identified the slope angle as one of the most important parameters for landslide analysis [6]. The slope angle determines the ability of pressure to cause movement; the steeper the area, equivalent to a higher slope, the greater the gravitational pressure component that causes the object to slide [36]. Aspect is another significant component determining slope instability because it regulates topographic moisture due to solar radiation and rainfall impaction [37]. The direction of the slope is represented by aspect, and it has an impact on hydrological processes, weathering, and soil development. Hence, many researchers have used this component extensively in their investigations of landslides [38,39].

Curvature is a conditioning factor that depicts the shape of the terrain surface and represents variations in slope angles over a very small arc of the curve, making it vulnerable

to slope instability [40]. It is one of the most significant contributing factors affecting mass flow down slopes, erosion, and weathering. Curvature is a three-dimensional component of a two-dimensional surface that describes the aberration from a horizontal surface; it is also known as the convex and concave slope [36]. TWI is a significant conditioning factor in landslide occurrence and is frequently used in hydrological process investigations. TWI is a secondary geomorphometric parameter used to describe and quantify local relief [41]. This conditioning factor is widely used to forecast catchment-scale soil moisture and allows topographic control on hydrologic response analyses of the watershed to be investigated [42]. TWI can be represented by Equation (1) as follows: [43]

$$TWI = \ln(\alpha / \tan \beta) \tag{1}$$

where α is cumulative upslope area drainage through a point (per unit contour length) and β is the angle of the slope at the point. The TWI map is shown in Figure 4e.
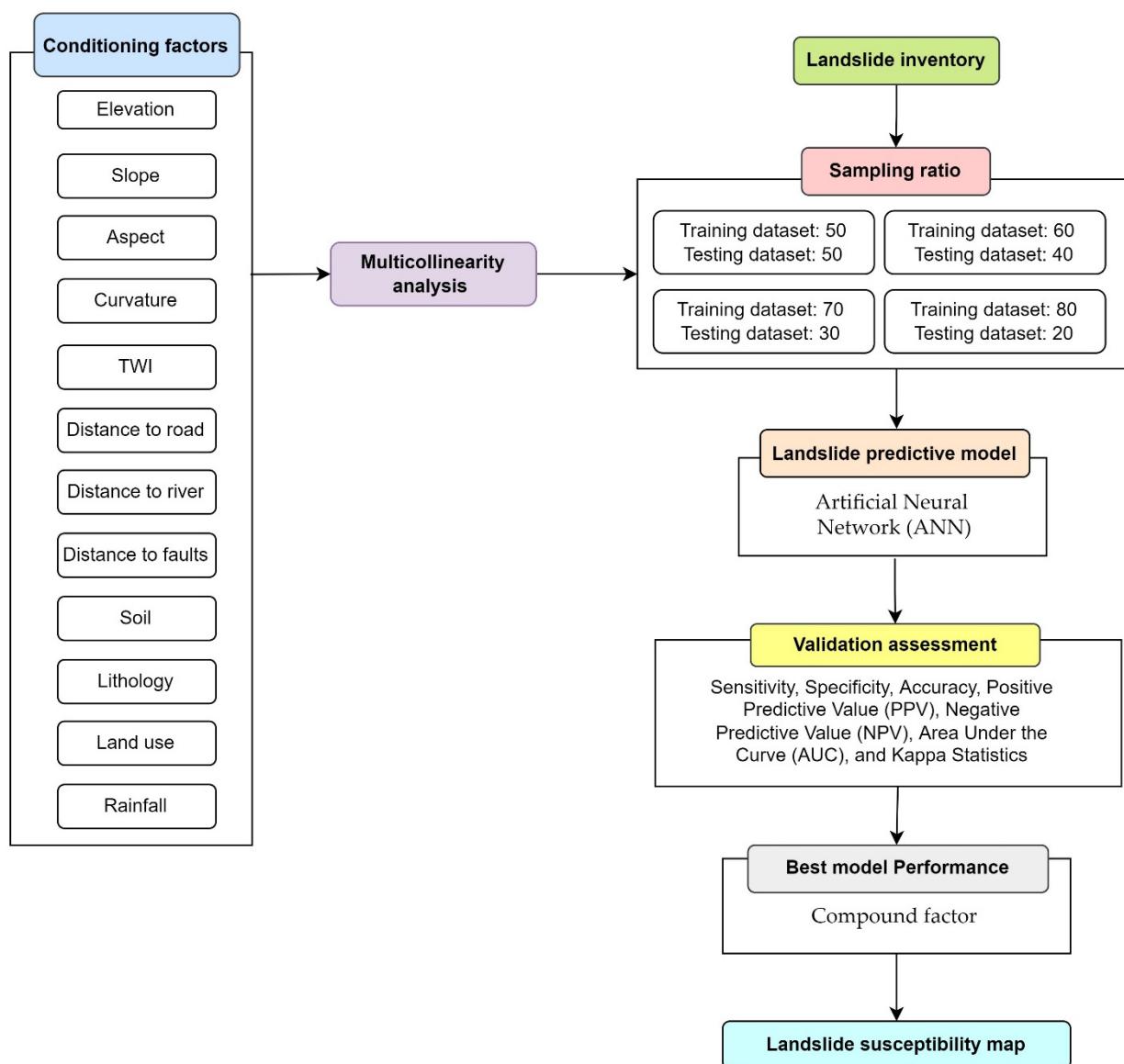

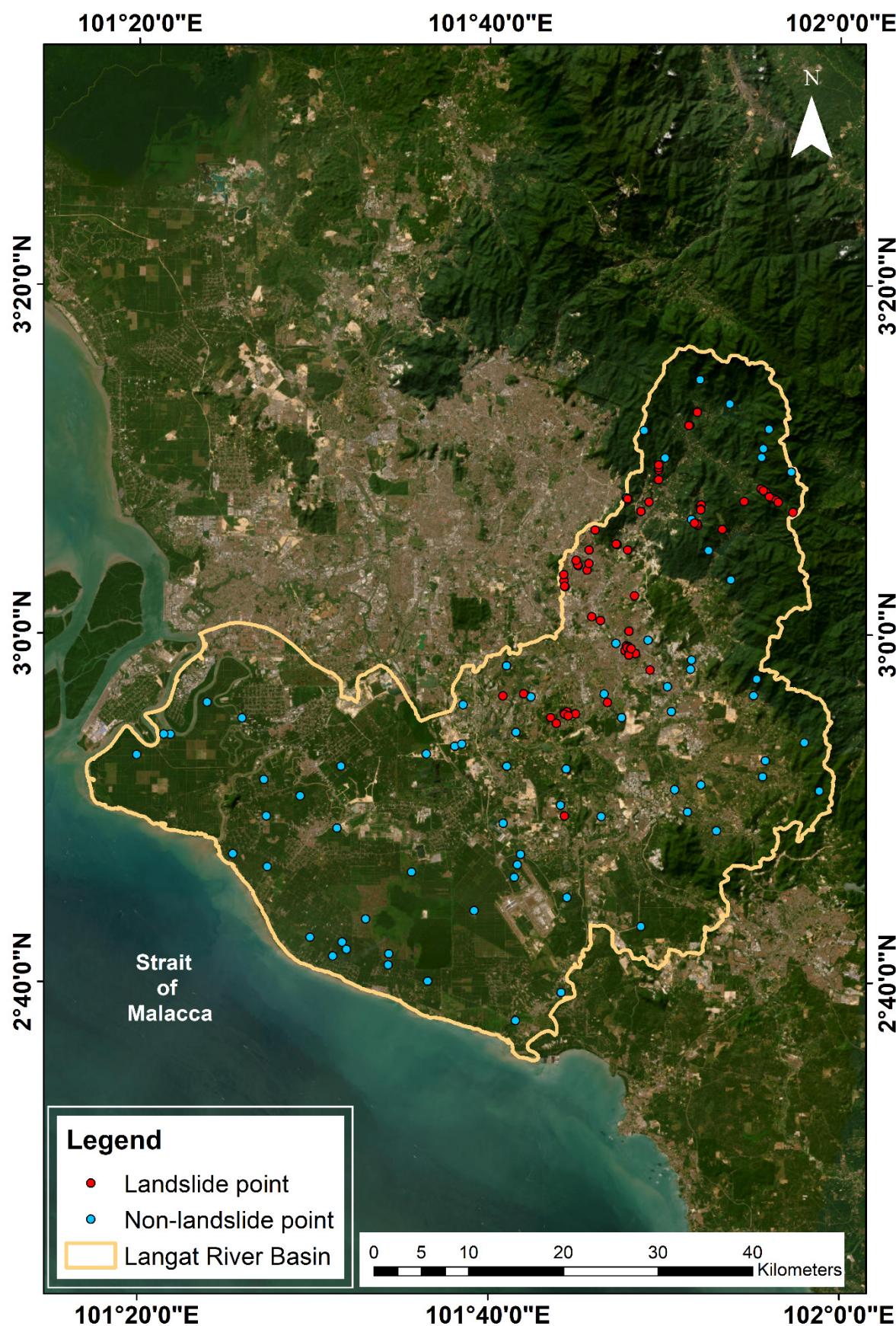**Figure 2.** Flowchart of research method of this study.

Figure 3. Map of landslide inventory at Langat River Basin.
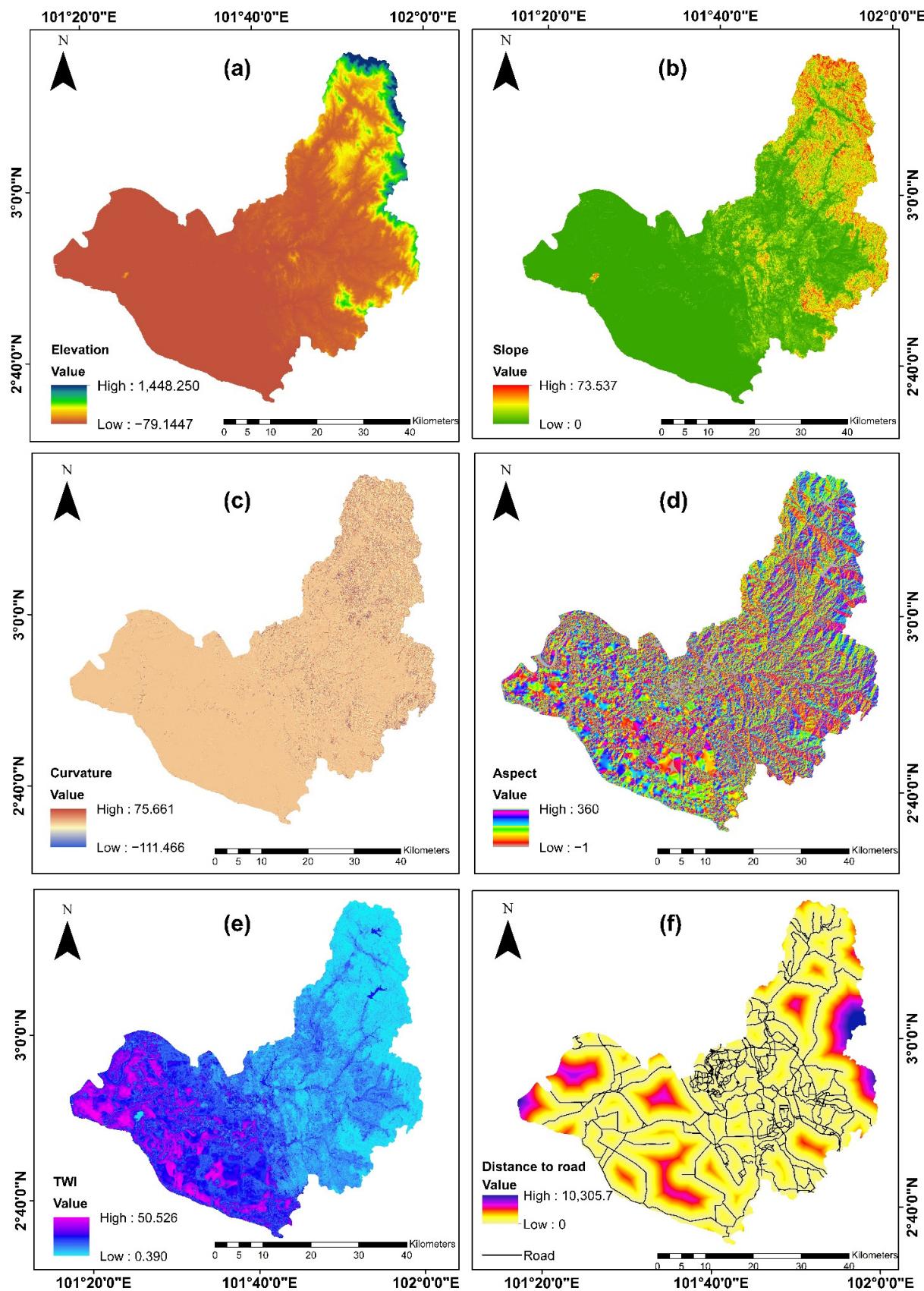
Figure 4. *Cont.*
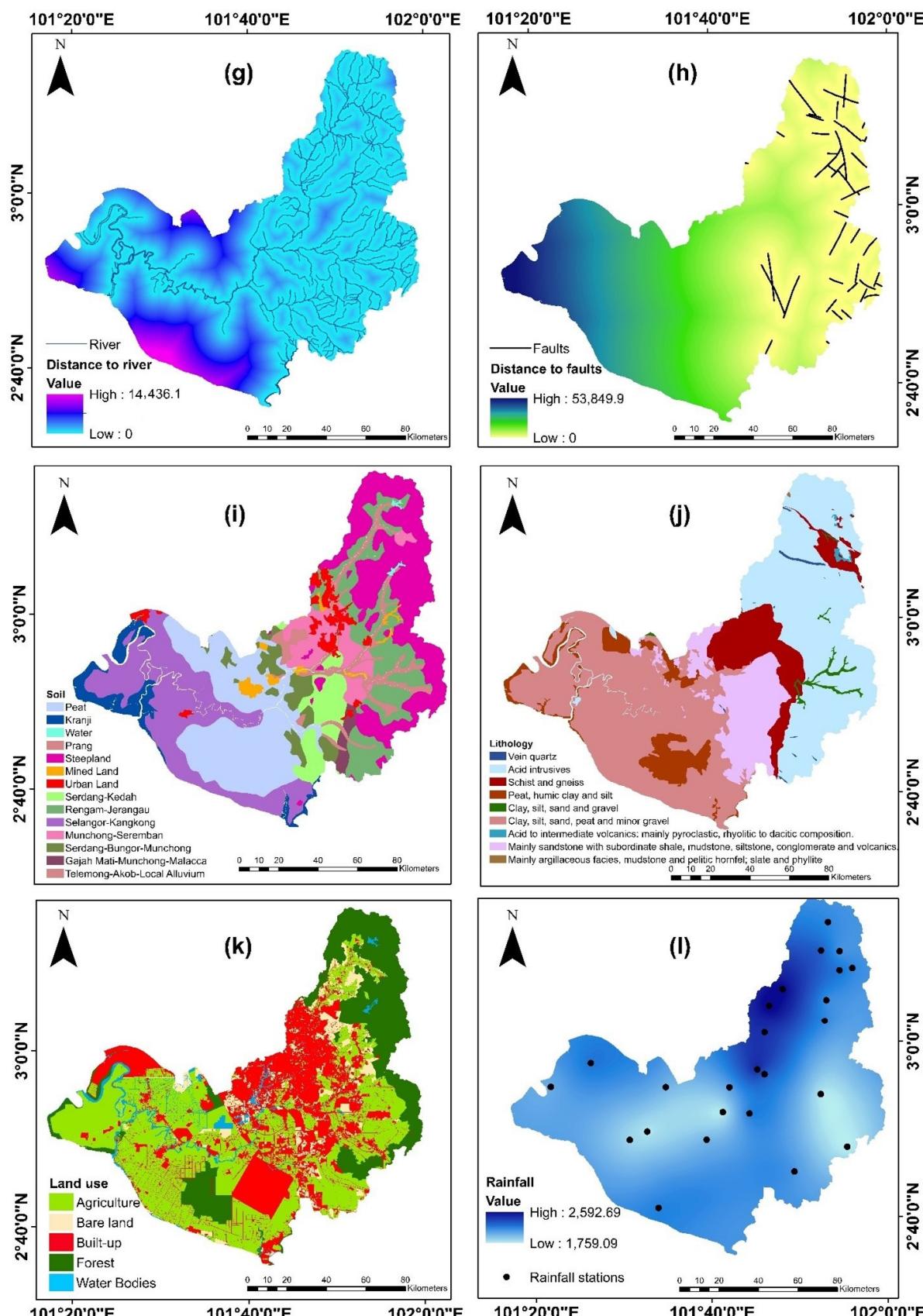
Figure 4. Landslide conditioning factors: (**a**) Elevation, (**b**) slope, (**c**) curvature, (**d**) aspect, (**e**) TWI, (**f**) distance to road, (**g**) distance to river, (**h**) distance to fault (**i**) soil, (**j**) lithology, (**k**) land use, and (**l**) rainfall.

In general, the distances to the road (Figure 4f), river (Figure 4g), and fault (Figure 4h) tend to be associated with a higher risk of landslide occurrences. The nearest distance to the river increases the risk of landslides since the river flow tends to deteriorate the slope material and move it from its original location [44]. Road construction as an alternative human activity network in hilly areas was classified as a human activity causing landslides. A decrease in rock strength due to tectonic movement and the formation of water-permeable fissure zones increases the landslide risk along faults. In this study, the distance to the river, the distance to the road, and the distance to the fault maps were constructed by Euclidean analysis.

Soil formation is strongly influenced by the geological structure of the surrounding bedrock [38]. The physical features of soils play an essential role in preventing mudslides and soil erosion [45]. Thus, the soil series has been identified as an important factor in the study of landslide occurrence due to slope instability. Lithology is also viewed as an important landslide conditioning factor [46]. Internal structures and mineral compositions of rocks and soils are different based on types of lithology [40]. The consequences, rock and soil strata strength, and permeability varied based on the studied area. This conditioning factor plays a very important role in providing valuable information about the physical characteristics of rock and soil. The LRB soil map and lithology map are shown in Figure 4i,j, respectively.

Nowadays, human activity has a direct impact on land use. Land-use changes can have an impact on slope stability, especially with regard to development activities in hilly areas such as road construction and housing development. In this study, land use activities were prepared using a land use map from Plan Malaysia as shown in Figure 4k. Rainfall was categorized as a major triggering factor. Previous studies have been conducted to identify the relationship between landslides and rainfall factors [2,47]. A total of 20 rainfall stations with 30 years of annual average rainfall data in the LRB were used in this study. The annual average rainfall was analyzed using Kriging interpolation as shown in Figure 4l.

2.3.3. Multicollinearity Analysis

Multicollinearity is a condition in which there is a significant correlation between the conditioning factor and other factors. This will have a negative impact on the reduction accuracy and quality of model predictions. In order to overcome this issue, multicollinearity analysis needs to be conducted to analyze the correlation between landslide conditioning factors. In addition, the multicollinearity test is an important step in determining if there was a strong correlation between the conditioning factors using multiple regression [48]. Among the various multicollinearity assessment techniques, Pearson correlation, Variance Inflation Factor (VIF), and Tolerance (TOL) are frequently used in landslide studies [30,33]. Consequently, a multicollinearity analysis was carried out using those techniques for this study.

Pearson correlation analysis was used to determine the correlation between each landslide conditioning factor. The Pearson correlation coefficient (r) is calculated by dividing the covariance into two factors according to the product of their standard deviations [33]:

$$\mathrm{rxy} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{2}$$

where X and Y are the landslide conditioning factors, while $\overline{X}$ and $\overline{Y}$ represent the mean landslide conditioning factors (X and Y). The r value represents the correlation coefficient between landslide conditioning factors. An r value higher than 0.7 indicates a high correlation between the X and Y factors, while an *r* value lower than 0.3 indicates that there is a low correlation between each factor.

VIF is a measure of multicollinearity in a multiple regression landslide conditioning factor. It is calculated by taking the ratio of variance, as follows:

$$\text{VIF} = \frac{1}{\text{Tolerance}} = \frac{1}{1 - R^2} \tag{3}$$

where R^2 is the value of coefficient determination. Thus, the threshold of the multicollinearity issue will occur when the value of VIF is >10 and the value of TOL is <0.10 [33].

2.3.4. Artificial Neural Network (ANN)

The ANN is a type of computer program that replicates the structure of the human brain's neural networks. The ANN is a well-known machine learning technology that has been effectively used to address a wide range of practical challenges, including the issue of landslides. The ANN model can be used to predict future landslides based on the historical distribution of landslide occurrence, thus making it a valuable model for assessing the probability and risk of landslides. Therefore, this model has been widely used in landslide predictive studies [19,49].

The ANN is a computer process that can learn, present, and make a prediction based on the input dataset [18]. The ANN learning process continuously adjusted the network parameter, where all the layers were connected to each other and assigned weighted values from layer to layer [50]. It is a very intricate network of neurons that analyzes information in accordance with the connection weight and transfers the results to the next layer [51]. The multi-layer perceptron (MLP) is the most frequently used architecture for network structure [52,53]. The ANN structure will be divided into three phases, including the input layer, hidden layer, and output layer. All the layers have their own function. In this study, an input layer represents the landslide conditioning factors that were selected for the model development process. The input layer will connect to the hidden layer by their own neurons. Next, the hidden layer was designed to transmit the information using the activation function, control the input neurons, and make predictions about the output neurons. Meanwhile, the output layer represents the landslide prediction, which in this study was used to classify the landslide and non-landslide areas. The model structure is displayed in Figure 5.

2.3.5. Validation Assessment

There are various types of statistical measures to evaluate the performance of the landslide predictive model. In this study, the sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), area under the curve (AUC), and kappa statistics were used to assess the overall performance of landslide prediction. The sensitivity assessment represents the proportion of landslide locations that are accurately classified as landslide occurrences. The specificity assessment represents the proportion of non-landslide locations that are accurately classified as non-landslide occurrences. The accuracy assessment shows the landslide and non-landslide locations that are correctly identified. Moreover, the PPV assessment is measured to determine the probability of actual landslide occurrences at a predicted landslide location. Meanwhile, the NPV assessment is measured to determine the probability that a predicted non-landslide location will have actual non-landslide occurrences. The statistical measures were calculated using the following equations [30]:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\text{Positive Predictive Value (PPV)} \ = \ \frac{\text{TP}}{\text{FP} \ + \ \text{TP}} \tag{7}$$

$$\text{Negative Predictive Value (NPP)} \ = \ \frac{\text{TN}}{\text{FN} \ + \ \text{TN}} \tag{8}$$

where the number of landslide points that were successfully identified as landslides is the true positive (TP) and the number of landslide points that were correctly identified as non-landslide points is the true negative (TN). Meanwhile, a false positive (FP) and a false negative (FN) refer to the number of points associated with landslides that were incorrectly identified as either FP or FN. The AUC assessment was computed using Equation (9):

$$\text{AUC} \ = \ \frac{(\sum \text{TF} \ + \ \sum \text{TN})}{(\text{L} \ + \ \text{N})} \tag{9}$$

where L represents the total number of landslides while N represents the total number of non-landslides.
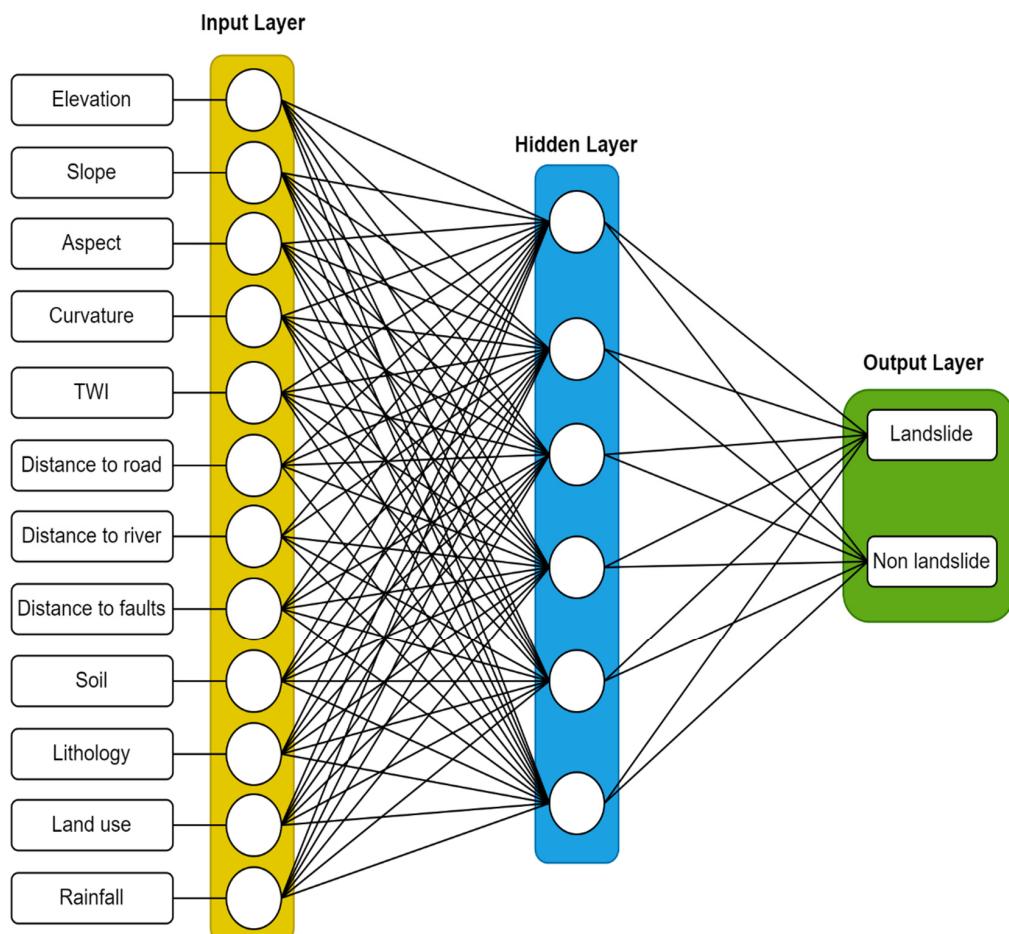

Figure 5. The structure of Artificial Neural Network (ANN) model.

2.3.6. Compound Factor

The compound factor (CF) method attempts to assign consecutive ranks to variables based on their aim of achieving representative relevance [54]. In this study, the CF method was used to choose the best model performance based on the validation results. The relative importance is represented by assigned ranks using an average value of the variable [55]. An expression of this method is seen in Equation (10):

$$\text{CF} \ = \ \frac{1}{\text{n}} \sum_{\text{i}=1}^{\text{n}} \text{R} \tag{10}$$

where n is the number of variables and R is the variable rank. To identify the best fit model for the landslide prediction model, the CF was performed using the evaluation rank of sensitivity, specificity, accuracy, PPV, NPV, AUC, and kappa statistics among four models.

3. Results

3.1. Multicollinearity Analysis

In this study, two types of multicollinearity assessment, such as Pearson correlation, VIF, and TOL, were used to test the correlation between landslide conditioning factors. Pearson correlations were conducted using a pair of landslide conditioning factors. The results show high correlation values between elevation and slope (0.619), slope and TWI (0.662), TWI and distance to the river (0.613), TWI and soil (0.690), and distance to faults and soil (0.658). Based on the results, it was found that there is no strong correlation between each of the landslide conditioning factors, as shown in Table 2. The Pearson correlation value is still under the tolerance limit, which is no more than 0.7. The next step for the multicollinearity assessment for landslide conditioning factors was employing VIF and TOL. The results show that the highest VIF value is 5.683 and the lowest TOL value is 0.176, as shown in Table 3. All the landslide conditioning factors are within the critical thresholds, which are a value of VIF of no more than 10 and a TOL value of no less than 0.1. Therefore, all the landslide conditioning factors represent no multicollinearity and were selected for the landslide modelling process.

Table 2. Pearson correlation between landslide conditioning factors.

Conditioning Factors	Elevation	Slope	Aspect	Curvature	TWI	Dist. to River	Dist. to Road	Dist. to Faults	Land Use	Lithology	Soil	Rainfall
Elevation	1	-	-	-	-	-	-	-	-	-	-	-
Slope	0.619	1	-	-	-	-	-	-	-	-	-	-
Aspect	−0.095	−0.141	1	-	-	-	-	-	-	-	-	-
Curvature	−0.067	−0.039	0.468	1	-	-	-	-	-	-	-	-
TWI	−0.470	−0.662	0.054	−0.050	1	-	-	-	-	-	-	-
Distance to river	−0.288	−0.344	0.044	−0.058	0.613	1	-	-	-	-	-	-
Distance to road	0.017	−0.240	0.088	−0.074	0.417	0.303	1	-	-	-	-	-
Distance to faults	−0.469	−0.517	0.085	−0.005	0.827	0.513	0.413	1	-	-	-	-
Land use	−0.409	−0.262	0.118	0.071	−0.044	−0.032	−0.242	−0.075	1	-	-	-
Lithology	−0.107	−0.298	−0.006	−0.108	0.541	0.328	0.416	0.467	−0.290	1	-	-
Soil	−0.548	−0.541	0.109	−0.033	0.690	0.555	0.288	0.658	0.134	0.270	1	-
Rainfall	0.022	0.277	−0.142	0.132	−0.393	−0.162	−0.188	−0.343	0.183	−0.214	−0.447	1

Table 3. Variance Inflation Factors (VIF) and Tolerance (TOL) analysis.

Conditioning Factors	Collinearity Statistics	
	Tolerance	VIF
Elevation	0.408	2.452
Slope	0.383	2.612
Aspect	0.701	1.427
Curvature	0.718	1.393
TWI	0.176	5.683
Distance to river	0.532	1.880
Distance to road	0.672	1.487
Distance to faults	0.214	4.665
Land use	0.612	1.634
Lithology	0.597	1.676
Soil	0.264	3.783
Rainfall	0.578	1.730

3.2. Important Landslide Conditioning Factors

In this study, the Relief-F method was used to analyze the importance of landslide conditioning factors at LRB using training and testing datasets divided by the ratios of

50:50, 60:40, 70:30, and 80:20. Figure 6 shows the important landslide conditioning factors for model prediction for all models. Based on the findings, rainfall is the most significant landslide conditioning factor for all predictive models, followed by soil and lithology.
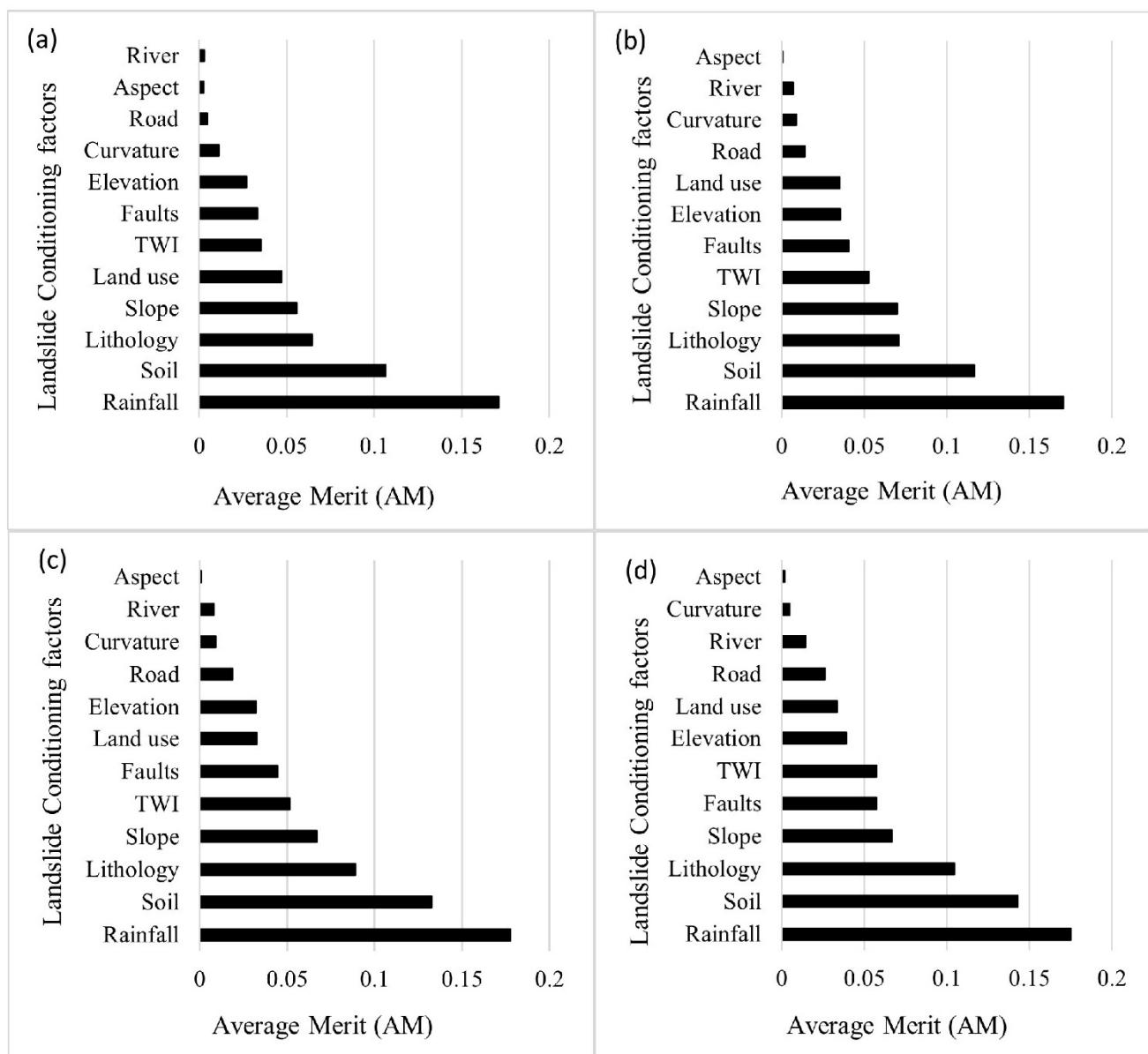

Figure 6. Average merit of landslide conditioning factor. (**a**) Model 50:50, (**b**) Model 60:40, (**c**) Model 70:30, and (**d**) Model 80:20.

3.3. Validation Models Performance

In this study, the landslide predictive model performed a validation assessment for the training and testing datasets as shown in Table 4. Based on the results of validation, model performance for the training dataset shows that model 80:20 was the highest for AUC with 0.931, followed by 70:30 (AUC = 0.918), 60:40 (AUC = 0.872), and 50:50 (AUC = 0.829). Meanwhile, the highest AUC value for the testing dataset is model 60:40 with 0.977, followed by 50:50 (AUC = 0.976), 80:20 (AUC = 0.964), and 70:30 (AUC = 0.957). The CF approach was used in this study to determine the relative priority ranking of model selection by considering all the statistical measures results obtained from training and testing datasets. The training dataset showed the highest-priority model was assigned to

model 80:20 at rank 1, followed by 70:30 (rank 2), 60:40 (rank 3), and 50:50 (rank 4). The training dataset also showed the same results.

Table 4. Validation model performance for training and testing dataset.

Statistical Measure	Training Dataset							
	Model Ratio				Rank			
	50:50	60:40	70:30	80:20	50:50	60:40	70:30	80:20
Sensitivity	0.806	0.854	0.878	0.873	4	3	1	2
Specificity	0.744	0.860	0.878	0.959	4	3	2	1
Accuracy	0.771	0.857	0.878	0.911	4	3	2	1
Positive Predictive Value	0.714	0.854	0.878	0.965	4	3	2	1
Negative predictive value	0.829	0.860	0.878	0.855	4	2	1	3
AUC	0.829	0.872	0.918	0.931	4	3	2	1
Kappa statistic	0.543	0.714	0.755	0.821	4	3	2	1
Rank Total					28	20	12	10
Compound factor (CF)					4.00	2.86	1.71	1.43
Priority Rank					4	3	2	1
	Testing dataset							
Sensitivity	0.906	0.923	0.909	0.923	3	1	2	1
Specificity	0.842	0.833	0.950	0.933	3	4	1	2
Accuracy	0.871	0.875	0.929	0.929	3	2	1	1
Positive Predictive Value	0.829	0.828	0.952	0.923	3	4	1	2
Negative predictive value	0.914	0.926	0.905	0.933	3	2	2	1
AUC	0.976	0.977	0.957	0.964	2	1	4	3
Kappa statistic	0.743	0.751	0.857	0.858	4	3	2	1
Rank Total					21	17	13	11
Compound factor (CF)					3.00	2.43	1.86	1.57
Priority Rank					4	3	2	1

3.4. Landslide Susceptibility Maps (LSMs)

The LSMs were prepared by using an ANN predictive model with different sampling ratios as shown in Figure 7. The LSMs were divided into five landslide susceptibility categories such as very low, low, moderate, high, and very high. The classification was performed using the natural break method of Jenk. This technique has been used extensively in previous studies to classify landslide susceptibility maps [56]. Based on the results of LSM for the landslide predictive model, the sampling ratio of 50:50 showed that the percentage of the very-low-susceptibility category was 27.92%, low susceptibility was 11.63%, moderate was 17.86%, high susceptibility was 17.71, and very high susceptibility was 24.89%. Next, LSM for the landslide predictive model with a 60:40 sampling ratio showed that very low, low, moderate, high, and very high susceptibility categories were 20.76%, 12.98%, 24.42%, 25.87%, and 15.98%, respectively. Meanwhile, LSM for the landslide predictive model with a 70:30 sampling ratio showed 26.11% as very low, 11.32% as low, 19.49% as moderate, 17.67% as high, and 25.40% as very high susceptibility of landslide categories. Lastly, LSM for the landslide predictive model with an 80:20 sampling ratio showed 24.95%, 12.73%, 19.18%, 17.35%, and 25.79% of LRB area in very low, low, moderate, high, and very high landslide susceptibility categories. The percentage of landslide susceptibility area is shown in Figure 8.
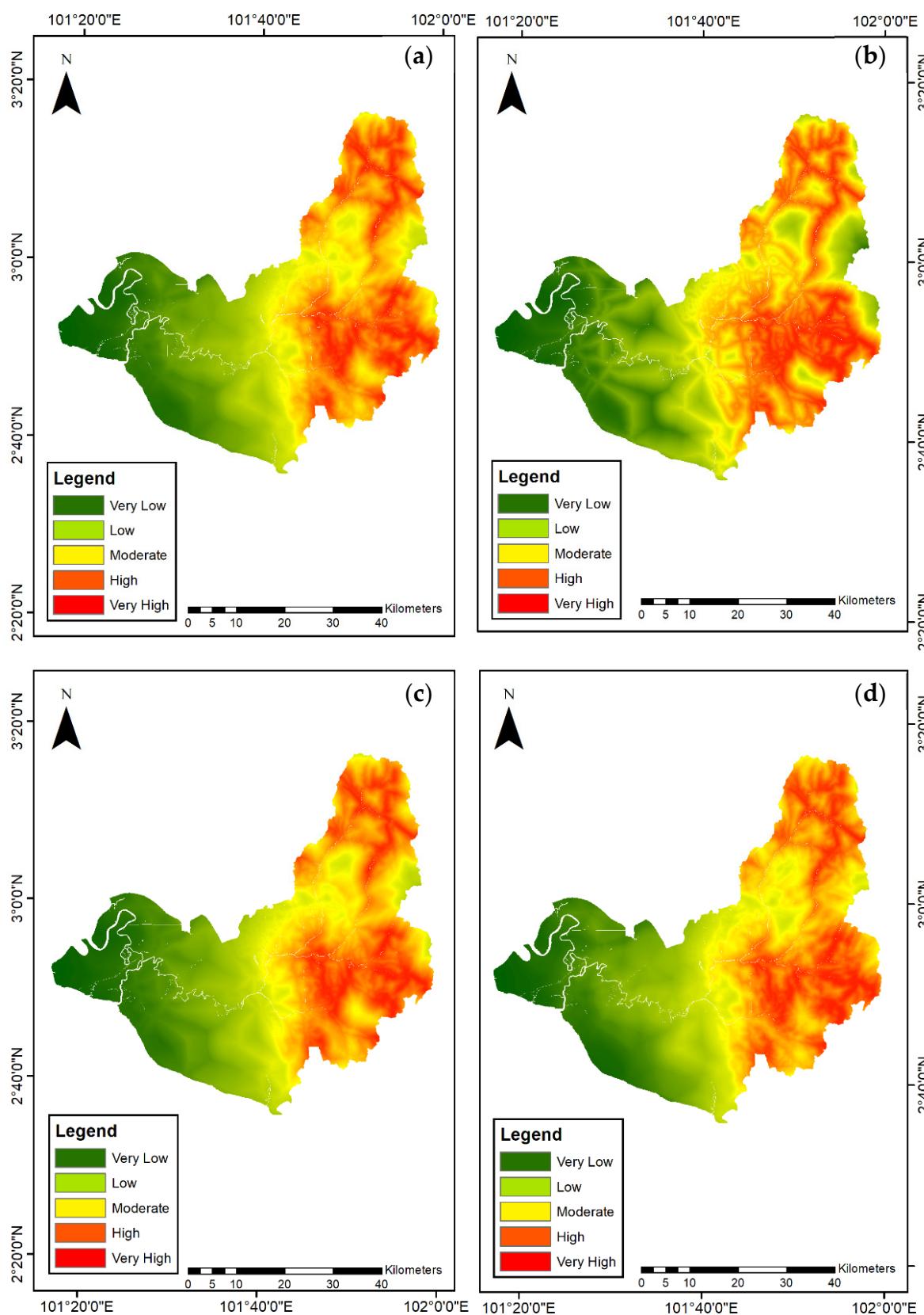
Figure 7. LSMs: (**a**) Model 50:50, (**b**) Model 60:40, (**c**) Model 70:30, and (**d**) Model 80:20.
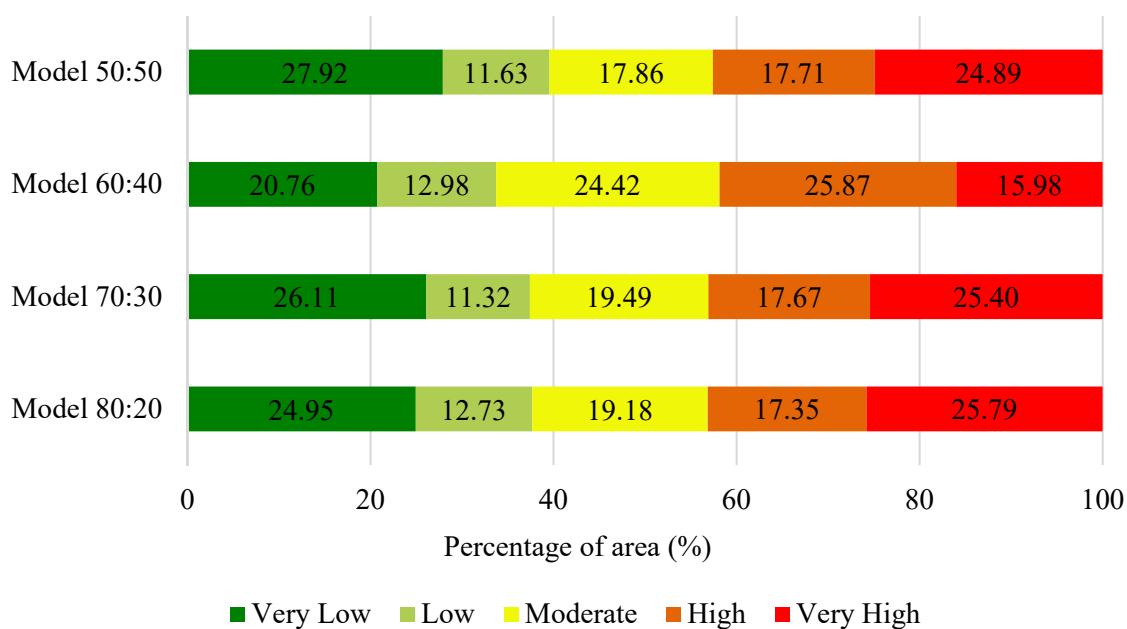
Figure 8. Percentage of landslide susceptibility area.

4. Discussion

The landslide predictive model was a useful instrument for identifying susceptible areas and making predictions about the possibility of landslides, and the result was a map illustrating the landslide susceptibility area. Generally, the landslide prediction model development applied a simple principle when attempting to predict the landslide, in which past and present information is the best indicator of future prediction [57]. The spatial and temporal probability of landslide occurrence must be quantified in order to estimate future landslide behavior, frequencies, extents, and impacts based on past and present landslide information. Many researchers have utilized the ANN as a landslide predictive model, and these researchers concur that it is a reliable prediction model [58–60].

A total of 12 landslide conditioning factors (elevation, slope, curvature, aspect, TWI, distance to the river, distance to the road, distance to the conditioning fault, soil, lithology, land use, and rainfall) were selected based on a literature review of existing landslide studies. The selection of factors for developing the landslide predictive model was an important task that quantifies the quality of the landslide predictive model [23]. Hence, all the selected conditioning factors underwent multicollinearity analysis using Pearson correlation, VIF, and TOL. Multicollinearity analysis is used to identify the correlation between the selected conditioning factors. Based on the results, the Pearson correlation analysis shows that all the selected factors were under the accepted tolerance, which is less than 0.7. A strong correlation exists when the absolute value of the correlation between two conditioning factors is more than 0.7 [61]. Thus, a strongly correlated conditioning factor exists when two independent variables have the same influence on a single independent response variable [62]. Meanwhile, the result of the VIF value is less than 10, and the TOL value is greater than 0.1, which are both considered under the tolerance limit. Therefore, all the landslide conditioning factors are appropriate for the landslide predictive model.

In this research, the ANN was chosen for landslide predictive model assessment with different sampling ratios in the LRB area. A comprehensive assessment was performed in order to determine the best accuracy of the landslide predictive model. The training and testing datasets were randomly divided into different ratios, namely, 50:50, 60:40, 70:30, and 80:20. The identification of the most important conditioning factors that allow for the correct interpretation of the spatial pattern of landslide susceptibility is crucial to spatially explicit landslide modelling [10]. In this study, the Relief-F method was used to determine the importance of landslide conditioning factors. Based on the results, all

the average merit values are greater than zero. The greater the AM value, the more significant the landslide factors that influence the occurrences. Based on the results of important landslide conditioning factors, it was found that rainfall, soil, and lithology are consistently recognized as the most important factors for all predictive models that influence the landslide occurrences in the LRB.

This finding indicates that rainfall was an important factor that influenced landslide occurrence, which is in line with Tajudin et al. [63], Maturidi et al. [64], and several other researchers. The annual rainfall in the LRB ranges from 144.58 mm to 296.254 mm on average. The month of November receives the highest annual average of rainfall with 354 mm, while the lowest was in February with 75.84 mm. These extremes in rainfall occurred during the northeast monsoon and the southwest monsoon. The active pressure of a slope eventually increases due to an increase in pore water pressure that acts as lateral pressure due to rainfall [14]. In addition, rainwater infiltration acts as a softening agent, which increases the probability of slope instability and influences landslide occurrences. This finding found that soil and lithology is the important landslide conditioning factor for landslide occurrences, which also agrees with Yamusa et al. [65], Roslee et al. [66], and Sulaiman et al. [67]. The possibility of landslide occurrence is significant in arable Steep land, Rengam-Jerangau, and urban land for soil factor. Meanwhile, landslides frequently occurred in acid intrusive, and Schist and gneiss. Similar to the study by Sulaiman et al. [67], very high and critical risk categories for landslides are influenced by urban land due to the soil factor and acid intrusive due to the lithology factor.

An accurate landslide predictive model is capable of helping to produce a good-quality landslide susceptibility map. In this study, the accuracy of the landslide predictive model was evaluated using several statistical measures such as sensitivity, specificity, accuracy, PPV, NPV, AUC, and the Kappa statistic. Considering the various techniques of validation performance are more effective in resolving the issues compared to using a single validation approach [56], the CF approach was performed in this study to determine the most reliable predictive model by considering seven statistical measure assessments. This study has emphasized the importance of examining and comparing the predictive landslide model with different sampling ratios for training and testing data because a small value of validation performance can increase the quality of the landslide susceptibility map. Different sample ratios for training and testing datasets were performed to analyze the accuracy and reliability of landslide predictive models.

Based on all statistical measurements, the landslide predictive model with a 50:50 sample ratio had the lowest validation performances for the training dataset when compared to other models. Meanwhile, this model had the second highest AUC value of 0.976 for the testing dataset out of all the models. Next, the landslide predictive model with a 60:40 sample ratio had the second-highest NPV value of 0.860 for the training dataset. In contrast, this model had the highest sensitivity and AUC values of 0.923 and 0.977, respectively, for the testing dataset. The landslide predictive model with a 70:30 sample ratio had the highest value of sensitivity and NPV with 0.878 for the training dataset. Hence, this model had the highest specificity, accuracy, and PPV values of 0.950, 0.929, and 0.952, respectively, for the testing dataset. Lastly, the predictive model with an 80:20 sample ratio had the highest specificity, accuracy, PPV, AUC, and kappa statistic values with 0.959, 0.911, 0.965, 0.931, and 0.821, respectively, for the training dataset. Furthermore, this model had the highest values of sensitivity, accuracy, NPV, and kappa statistic with 0.923, 0.929, 0.933, and 0.858, respectively, for the testing dataset. The CF approach was used to rank the accuracy and reliability of model selection in the landslide predictive model based on validation model performance. The sampling ratio of 80:20 achieved the highest predictive accuracy and reliability among the landslide predictive models, followed by 70:30, 60:40, and 50:50 in training and testing datasets. This finding aligns with the studies by Roslee et al. [66], Shirzadi et al. [68], and Su et al. [69], which found that an 80:20 sampling ratio was able to obtain the highest accuracy of the landslide predictive model.

Malaysia has been committed to sustainable development since the 1970s. The Sustainable Development Goals (SDGs) of Malaysia are carried on by the Millennium Development Goals (MDGs), which were completed in 2015. The 17 SDGs, each with a distinct target, are anticipated to be completed by 2030. The 13th goal is to immediately begin countering the effects of climate change. In the LRB, landslides are natural disasters that occur due to environmental factors and can be accelerated by anthropogenic activities. On the other hand, slope stability is drastically and regularly impacted by human activities, and slope movements are triggered by construction sites that are unsuitable in high-susceptibility areas. The risk of landslides increases mostly due to uncontrolled urbanization, inappropriate land use, and continuous deforestation [70,71]. Landslides are affected by changes in climate conditions, especially the distribution and amount of precipitation and the occurrence of seasonal rain cycles. Researchers have offered a variety of approaches to mitigate the negative consequences of landslide occurrences [72]. The state government will be able to plan and manage processes better with the help of the landslide prediction model. In addition, local governments should embrace and incorporate disaster risk reduction strategies specific to their communities. The impacts of disasters are felt most profoundly and instantly at the local level. Land use legislation, building code enforcement, and basic environmental protection and regulatory compliance functions are crucial for effective disaster risk reduction. Governments and communities should cooperate better on disaster risk reduction, sustainable growth, and environmental management.

5. Conclusions

Landslides constitute a natural hazard that continues to occur in the LRB area that causes property damage and fatalities. However, the development of a landslide prediction model can reduce the risk of this disaster by preventing any human activity, such as road construction and housing development, in areas prone to landslide risk. In this study, a comprehensive assessment of sampling ratios using an ANN for a landslide predictive model was performed. Different sampling ration for training and testing datasets were used in this study, including 50:50, 60:40, 70:30, and 80:20. According to the findings, all sampling ratios have good competency for the predictive of landslide model in LRB; however, the sampling ratio 80:20 was found to be the best model in terms of comprehensive model performance employing compound factor. Rainfall, soil, and lithology were identified as major conditioning elements influencing landslide occurrences by the landslide predictive model. In accordance with the landslide susceptibility map, the east part of the LRB is prone to landslide occurrences. Therefore, the landslide predictive model will help state authorities in their planning and managing process. It is in line with the Agenda 2030 for Sustainable Development Goals (SDGs) in Malaysia and was aimed to reduce the risk of natural disasters. Achieving sustainable development is the best strategy for the environment and future generations.

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

References

1. Cui, Y.; Cheng, D.; Choi, C.E.; Jin, W.; Lei, Y.; Kargel, J.S. The cost of rapid and haphazard urbanization: Lessons learned from the Freetown landslide disaster. *Landslides* **2019**, *16*, 1167–1176. [CrossRef]
2. Abd Majid, N.; Rainis, R. Aplikasi Sistem Maklumat Geografi (GIS) dan Analisis Diskriminan dalam Pemodelan Kejadian Kegagalan Cerun di Pulau Pinang, Malaysia. *Sains Malays.* **2019**, *48*, 1367–1381. [CrossRef]
3. Hasnat, G.T.; Kabir, M.A.; Hossain, M.A. Major environmental issues and problems of South Asia, particularly Bangladesh. In *Handbook of Environmental Materials Management*; Springer: Cham, Switzerland, 2018; pp. 1–40.
4. Abd Majid, N.; Rainis, R.; Ibrahim, W.M.M.W. Pemodelan ruangan pelbagai jenis kegagalan cerun di Pulau Pinang menggunakan kaedah nisbah kekerapan. *Geografi* **2017**, *5*, 13–26.
5. Majid, N.; Taha, M.; Selamat, S. Historical landslide events in Malaysia 1993–2019. *Indian J. Sci. Technol.* **2020**, *13*, 3387–3399. [CrossRef]
6. Ha, N.D.; Sayama, T.; Sassa, K.; Takara, K.; Uzuoka, R.; Dang, K.; Van Pham, T. A coupled hydrological-geotechnical framework for forecasting shallow landslide hazard—A case study in Halong City, Vietnam. *Landslides* **2020**, *17*, 1619–1634. [CrossRef]
7. Li, Y.; Wang, X.; Mao, H. Influence of human activity on landslide susceptibility development in the Three Gorges area. *Nat. Hazards* **2020**, *104*, 2115–2151. [CrossRef]
8. Lembaga Urus Air Selangor (LUAS). *Langat River Basin Management Plan 2015–2020*; Lembaga Urus Air Selangor (LUAS): Selangor, Malaysia, 2015.
9. Department of Statistics Malaysia. Available online: https://www.dosm.gov.my/v1/index.php?r=column/cthree&menu_id=UmtzQ1pKZHBjY1hVZE95R3RnR0Y4QT09 (accessed on 2 December 2022).
10. 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. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* **2020**, *188*, 104451.
11. Wong, J.L.; Lee, M.L.; Teo, F.Y.; Liew, K.W. A Review of Impacts of Climate Change on Slope Stability. In *Climate Change and Water Security*; Springer: Singapore, 2022; pp. 157–178.
12. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **2018**, *165*, 520–529. [CrossRef]
13. Liu, J.; Wu, Z.; Zhang, H. Analysis of Changes in Landslide Susceptibility according to Land Use over 38 Years in Lixian County, China. *Sustainability* **2021**, *13*, 10858. [CrossRef]
14. Rosly, M.H.; Mohamad, H.M.; Bolong, N.; Harith, N.S.H. An Overview: Relationship of Geological Condition and Rainfall with Landslide Events at East Malaysia. *Trends Sci.* **2022**, *19*, 3464. [CrossRef]
15. Baig, M.F.; Mustafa, M.R.U.; Baig, I.; Takaijudin, H.B.; Zeshan, M.T. Assessment of land use land cover changes and future predictions using CA-ANN simulation for selangor, Malaysia. *Water* **2022**, *14*, 402. [CrossRef]
16. Talukdar, S.; Singha, P.; Mahato, S.; Pal, S.; Liou, Y.-A.; Rahman, A. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sens.* **2020**, *12*, 1135. [CrossRef]
17. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
18. Moayedi, H.; Osouli, A.; Tien Bui, D.; Foong, L.K. Spatial Landslide Susceptibility Assessment Based on Novel Neural-Metaheuristic Geographic Information System Based Ensembles. *Sensors* **2019**, *19*, 4698. [CrossRef]
19. Abd Majid, N.; Rainis, R.; Ibrahim, W.M.M.W. Spatial Modeling Various Types of Slope Failure Using Artificial Neural Network (ANN) In Pulau Pinang, Malaysia. *J. Teknol.* **2018**, *80*, 135–146.
20. Bui, D.T.; Tsangaratos, P.; Nguyen, V.-T.; Liem, N.V.; 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]
21. Kumar, D.; Thakur, M.; Dubey, C.S.; Shukla, D.P. Landslide susceptibility mapping & prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India. *Geomorphology* **2017**, *295*, 115–125. [CrossRef]
22. Hong, H. Landslide Susceptibility Mapping in the Youfang area (China) using Dagging-Random Forest model. In Proceedings of the AGU Fall Meeting Abstracts, Washington, DC, USA, 10–14 December 2018; p. NH21C-0848.
23. Kavzoglu, T.; Teke, A. Predictive Performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arab. J. Sci. Eng.* **2022**, *47*, 7367–7385. [CrossRef]
24. He, Q.; Xu, Z.; Li, S.; Li, R.; Zhang, S.; Wang, N.; Pham, B.T.; Chen, W. Novel entropy and rotation forest-based credal decision tree classifier for landslide susceptibility modeling. *Entropy* **2019**, *21*, 106. [CrossRef]
25. Arabameri, A.; Chandra Pal, S.; Rezaie, F.; Chakrabortty, R.; Saha, A.; Blaschke, T.; Di Napoli, M.; Ghorbanzadeh, O.; Thi Ngo, P.T. Decision tree based ensemble machine learning approaches for landslide susceptibility mapping. *Geocarto Int.* **2022**, *37*, 4594–4627. [CrossRef]
26. Abu El-Magd, S.A.; Ali, S.A.; Pham, Q.B. Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain. *Earth Sci. Inform.* **2021**, *14*, 1227–1243. [CrossRef]

27. Youssef, A.M.; Pourghasemi, H.R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geosci. Front.* **2021**, *12*, 639–655. [CrossRef]

28. 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]

29. Oh, H.-J.; Lee, S. Shallow landslide susceptibility modeling using the data mining models artificial neural network and boosted tree. *Appl. Sci.* **2017**, *7*, 1000. [CrossRef]

30. Gautam, P.; Kubota, T.; Sapkota, L.M.; Shinohara, Y. Landslide susceptibility mapping with GIS in high mountain area of Nepal: A comparison of four methods. *Environ. Earth Sci.* **2021**, *80*, 1–18. [CrossRef]

31. Manap, M.A.; Nampak, H.; Pradhan, B.; Lee, S.; Sulaiman, W.N.A.; Ramli, M.F. Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. *Arab. J. Geosci.* **2014**, *7*, 711–724. [CrossRef]

32. Amirabadizadeh, M.; Huang, Y.F.; Lee, T.S. Recent trends in temperature and precipitation in the Langat River Basin, Malaysia. *Adv. Meteorol.* **2015**, *2015*, 579437. [CrossRef]

33. Lee, D.-H.; Kim, Y.-T.; Lee, S.-R. Shallow landslide susceptibility models based on artificial neural networks considering the factor selection method and various non-linear activation functions. *Remote Sens.* **2020**, *12*, 1194. [CrossRef]

34. Nhu, V.-H.; Mohammadi, A.; Shahabi, H.; Ahmad, B.B.; Al-Ansari, N.; Shirzadi, A.; Clague, J.J.; Jaafari, A.; Chen, W.; Nguyen, H. Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4933. [CrossRef]

35. Deng, X.; Li, L.; Tan, Y. Validation of spatial prediction models for landslide susceptibility mapping by considering structural similarity. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 103. [CrossRef]

36. Kornejady, A.; Ownegh, M.; Bahremand, A. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena* **2017**, *152*, 144–162. [CrossRef]

37. Sadr, M.P.; Maghsoudi, A.; Saljoughi, B.S. Landslide susceptibility mapping of Komroud sub-basin using fuzzy logic approach. *Geodyn. Res. Int. Bull.* **2014**, *2*, 16–28.

38. 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]

39. Li, Y.; Chen, W. Landslide susceptibility evaluation using hybrid integration of evidential belief function and machine learning techniques. *Water* **2020**, *12*, 113. [CrossRef]

40. Pham, B.T.; Bui, D.T.; Dholakia, M.B.; Prakash, I.; Pham, H.V.; Mehmood, K.; Le, H.Q. A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. *Geomat. Nat. Hazards Risk* **2016**, *8*, 649–671. [CrossRef]

41. Różycka, M.; Migoń, P.; Michniewicz, A. Topographic Wetness Index and Terrain Ruggedness Index in geomorphic characterisation of landslide terrains, on examples from the Sudetes, SW Poland. *Z. Für Geomorphol. Suppl. Issues* **2017**, *61*, 61–80. [CrossRef]

42. Saleem, N.; Huq, M.E.; Twumasi, N.Y.D.; Javed, A.; Sajjad, A. Parameters derived from and/or used with digital elevation models (DEMs) for landslide susceptibility mapping and landslide risk assessment: A review. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 545. [CrossRef]

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

44. Yan, G.; Liang, S.; Gui, X.; Xie, Y.; Zhao, H. Optimizing landslide susceptibility mapping in the Kongtong District, NW China: Comparing the subdivision criteria of factors. *Geocarto Int.* **2019**, *34*, 1408–1426. [CrossRef]

45. Xia, J.; Dong, P. Spatial characteristics of physical environments for human settlements in Jinsha River watershed (Yunnan section), China. *Geomat. Nat. Hazards Risk* **2019**, *10*, 544–561. [CrossRef]

46. Chen, W.; Sun, Z.; Han, J. Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. *Appl. Sci.* **2019**, *9*, 171. [CrossRef]

47. Long, N.; De Smedt, F. Analysis and Mapping of Rainfall-Induced Landslide Susceptibility in A Luoi District, Thua Thien Hue Province, Vietnam. *Water* **2018**, *11*, 51. [CrossRef]

48. Shrestha, N. Detecting multicollinearity in regression analysis. *Am. J. Appl. Math. Stat.* **2020**, *8*, 39–42. [CrossRef]

49. Hu, X.; Mei, H.; Zhang, H.; Li, Y.; Li, M. Performance evaluation of ensemble learning techniques for landslide susceptibility mapping at the Jinping county, Southwest China. *Nat. Hazards* **2021**, *105*, 1663–1689. [CrossRef]

50. Selamat, S.N.; Abd Majid, N.; Taha, M.R.; Osman, A. Application of geographical information system (GIS) using artificial neural networks (ANN) for landslide study in Langat Basin, Selangor. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *1064*, 012052. [CrossRef]

51. Li, B.; Wang, N.; Chen, J. GIS-based landslide susceptibility mapping using information, frequency ratio, and artificial neural network methods in Qinghai Province, Northwestern China. *Adv. Civ. Eng.* **2021**, *2021*, 4758062. [CrossRef]

52. Hong, H.; Tsangaratos, P.; Ilia, I.; Loupasakis, C.; Wang, Y. Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping. *Sci. Total Environ.* **2020**, *742*, 140549. [CrossRef]

53. Li, D.; Huang, F.; Yan, L.; Cao, Z.; Chen, J.; Ye, Z. Landslide susceptibility prediction using particle-swarm-optimized multilayer perceptron: Comparisons with multilayer-perceptron-only, bp neural network, and information value models. *Appl. Sci.* **2019**, *9*, 3664. [CrossRef]

54. 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]

55. Hembram, T.K.; Saha, S. Prioritization of sub-watersheds for soil erosion based on morphometric attributes using fuzzy AHP and compound factor in Jainti River basin, Jharkhand, Eastern India. *Environ. Dev. Sustain.* **2020**, *22*, 1241–1268. [CrossRef]

56. Saha, S.; Roy, J.; Hembram, T.K.; Pradhan, B.; Dikshit, A.; Abdul Maulud, K.N.; Alamri, A.M. Comparison between Deep Learning and Tree-Based Machine Learning Approaches for Landslide Susceptibility Mapping. *Water* **2021**, *13*, 2664. [CrossRef]

57. Ma, Z.; Mei, G.; Piccialli, F. Machine learning for landslides prevention: A survey. *Neural Comput. Appl.* **2021**, *33*, 10881–10907. [CrossRef]

58. Jacinth Jennifer, J.; Saravanan, S. Artificial neural network and sensitivity analysis in the landslide susceptibility mapping of Idukki district, India. *Geocarto Int.* **2022**, *37*, 5693–5715. [CrossRef]

59. Orhan, O.; Bilgilioglu, S.S.; Kaya, Z.; Ozcan, A.K.; Bilgilioglu, H. Assessing and mapping landslide susceptibility using different machine learning methods. *Geocarto Int.* **2022**, *37*, 2795–2820. [CrossRef]

60. Mehrabi, M.; Moayedi, H. Landslide susceptibility mapping using artificial neural network tuned by metaheuristic algorithms. *Environ. Earth Sci.* **2021**, *80*, 1–20. [CrossRef]

61. Qin, Y.; Yang, G.; Lu, K.; Sun, Q.; Xie, J.; Wu, Y. Performance Evaluation of Five GIS-Based Models for Landslide Susceptibility Prediction and Mapping: A Case Study of Kaiyang County, China. *Sustainability* **2021**, *13*, 6441. [CrossRef]

62. Kalantar, B.; Ueda, N.; Saeidi, V.; Ahmadi, K.; Halin, A.A.; Shabani, F. Landslide Susceptibility Mapping: Machine and Ensemble Learning Based on Remote Sensing Big Data. *Remote Sens.* **2020**, *12*, 1737. [CrossRef]

63. Tajudin, N.; Ya'acob, N.; Ali, D.M.; Adnan, N.; Naim, N.F. Rainfall–landslide potential mapping using remote sensing and GIS at Ulu Kelang, Selangor, Malaysia. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *169*, 012080. [CrossRef]

64. Maturidi, A.M.A.M.; Kasim, N.; Taib, K.A.; Azahar, W.N.A.W.; Tajuddin, H.B.A. Empirically Based Rainfall Threshold for Landslides Occurrence in Peninsular Malaysia. *KSCE J. Civ. Eng.* **2021**, *25*, 4552–4566. [CrossRef]

65. Yamusa, I.B.; Ismail, M.S.; Tella, A. Geospatial Detection of Hidden Lithologies along Taiping to Ipoh Stretch of the Highway Using Medium Resolution Satellite Imagery in Malaysia. *J. Adv. Geospat. Sci. Technol.* **2021**, *1*, 19–37.

66. Roslee, R.; Sharir, K.; Lai, G.T.; Simon, N.; Ern, L.K.; Madran, E.; Saidin, A.S. Application of Analytical Hierarchy Process (AHP) for Landslide Hazard Analysis (LHA) in Kota Kinabalu area, Sabah, Malaysia. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *1103*, 012031. [CrossRef]

67. Sulaiman, M.S.; Nazaruddin, A.; Salleh, N.M.; Abidin, R.Z.; Miniandi, N.D.; Yusoff, A.H. Landslide occurrences in Malaysia based on soil series and lithology factors. *Int. J. Adv. Sci. Technol.* **2019**, *28*, 1–26.

68. Shirzadi, A.; Soliamani, K.; Habibnejhad, M.; Kavian, A.; Chapi, K.; Shahabi, H.; Chen, W.; Khosravi, K.; Thai Pham, B.; Pradhan, B. Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* **2018**, *18*, 3777. [CrossRef] [PubMed]

69. Su, Q.; Zhang, J.; Zhao, S.; Wang, L.; Liu, J.; Guo, J. Comparative assessment of three nonlinear approaches for landslide susceptibility mapping in a coal mine area. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 228. [CrossRef]

70. Barancokova, M.; Sosovicka, M.; Barancok Jr, P.; Barancok, P. Predictive Modelling of Landslide Susceptibility in the Western Carpathian Flysch Zone. *Land* **2021**, *10*, 1370. [CrossRef]

71. Abd Majid, N.; Zulkafli, S.A.; Zakaria, S.Z.S.; Razman, M.R.; Ahmed, M.F. Spatial Pattern Analysis on Landslide Incidents in Kuala Lumpur, Malaysia. *Ecol. Environ. Conserv.* **2022**, *28*, 1624–1627. [CrossRef]

72. Chen, X.; Chen, W. GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. *Catena* **2021**, *196*, 104833. [CrossRef]