MDPI

Article

Landslide Susceptibility Assessment for Maragheh County, Iran, Using the Logistic Regression Algorithm

Ahmed Cemiloglu [1], Licai Zhu [1], Agab Bakheet Mohammednour [2], Mohammad Azarafza [3] and Yaser Ahangari Nanehkaran [1,*]

[1] School of Information Engineering, Yancheng Teachers University, Yancheng 224002, China; ahmed@yctu.edu.cn (A.C.); zhulc01@yctu.edu.cn (L.Z.)

[2] Department of Control System Engineering, Al-Neelain University, Khartoum 12702, Sudan; agab.bakheet@neelain.edu.sd

[3] Geotechnical Department, Faculty of Civil Engineering, University of Tabriz, Tabriz 5166616471, Iran; m.azarafza.geotech@gmail.com

[*] Correspondence: yaser@yctu.edu.cn

Abstract: Landslide susceptibility assessment is the globally approved procedure to prepare geo-hazard maps of landslide-prone areas, which are highly used in urban management and minimizing the possible disasters due to landslides. Multiple approaches to providing susceptibility maps for landslides have one specification. Logistic regression is a statistical-based model that investigates the probabilities of the events which is received extensive success in landslide susceptibility assessment. The presented study attempted to use a logistic regression application to prepare the Maragheh County hazard risk map. In this regard, several predisposing factors (e.g., elevation, slope aspect, slope angle, rainfall, land use, lithology, weathering, distance from faults, distance from the river, distance from the road, and distance from cities) are identified as main responsible for landslide occurrence and 20 historical sliding events which used to prepare hazard risk maps. As verification, the models were controlled by operating relative characteristics (ROC) curves which reported the overall accuracy for susceptibility assessment. According to the results, the region is located in a moderate to high-hazard risk zone. The north and northeast parts of Maragheh County show high suitability for landslides. Verification results of the model indicated that the AUC estimated for the training set is 0.885, and the AUC estimated for the testing set is 0.769. To justify the model, the results of the LR were comparatively checked with several benchmark learning models. Results indicated that LR model performance is reasonable.

Keywords: landslides; susceptibility analysis; logistic regression; hazard mapping; geo-hazards

1. Introduction

Landslides are the second most hazardous geo-hazard phenomenon worldwide that cause countless damages to infrastructures and loss of lives. Geologically, the wide variety of mass movements on the earth-surface that triggered by certain influence factors where the movable mass moved into the downward slope by static and/or dynamic forces [1–4]. The cause of the movement of the mass can be gravity, earthquake, road construction, heavy rain or water-pore pressure, lightning downstream of the slope, and so on [5]. These movements can occur very slowly (only a few millimeters per year), or they can happen very quickly and have disastrous effects. Landslides can even appear on the seabed and underwater, creating tidal waves that destroy coastal areas [6]. According to the European Geotechnical Thematic Network (EGTN), landslides account for about 17% of the world's natural disasters [7]. Therefore, identifying a suitable set of instability factors related to slope failures requires prior knowledge of the main causes of landslides [8]. Landslide susceptibility analysis provides appropriate information about the high potential regions regarding the landslide occurrence probability [9]. Landslide susceptibility mapping

relies on relatively sophisticated knowledge of slope movements and their controlling factors. The reliability of landslide susceptibility maps mostly depends on the amount and quality of available data, the working scale, and the selection of the appropriate analysis procedure [10–14]. The resulting need to predict the occurrence of landslides has led to the development of several stochastic and process-based models that emphasize the use of geographic information systems (GIS). To solve this problem, a scientific evaluation of the landslide-prone area is necessary. In the last decade, the mapping system of remote sensing techniques has facilitated the preparation of geographic information and landslide susceptibility maps [15–18].

Iran is one of the countries involved in landslides which can be triggered by rainfall, unprincipled human activities, or earthquake events. For example, the Kejur earthquake in Mazandaran province in 2013 did cause extensive rockfalls and brought a lot of financial costs for the region [19]. Considering Iran's geological complexity and active tectonically conditions regarding earthquake events, the country is considered as high seismic activity [14]. So, providing landslide susceptibility mapping for different regions is inevitable. Nevertheless, there are specific predisposing factors that are responsible for landslide occurrences (regardless of the scale and mechanism) which provide suitable conditions for massive movements [20], especially in mountainous regions [21]. In such regions, implementing projects related to water pipelines, transportation networks and environmental management, and urban planning must consider requirements to identify landslide risks [13].

Considering the complexity of Iran's platform regarding landslide occurrence, providing a detailed and reliable procedure to provide the susceptibility maps for landslides. In general, different methods have been applied by researchers around the world for landslide hazard susceptibility assessment. These methods can be classified as qualitative or quantitative approaches. Qualitative methods are subjective and are based on the experts' opinions and depict risk zoning descriptively. Quantitative methods can be established based on ground surveys and remote sensing applications that lead to producing quantities. By using specialized knowledge and precise procedure, the direct relationship between the landslides and the triggering parameters responsible for landslides occurrences can be determined [22]. Although each method has its advantages and disadvantages, due to the lack of international standards and uniform instructions, researchers widely use them [23,24]. Among the wide range of quantitative which can be categorized into deterministic, statistical, heuristic, inventory-based, geostatistic, and knowledge-based methods [25–28]. In the meantime, statistical methods have received strong attention from professionals regarding landslide susceptibility [29]. In statistical methods proposed for assessing landslide susceptibility, Logistic Regression (LR) analysis has proven to be one of the most reliable approaches [30–32]. LR is a generalized linear model type which is very suitable for analyzing the presence or absence of a dependent variable and has been used to predict the susceptibility of landslides [33]. LR uses discriminant analysis and likelihood ratio to stepwise variable selection and provides more accurate landslide susceptibility by prediction process [34].

LR is a valuable tool for landslide susceptibility analysis due to several key reasons. Firstly, landslide susceptibility analysis often involves binary classification, where areas are categorized as either susceptible or not susceptible to landslides. LR is specifically designed for binary classification problems, making it well-suited. Secondly, LR provides estimates of the probability that an area belongs to a specific class, which is particularly useful in landslide susceptibility analysis. By estimating the probability of an area being susceptible to landslides based on various input factors, LR enables a more nuanced understanding of the likelihood of landslides occurring in different areas. This probability information is crucial for informed decision-making and prioritizing mitigation efforts [30,31]. Also, LR offers interpretability, allowing us to understand the relative importance and influence of different predisposing factors in landslide susceptibility [34]. The coefficients associated

with each independent variable provide insights into the strength and direction of their impact on the probability of landslide occurrence [32].

Landslide susceptibility analysis considers multiple input factors, such as predisposing factors and historical landslide records. LR can handle continuous and categorical variables, accommodating a wide range of data types and allowing for a comprehensive analysis. Moreover, LR models are transparent and understandable, expressing the relationship between the input factors and the probability of landslide susceptibility in a straightforward manner. This transparency enables stakeholders to grasp the underlying mechanisms and potentially validate the model against their domain knowledge. Lastly, logistic regression has a well-established methodology with numerous techniques for model assessment, including goodness-of-fit tests and evaluation metrics like operating relative characteristics (ROC) analysis [32,33]. This rich body of knowledge surrounding logistic regression provides confidence in its applicability and reliability for landslide susceptibility analysis. In conclusion, logistic regression is a valuable and appropriate choice for landslide susceptibility analysis due to its compatibility with binary classification, probability estimation, interpretability, ability to handle multiple input factors, transparency, and established methodology. Its utilization can enhance our understanding of landslide susceptibility and inform effective mitigation strategies [31–34].

Lee et al. [35] used frequency ratio and LR models to prepare landslide susceptibility maps in GIS. The variables used by LR are slope angle, topography, elevation, land-land use, soil material, drainage pattern, effective soil thickness, forest type, tree diameter, tree age, and forest density. The results showed the high accuracy of the logistic regression model prediction compared to the frequency ratio model. Ayalew et al. [36] have studied landslide susceptibility conditions for Sado Island landslides in Japan. The authors used lithology, topography and slope angle as basics and implemented logistic regression and analysis hierarchy (AHP) to cover the parametric analysis of predisposing factors. Greco et al. [37] presented landslide susceptibility analysis based on LR in Cambria, Italy. The researchers used six main predisposing factors to develop the landslide risk maps. As a result, the logistic regression can predict and detect the prone area regarding land sliding. Yalcin et al. [38] provide a comparative analysis based on LR, AHP and frequency ratio to estimate the suitable methodology for landslide susceptibility mapping performed on 20 active landslide areas in Turkey. The results indicated that logistic regression is getting more accurate than other approaches.

Ozdemir [39] used frequency ratio, weights of evidence and LR methods in landslide susceptibility assessment in the Sultan Mountains of Turkey. The study aims to establish a comparative study to understand the capability of these procedures. As a result, LR reaches the highest performance of other methods. Shahabi et al. [40] prepare research to analyse the AHP, LE and frequency ratio capability to develop landslide susceptibility maps in the central Zab basin of Iran by considering eight different predisposing factors. Researchers used the coefficient of determination (R^2) and ROC curve to evaluate the predictive model's performance. As a result, the LR provided better predictions. In their study, Chen et al. [30] introduced a new kernel logistic regression-based model approach named the 'BKLR model', developed for spatial susceptibility analysis of landslides in Shangnan, China. In this research, 15 conditioning factors were selected and entered into the process, which leads to providing landslide susceptibility maps of the studied area. Tekin [41] provides a study on landslide susceptibility mapping by using LR and landslide inventory methods which were implemented in the Ceyhan watershed in Turkey. The process applied to several predisposing factors and was controlled by the ROC curve and R2. Results indicate that the LR model operated with 84.2% overall accuracy to obtain the final susceptibility map. Nwazelibe et al. [42] utilized a comparative study on LR and weight-of-evidence (WoE) algorithms to provide Orumba North region (Nigeria) suitability for landslides. The models were verified and controlled by ROC overall accuracy. As a result, LR provided more accurate results than WoE (WoE and LR results as 0.986 and 0.995, respectively). Abeysiriwardana and Gomes [6] used GIS and LR to analyse the impact

of vegetation on soil properties (moisture, compaction, ...) and landslide susceptibility conditions. The model was validated by confusion matrix and ROC overall accuracy. Gu et al. [43] used a geographically weighted logistic regression model (GWLR) to provide the landslide susceptibility maps for Zhenxiong County, China. The authors operated the model on 2015 historical data of occurred landslides and about ten predisposing factors, which led to preparing landslide susceptibility maps. The model was implemented with 90.4% accuracy to provide results controlled by ROC curve analysis.

It is important to note that applying LR to landslide susceptibility, several key points should be considered. LR is particularly suitable for binary classification problems, making it well-suited for distinguishing between areas that are susceptible or not susceptible to landslides. By modeling the relationship between input factors and the binary outcome, LR helps identify factors significantly contributing to landslide susceptibility [34]. Also, LR provides interpretable results [31,32]. The LR model's coefficients associated with each input factor can be analyzed to understand the magnitude and direction of their impact on landslide susceptibility [33,35]. This interpretability allows researchers to gain insights into the relative importance of different triggering factors in the study area [36]. LR can accommodate various types of input variables. It can handle continuous and categorical factors, which is crucial when considering the diverse predisposing factors in landslide susceptibility analysis [35].

Furthermore, LR can effectively handle small datasets, which is often the case in landslide susceptibility analysis due to the limited availability of historical landslide events. Despite the small number of data points, LR can still provide meaningful predictions and insights into the susceptibility of different areas. Also, LR results can be integrated into a GIS environment, allowing for spatial visualization and analysis. By mapping the predicted landslide susceptibility, decision-makers can identify high-risk areas and incorporate this information into urban planning and land management strategies [41,42].

The presented study attempted to use the LR analysis method to develop the landslide susceptibility maps for Maragheh County, northwest of Iran. The county has a complex geological history that leads to occurred local scale landslides which recorded 20 historical events during field surveys and remote-sensing studies. The model was implemented in Python high-level programming language, and results are entered as information layers in the GIS. Ultimately, these information layers represent the predisposing factors for the studied region susceptibility assessment. The motivation of this study regarding using the LR application for landslide susceptibility analysis for Maragheh County can be classified as its compatibility with binary classification, probability estimation, interpretability, ability to handle multiple input factors, transparency, and established methodology. Its utilization can enhance our understanding of landslide susceptibility and inform effective mitigation strategies.

Furthermore, integrating remote sensing data with ground surveys in Maragheh County has improved the precision and resolution of the input variables in LR modeling. These high-resolution datasets provide valuable information on predisposing factors and historical records. Another significant advancement is incorporating the LR algorithm for the studied area. LR can combine with other algorithms employed to improve predictive accuracy. Also, these advancements contribute to better-informed decision-making processes, allowing for effective mitigation strategies and reducing the impact of landslides on human lives and infrastructure.

The hypothesis underlying the application of LR in landslide susceptibility analysis for Maragheh County suggests a relationship between the identified predisposing factors and the occurrence of landslides. The null hypothesis posits that these factors do not have a significant impact on the susceptibility of an area to landslides. Conversely, the alternative hypothesis proposes that the combination of factors such as elevation, slope aspect, slope angle, rainfall, land use, lithology, weathering, distance from faults, distance from rivers, distance from roads, and distance from cities plays a crucial role in determining the susceptibility of an area to landslides within Maragheh County. By employing LR

and analyzing the relationships between these factors and historical landslide events, the study aims to provide empirical evidence supporting a significant relationship. The goal is to develop a reliable landslide susceptibility map to identify and delineate county areas prone to landslides. This hypothesis-driven approach will better understand the factors influencing landslide susceptibility in Maragheh County and facilitate informed decision-making for land use planning and sustainable development initiatives.

In the context of landslide susceptibility assessment in Maragheh County, LR offers several advantages that make it a suitable approach. Firstly, LR provides interpretable results, allowing for an understanding of the relationship between input variables and the likelihood of landslide occurrence. This interpretability is valuable in identifying the key factors contributing to landslide susceptibility in the specific region. Additionally, LR can effectively handle situations where data availability is limited or of low resolution, which is often the case in landslide studies. It can still yield meaningful insights based on the available data without being overly sensitive to data limitations.

Moreover, LR produces probability estimates that quantify the uncertainty associated with landslide predictions, aiding in decision-making and risk management efforts. The probabilistic nature of LR enables a more nuanced understanding of landslide susceptibility in the studied County. The validation of LR models can be easily conducted using evaluation metrics such as the ROC curve and AUC, providing a measure of the model's performance in discriminating between landslide and non-landslide areas. Suppose LR has been successfully applied in previous studies or similar geospatial analyses in the region. In that case, it justifies its use in subsequent landslide susceptibility assessments, ensuring consistency and facilitating comparisons. Ultimately, the choice of application should be carefully considered, considering the specific characteristics of Maragheh County, available data, and research objectives.

2. Materials and Methods

2.1. Studied Case Location

Maragheh County, with 65.2185 km^2 occupies about 8.4 percent of East Azerbaijan province's total area, which is located in northwest Iran. Maragheh city (capital of Maragheh County) is located at 37°23′21″ N 46°14′15″ E, where limited by Urmia Lake in the west, Tabriz in the north, and Hashtroud in the east [44]. Figure 1 presents the location of the studied region. The climate of Maragheh is moderate. It tends to be cold and relatively humid. The maximum temperature in this city in summer is about 35 degrees Celsius, and its minimum in winter is about 20 degrees Celsius. The annual rainfall in Maragheh is about 330 mm, and its frost days are about 114 days a year. The highest maximum precipitation is in March and April, and the lowest is related to summer [45]. As a topographical condition, Maragheh is located in the Sofichay River Valley (unearthed waterway of Sahand Mountain foothill), a Sofichay River basin. Geologically, Maragheh has been made in alluvial deposits of Sofichay River and Sahand Stratovolcan Mountain activity, located in northern Maragheh. The Sahand Mountain comprises pyroclastic, ignimbrites, dacite, felsic rocks and lavas. So, the north and east parts of the County are on volcanic ashes and the west part is covered by alluvium [46]. Geographical formations from the studied region have varied geological histories and reflect the region's complex tectonic history and tectonic earthquakes. Maragheh city is affected by large and active faults in the region, such as the Tabriz fault, Urmia-Zarinehrud fault, and north and south Maragheh faults, which have caused various earthquakes in the region [46]. A geological map of the studied region is presented in Figure 2.
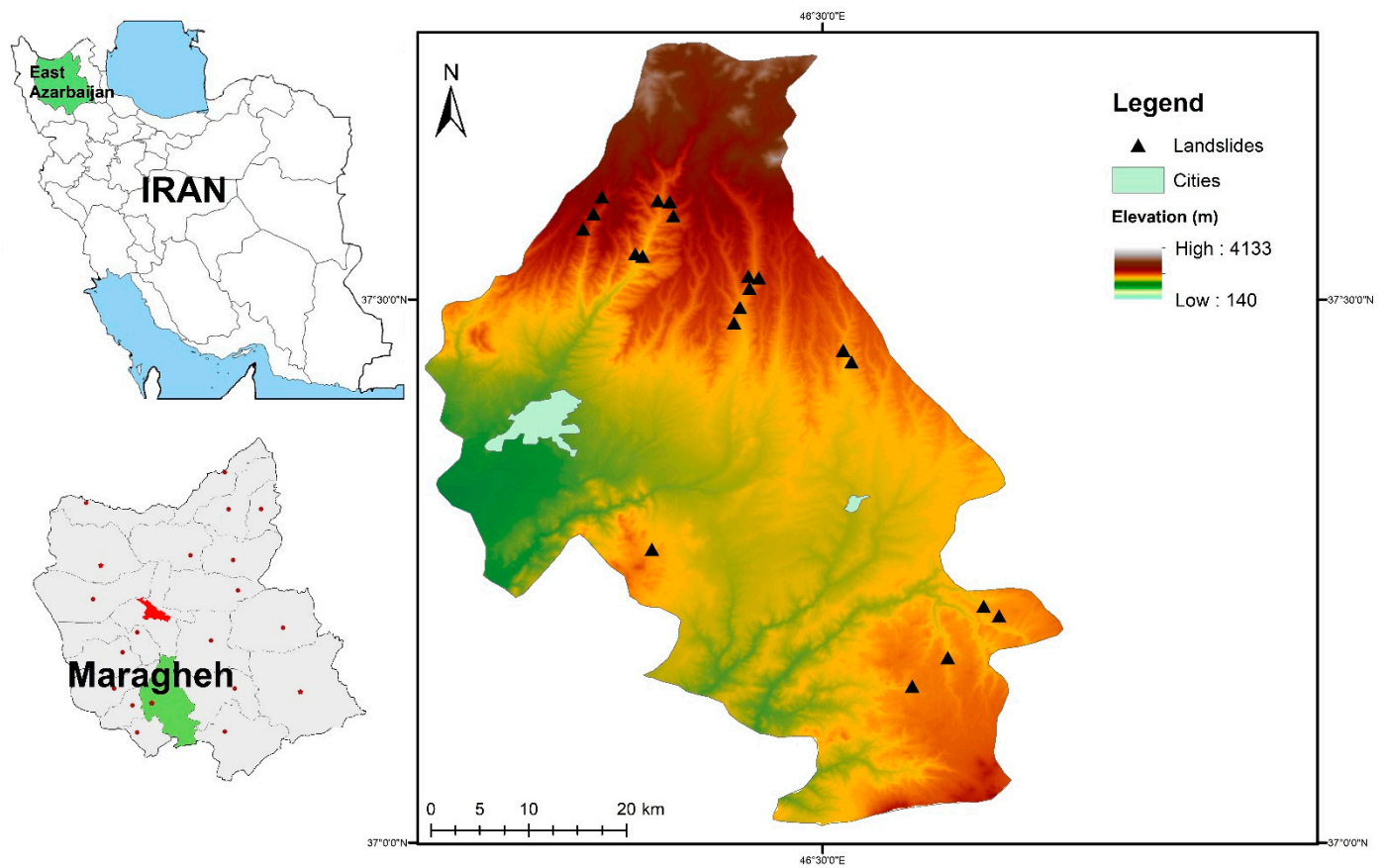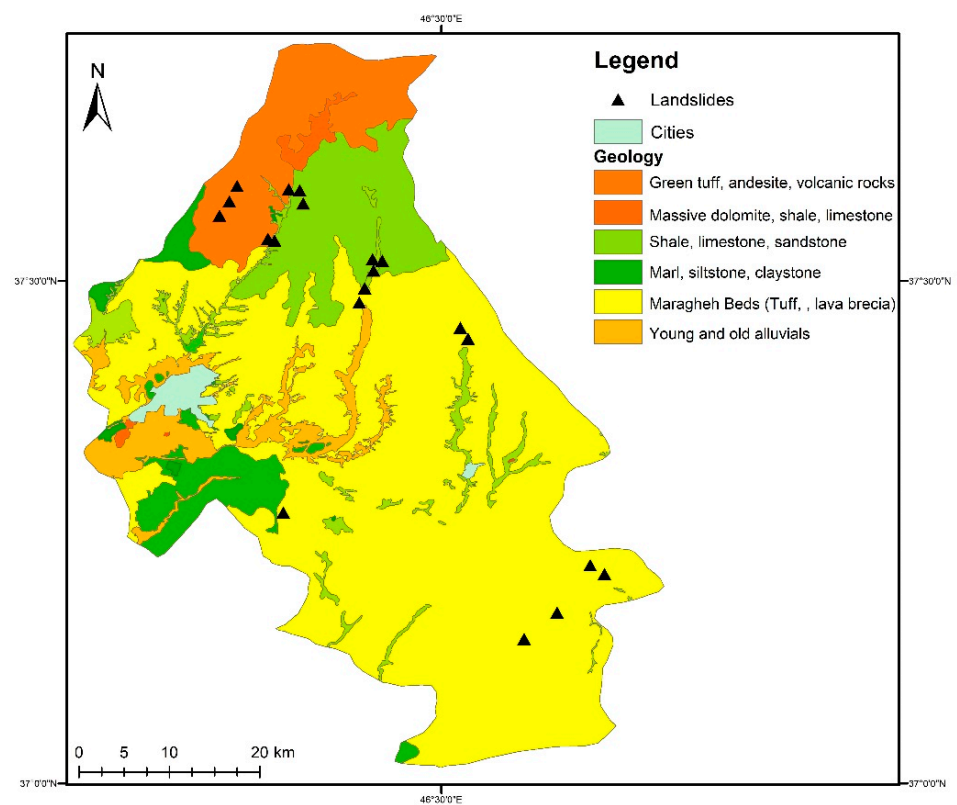
Figure 1. Location of the studied region.

Figure 2. Geological map of the studied region (adapted from Ref. [46]).

Regarding the complexity of the geological background of the Maragheh region. There are several various predisposing factors, including elevation, slope aspect, slope angle, rainfall, land use, lithology, weathering, distance from faults, distance from the river, distance from road and distance from cities were used as main influence parameters in landslide occurrence which are identified as main responsible for landslides. Also, 20 historical landslide events are recorded during the ground survey and remote-sensing assessment. The location of the historical landslides is shows in Figure 1. Generally, selecting predisposing factors plays a key role in landslide susceptibility analysis, considered an essential stage in susceptibility mapping [47]. To estimate proper parameters and have a good understanding of predisposing factors, field surveys and remote sensing observations are required, which are done in the studied region. The selection of the predisposing factors requires several considerations about the triggering elements' dependency, measurability, non-redundancy, and relevance of geological characteristics, which leads to logical and reliable preparations. Table 1 provides relevant information about the selected predisposing factors used in this study. This information was used to develop the parametric maps in GIS regarding the predisposing factors. Figure 3 provides the relative maps regarding estimated predisposing factors in the study. Regarding the multicollinearity, the applied variables in susceptibility analysis for landslides must be correlated. The multicollinearity term is the capability of a predictor variable in a regression model to predict linearly from others [48]. To test for multicollinearity variance, inflation factors (VIF) are common, and if VIF > 5, it indicates potential multicollinearity. In this article, all selected predisposing factors produced VIF values less than 5.00. Figure 4 provides the VIF variations for selected predisposing factors. VIF ratio quantifies the multicollinearity severity in an ordinary least squares regression analysis and presents how much larger the standard error increases. Low value of VIF indicates the less errors during calculations [49,50].

Table 1. A summary of landslide predisposing factors information was used in this study.

Class	Predisposing Factors	Resolution	Variables	Data Source
	Elevation	±30 m	Continuous	DEM
Morphologic	Slope aspect	±30 m	Continuous	DEM
	Slope angle	±30 m	Continuous	DEM
Climatologic	Rainfall	±30 m	Continuous	IMO *
	Land use	±30 m	Discrete	Geological data
	Lithology	±30 m	Discrete	Geological data
Geologic	Weathering	±30 m	Discrete	Landsat TM, ETM+
	distance from faults	±30 m	Continuous	DEM, Google Map
	distance from river	±30 m	Continuous	DEM, Google Map
	distance from road	±30 m	Continuous	DEM, Google Map
Human related	distance from cities	±30 m	Continuous	DEM, Google Map

* IMO: Iran Meteorological Organization.

While the dataset size of 20 landslides + predisposing factors identified for studied region may be limited, the LR algorithm still provides meaningful insights. To overcome the challenge of a small dataset, a data augmentation technique was utilized. Data augmentation involves generating synthetic data points by applying transformations or perturbations to the existing dataset, effectively expanding the sample size [51,52]. This technique is helped mitigate the potential limitations arising from the small number of recorded landslides [53]. In this study, oversampling was considered for modeling. Oversampling addresses class imbalance between landslide occurrences/historical records (positive instances) and non-landslide areas (negative instances) in the dataset. The imbalance can lead to biased models that favor the majority class [54]. This allows logistic regression, a binary classification algorithm commonly used in landslide susceptibility assessment, to learn from a broader range of examples and improve its ability to predict landslide occurrences accurately. The oversampled dataset is used to train the logistic regression model, which

estimates the coefficients for independent variables and models the relationship between these variables and the binary outcome.
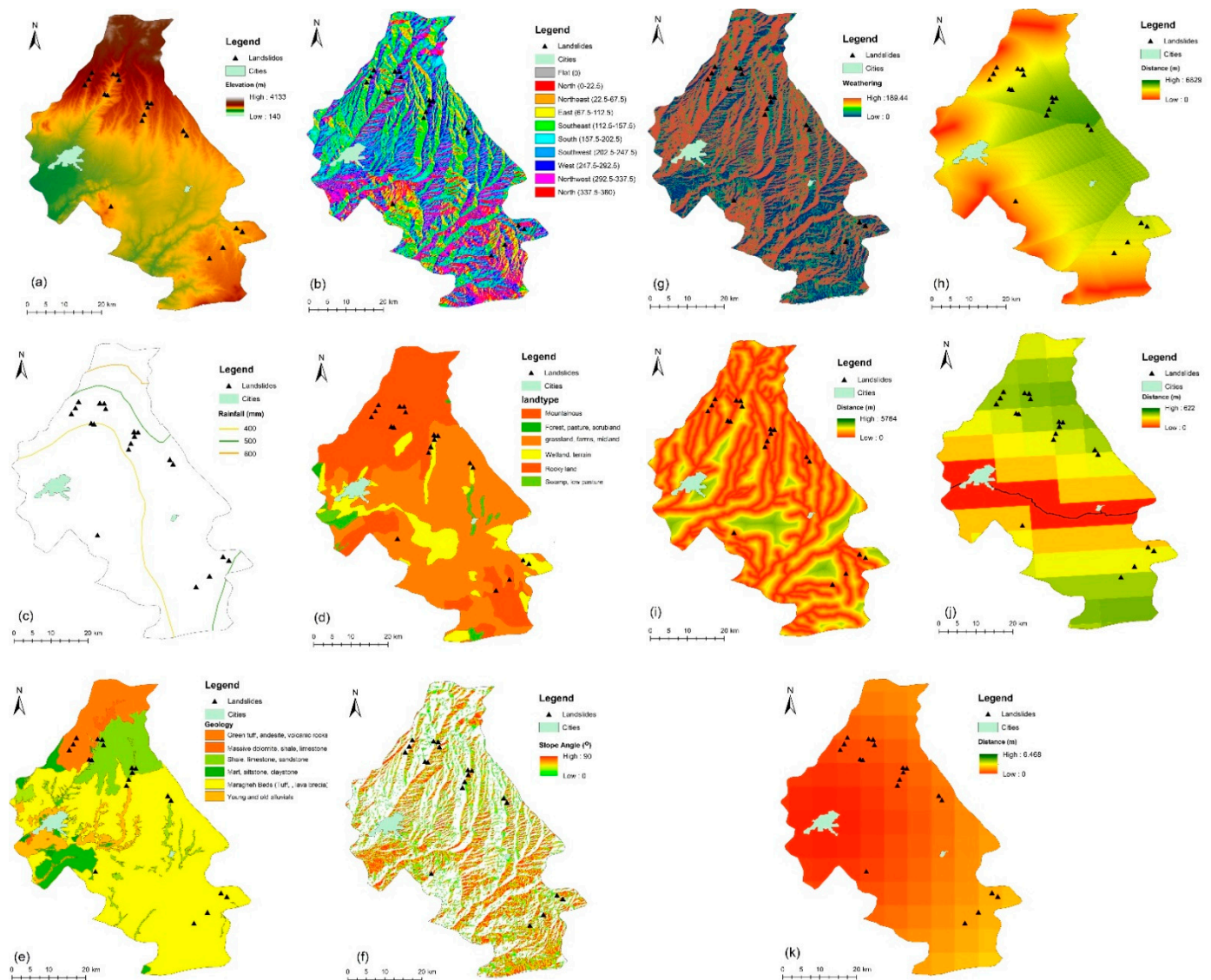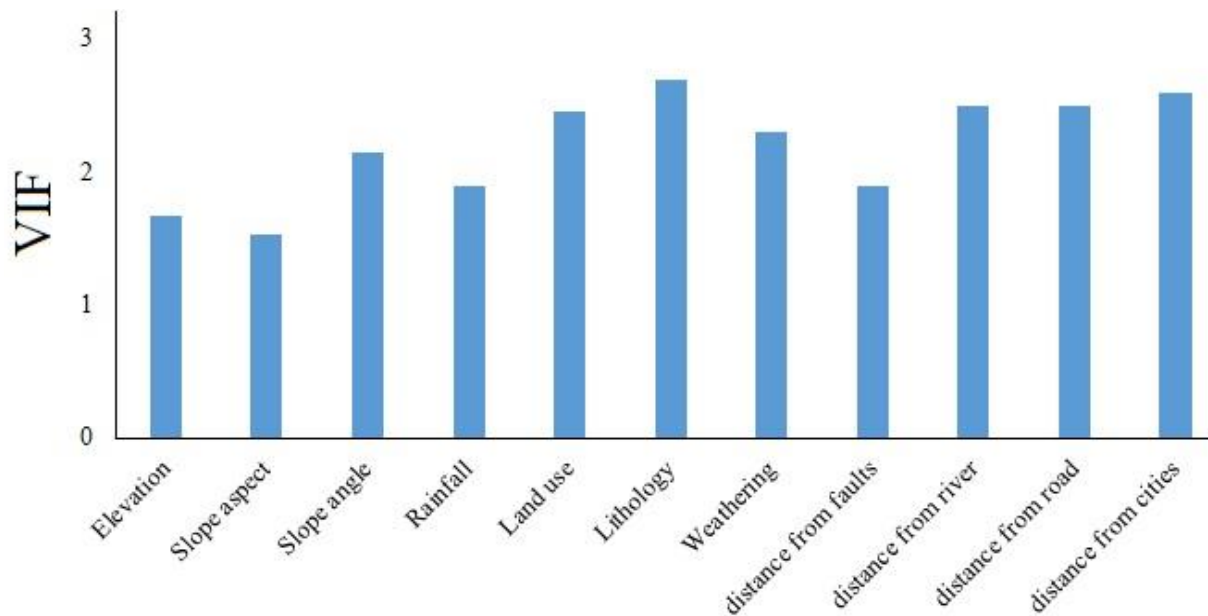

Figure 3. The main predisposing factors used in the analysis: (**a**) elevation, (**b**) slope aspect, (**c**) rainfall, (**d**) land use, (**e**) lithology, (**f**) slope angle, (**g**) weathering, (**h**) distance from faults, (**i**) distance from river, (**j**) distance from road, (**k**) distance from cities.

Regarding the spatial data quality, it is acknowledged that a ±30 m resolution may result in coarse-grained results. However, LR can still provide valuable insights even with low-resolution data by incorporating other relevant data sources, considering identified predisposing factors, and capturing meaningful patterns and relationships associated with landslide susceptibility. It should be noted that LR is advantageous due to its simplicity, interpretability, and ability to provide probabilistic outputs. It can offer valuable insights even with low-resolution data and is computationally efficient. LR can partially address the challenges posed by low-resolution data in landslide susceptibility assessment. Techniques such as feature engineering, regularization, ensemble methods, and feature selection can help overcome limitations. Feature engineering is considered to modify the LR model and improve predictive performance. During the feature engineering, relevant features from the database are selected and combined with approved features to capture the underlying patterns and relationships between the predisposing factors, recorded landslides, and non-

landslides. Although the study area may be small, logistic regression is applied effectively for localized analysis and proper preparation of the susceptibility map. It is essential to consider the specific context and scale of the study, ensuring that interpretations and conclusions are cautious and appropriately scaled to the study area.

Figure 4. The VIF versions for selected predisposing factors.

2.2. Principle of Logistic Regression

The main purpose of the LR method is to predict the probability of occurrence of a binary event from a set of variables that may be continuous, discrete or both in combination. The main difference between LR and other multiple statistical analyses is that the independent variables do not need to be normally distributed or linearly related, and the predicted values are converted to probabilities between 0 and 1. Therefore, many studies have used LR for evaluation [18]. LR allows to form a multivariate regression relationship between a dependent variable and several independent variables. LT is one of the multivariate analysis models that help predict the presence or absence of a characteristic or outcome based on the values of a set of predictor variables. Another advantage of logistic regression is that by adding an appropriate link function to the usual linear regression model, the variables may be continuous, discrete, or any combination of both and not necessarily normally distributed. In the case of multiple regression analysis, the factors must be numerical, and in the case of a similar statistical model, deterministic analysis, the variables must have a normal distribution [55]. The LR model is a generalization of the general linear model (GLM) that is compatible with the GLM and can also be compatible with the LR method, where its general model is as follows:

$$Y = \log \text{it}(p) = \ln\left(\frac{p}{1-p}\right) = c_0 + c_1 + c_2 + \cdots + c_{nx} \tag{1}$$

where Y is the probability of an accident (landslides), C_0 is the width from the origin or constant coefficient, $\left(\frac{p}{1-p}\right)$ is odds ratio, $c_1 + c_2 + \cdots + c_{nx}$ which are independent variables [56]. Mainly the p(x) is presented as logistic function and can be calculated as $p(x) = \frac{1}{1+e^{-[\frac{x-\mu}{s}]}}$ which μ is a location parameter (the midpoint of the LR curve), *s* is

a scale parameter, and $x_1 + x_2 + \cdots + x_n$ are coefficients of independent variables (c). This expression may be rewritten as [56]:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 - \beta_1 x)}} \qquad (2)$$

where β_0 (= $-\mu/s$) is the y-intercept, β_1 is the inverse scale parameter or rate parameter ($\beta_1 = 1/s$) which these are the y-intercept and slope of the log-odds as a function of x ($\mu = \beta_0/\beta_1$, and $s = 1/\beta_1$). So, after exponentiation of the log-odds functions, it can be stated $\frac{P(x)}{1 - P(x)} = e^{(\beta_0 - \beta_1 x)}$ [48].

The particular LR model various types of regressions used for binary-valued outcomes, which is the way the probability of a particular outcome is linked to the linear predictor function [48]:

$$\text{logit}\left(\varepsilon[Y_i \mid X_i]\right) = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta.X_i \qquad (3)$$

This formulation expresses logistic regression as a generalized linear model, which predicts variables with various types of probability distributions by fitting a linear predictor function of the above form to some sort of arbitrary transformation of the expected value of the variable. As the value of x ($x \to +\infty$) increases, the logistic function will reach 1.0. Also, by decreasing the value of x ($x \to -\infty$), the value of the function tends to zero. Suppose this function is used to express the probability of the dependent variable for LR.

Dai and Lee [57] stated that LR's main advantage compared to other multivariable statistical techniques is its capability to conduct multiple regression and diagnostic analysis on dependent variables, which can only have two values (one for occurrence probability and the other for non-occurrence). So, it can be useful to identify the variables based on the probability of landslide occurrence and non-occurrence in different regions.

2.3. Logistic Regression Verification

To evaluate the LR model's overall accuracy, mainly the operating relative characteristics (ROC) curves were considered in this article as well. Between floors in the logistic model, it is examined and confirmed, considering that the same slips that are used in zoning cannot be used to evaluate the model. ROC curve is one of the most efficient methods in providing the deterministic feature, probability identification and prediction of systems, which quantitatively estimates the accuracy of the machine learning-based predictive model. ROC curve is a graphical plot for binary classifiers' diagnostic ability to discriminate threshold is varied. The ROC curve is mainly created by plotting the true positive rate against the false positive rate at various threshold settings, known as evaluation criteria in the confusion matrix [58]. In ROC, the area under the curve (AUC) indicates the overall accuracy of predictive values by describing its ability to correctly estimation of events that have occurred (landslides) and not (non-occurrence of landslides). So, the AUC will vary from 0 to 1. So that the closer the numerical value of AUC is to 1, the higher the overall accuracy and the closer to 0, which indicates low accuracy and thus more error [59].

2.4. Comparative Justifications

As justification for the checking of predictive model performance, several benchmark learning classifiers are considered. Naïve Bayes (NB), Decision Trees (DT), Random Forests (RF), k-Nearest Neighbors (k-NN), and Liner Support Vector Machines (SVM) are the main concern for comparative justification in this study. NB is an effective probabilistic classifier that is based on Bayes' theorem. Given the class label, it assumes that features are conditionally independent, hence the 'naïve' assumption. NB works well with high-dimensional data and is computationally efficient. It has been widely used in text classification tasks and spam filtering due to its simplicity and decent performance. However, the naïve assumption may not hold in all scenarios, which can limit its accuracy compared to more complex models. DT is a popular class of classifiers that uses a hierarchical structure of

nodes to make decisions based on feature values. They recursively split the data into subsets based on the most informative features, creating a tree-like structure. DT is easy to interpret and visualize, making them helpful in understanding the decision-making process. However, they can be prone to overfitting if not properly pruned or regularized. Ensemble methods like RF, which combine multiple decision trees, can mitigate overfitting, and improve predictive accuracy by aggregating the decisions of individual trees within DT [58,59].

The k-NN is a non-parametric classifier that assigns labels to new instances based on the majority vote of their k-nearest neighbors in the feature space. It is simple and intuitive but computationally expensive for large datasets. k-NN is sensitive to the choice of k and the distance metric used for similarity calculation. It performs well when there are clear clusters in the data or when the decision boundary is nonlinear. Linear SVM aims to find an optimal hyperplane that separates the classes with the maximum margin. SVMs are particularly effective in high-dimensional spaces and can handle linear and non-linear classification tasks using kernel functions. They are robust to outliers and have a solid theoretical foundation. However, SVMs can be computationally intensive for large datasets and may struggle with datasets that have overlapping classes or complex decision boundaries [58,59].

In the context of landslide susceptibility assessment, these models offer several advantages. NB is computationally efficient and can handle high-dimensional data, making it suitable for datasets often encountered in landslide studies. DT provide interpretability, allowing insights into decision-making and identifying essential features. RF, as an ensemble of decision trees, improve generalization and capture complex interactions. k-NN can capture non-linear patterns and spatial clustering relevant to landslide analysis. SVM excel in handling high-dimensional data, dealing with overlapping classes, and capturing complex decision boundaries [60,61]. These models provide different strengths and capabilities, and the choice depends on factors such as data characteristics, interpretability needs, and the complexity of relationships in the specific landslide susceptibility assessment. It is essential to experiment and evaluate the performance of different models to select the most suitable one for a given scenario.

In the presented study, a comparative justification was undertaken by evaluating the mentioned classifiers in conjunction with LR. The study utilized the same input variables and requested outputs for all predictive models to ensure consistency. The performance evaluation was primarily controlled by the ROC curve analysis. The AUC was utilized as a metric to assess the overall accuracy of the models. The study aimed to provide a comparative analysis of the classifiers' predictive performance for the specific landslide susceptibility assessment by considering the AUC and overall accuracy.

2.5. Model Implementation

Since the LR model can examine and predict the impact of several predisposing factors on landslide susceptibility, using LR is beneficial in predicting the probability of landslide occurrence. The studied region is known as Maragheh County, located northwest of Iran. Both ground survey and remote-sensing analysis were used in the studied region to extract the proper features to identify the main predisposing factors. The assessment was improved by recording 20 historical landslides from the County, which shows the hazardous areas in the region. The information was prepared, classified, and used as input parameters in the LR model. Table 1 provides information about resources of the predisposing factors accumulations. The input parameters that gathers as an inventory dataset of landslide predisposing factors were randomly divided into two testing and training sets responsible for LR model learning and validating. The LR modeling was implemented in Python programming language. The training set is considered 70% of the main dataset, and other remaining 30% is selected for testing purposes. The model results are used as information layers to prepare a susceptibility map for landslides in the studied region. Figure 5 provides the process flowchart of model implication.

```
┌─────────────────────────┐
│ Providing the Predisposing│
│          factors          │
└─────────────────────────┘
            │
            ▼
    ┌─────────────────┐
    │ Landslide records│
    │   integration    │
    └─────────────────┘
            │
            ▼
    ┌─────────────────┐
    │ Preparation of database│
    └─────────────────┘
       │           │
       ▼           ▼
┌──────────────┐  ┌──────────────┐
│ Training dataset│  │  Test dataset  │
└──────────────┘  └──────────────┘
       │           │
       ▼           ▼
    ┌─────────────────┐
    │ Model verification and│
    │     validation    │
    └─────────────────┘
            │
            ▼
    ┌─────────────────┐
    │ GIS-based information│
    │       layer       │
    └─────────────────┘
            │
            ▼
    ┌─────────────────┐
    │ Landslide susceptibility│
    │        map        │
    └─────────────────┘
```
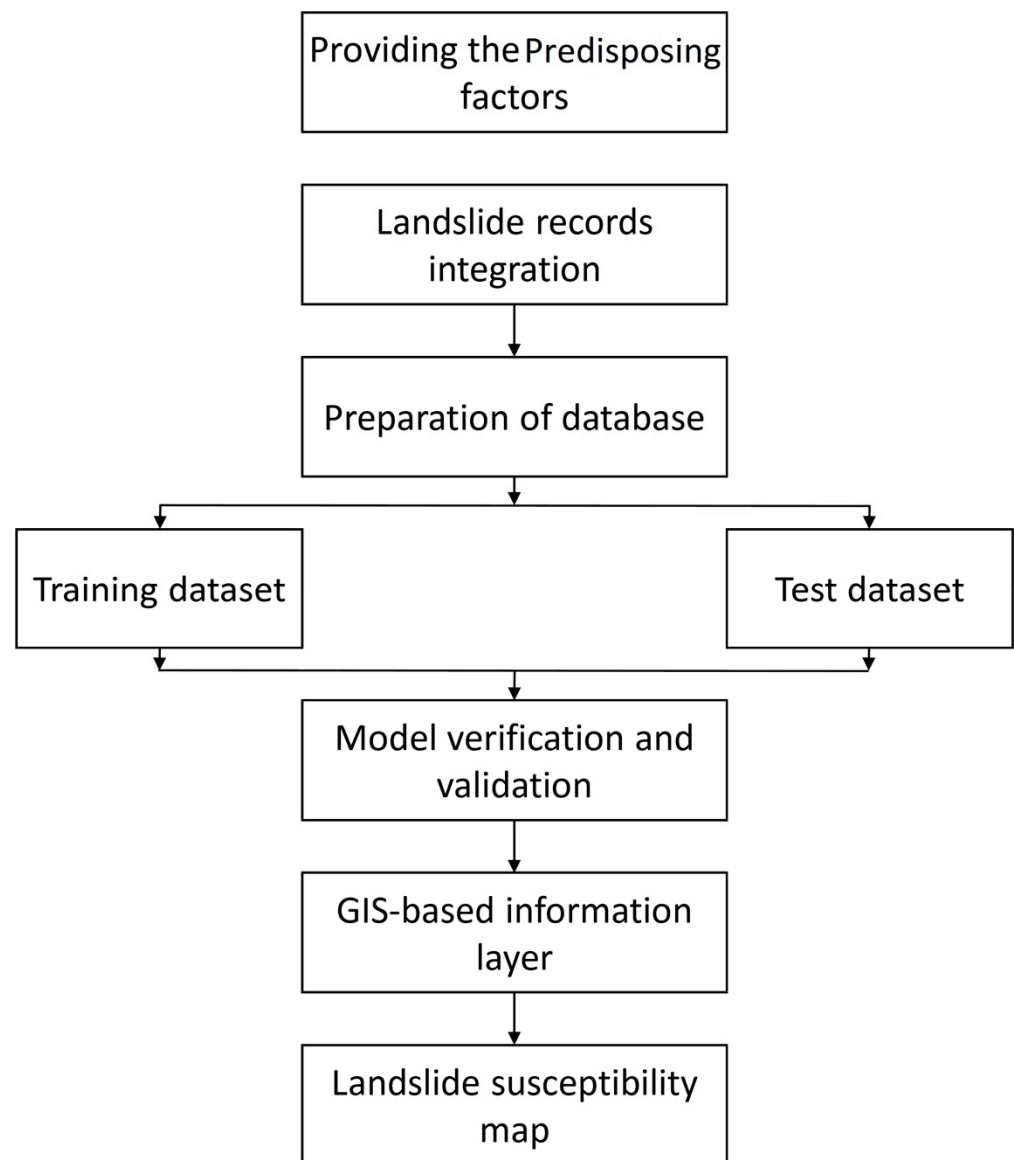
Figure 5. The process flowchart of the LR-based landslide susceptibility assessment.

3. Results

Referring to various studies on landslide susceptibility assessment, it can appear that there is no global framework for susceptibility analysis, and professionals are looking for more accurate procedures in this matter. The presented study used the LR predictive model regarding the high performance of the model to predict the probability of events. The model was applied to several selected predisposing factors (one of the most important steps in spatial prediction modelling) and 20 historical landslide records. The final susceptibility analysis maps are an effective tool for managing future landslide occurrences in Maragheh County. Figure 6 is presented the landslide susceptibility map using the LR method. Firstly, the entire study region was converted to an information layer as pixels using GIS software version 10.4. Next, all predisposing factors that entered the model were classified based on the natural breaks classification scheme [30].
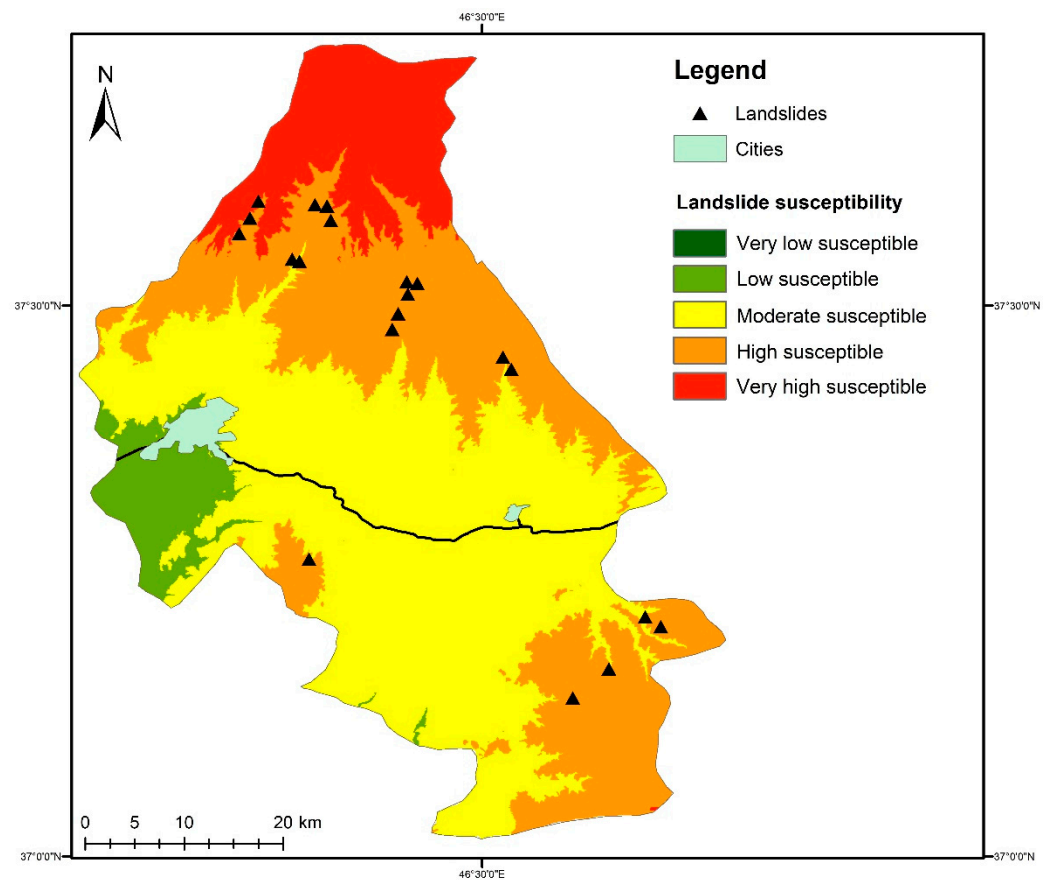
Figure 6. Landslide susceptibility map using the logistic regression model.

Finally, these factors merged and classified into five susceptibility classes: very low, low, moderate, high, and very high. As seen in Figure 6, it can be stated that the main focus of the landside hazards belongs to the north and northeast part of the studied region, showing the high and very high potential regarding landslide occurrence. From a general perspective, the primary susceptibility condition of Maragheh County is moderate hazard risk suitability. The second-ranked level of susceptibility regarding high potential which is covered the north and southeast of the region. Also, Figure 7 is presented the percent of various landslide susceptibility classes, which is estimated by the model.
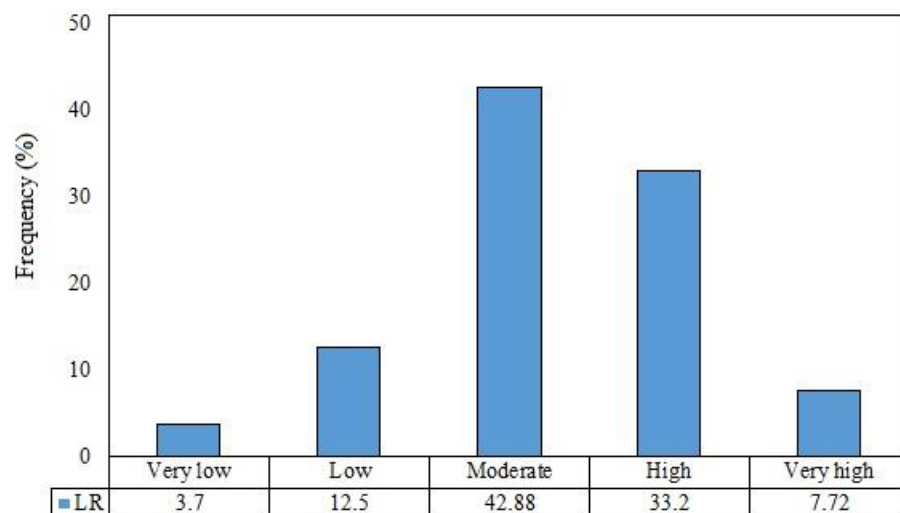
Figure 7. The distribution of various susceptibility classes for a landslide.

Regarding this figure, the region's moderate susceptibility class ranks at 42.88%, followed by high-risk potential at 33.2%. This variation indicates that Maragheh County has moderate to high hazard risk levels regarding ground movements and the probability of landslides. So, in sustainable development and urban management, landslide risk management has to be considered properly.

After providing the landslide susceptibility map for Maragheh County, the applied model must be validated and controlled regarding performance and capability. The results of the machine learning–based predictive models are commonly validated based on ROC curve analysis. Figure 8 illustrates the results of ROC curve analysis for LR models, which were implemented for testing and training datasets. The AUC result from ROC in the training set shows that the LR has good and reliable prediction capability with considerable accuracy and performance. The AUC estimated for the training set is 0.885, and the AUC estimated for the testing set is 0.769. A ROC provides the overall accuracy based on true positive and false positive rates. The rate is responsible for the model performance, which is reasonable and high in this case. Thus, the model is recommended for landslide susceptibility assessment in the studied region.

Regarding LR application in landslide susceptibility analysis which is considered one of the flexible machine learning-based classification models utilized for Maragheh County susceptibility assessment. In this regard, based on conducting field surveys and remote-sensing, several predisposing factors included elevation, slope aspect, slope angle, rainfall, land use, lithology, weathering, distance from faults, distance from rivers, distance from roads, and distance from cities was determinate for the studied region. In the meantime, during ground investigation, 20 historical landslide events are recorded and verified by satellite images to correct the prone areas of land sliding. These predisposing factors represent the morphologic, climatologic, geologic, and human works which is the primary classification for influencing elements in the studied area. The data on predisposing factors and the location of the historical landslide is used to provide the main dataset, which randomly divides into the testing and training sets. The training set is 70% of the main dataset, and the remaining 30% is selected for testing purposes. LR method was implemented to prepare landslide susceptibility assessment for Maragheh County, and the results were converted into the GIS environment. To evaluate the performance of the LR model, ROC curves and AUC are considered for both testing and training sets to calculate the overall accuracy of the modeling. The results confirmed that LR is a reliable method for landslide susceptibility mapping. As per the susceptibility assessment of landslide results for the Maragheh region, it appears the north and northeast parts are located the high and very high susceptible area and the general view of the County demonstrate the moderate to high classes for land sliding, which have to be considered in urban planning and sustainable developments.

It is important to note that the validation process, particularly using AUC, measures the modelling results and does not directly indicate the quality of the susceptibility map. A high AUC value does not guarantee that the model's results are reasonable or realistic in terms of identifying susceptible areas, but a high AUC value is indicating the model is implemented and operates properly. If the AUC is high, it can be concluded the model operates and act properly and functionally. Also, the relationship between AUC and the results of a landslide susceptibility map is indirect and requires careful interpretation. AUC is commonly used as a metric to evaluate the performance of a predictive model, including models used in landslide susceptibility analysis. It measures the model's ability to distinguish between positive and negative instances, where higher AUC values generally indicate better discrimination between the two classes. This study related to the 'landslide' and 'non-landslide' classes. However, it is important to note that AUC alone does not directly represent the quality or reliability of the susceptibility map itself. A high AUC value implies that the model can differentiate between areas with different susceptibilities. Nonetheless, the map's accuracy and validity depend on several factors, including the

quality and representativeness of the training data, the choice of input variables, and the model's assumptions and limitations.
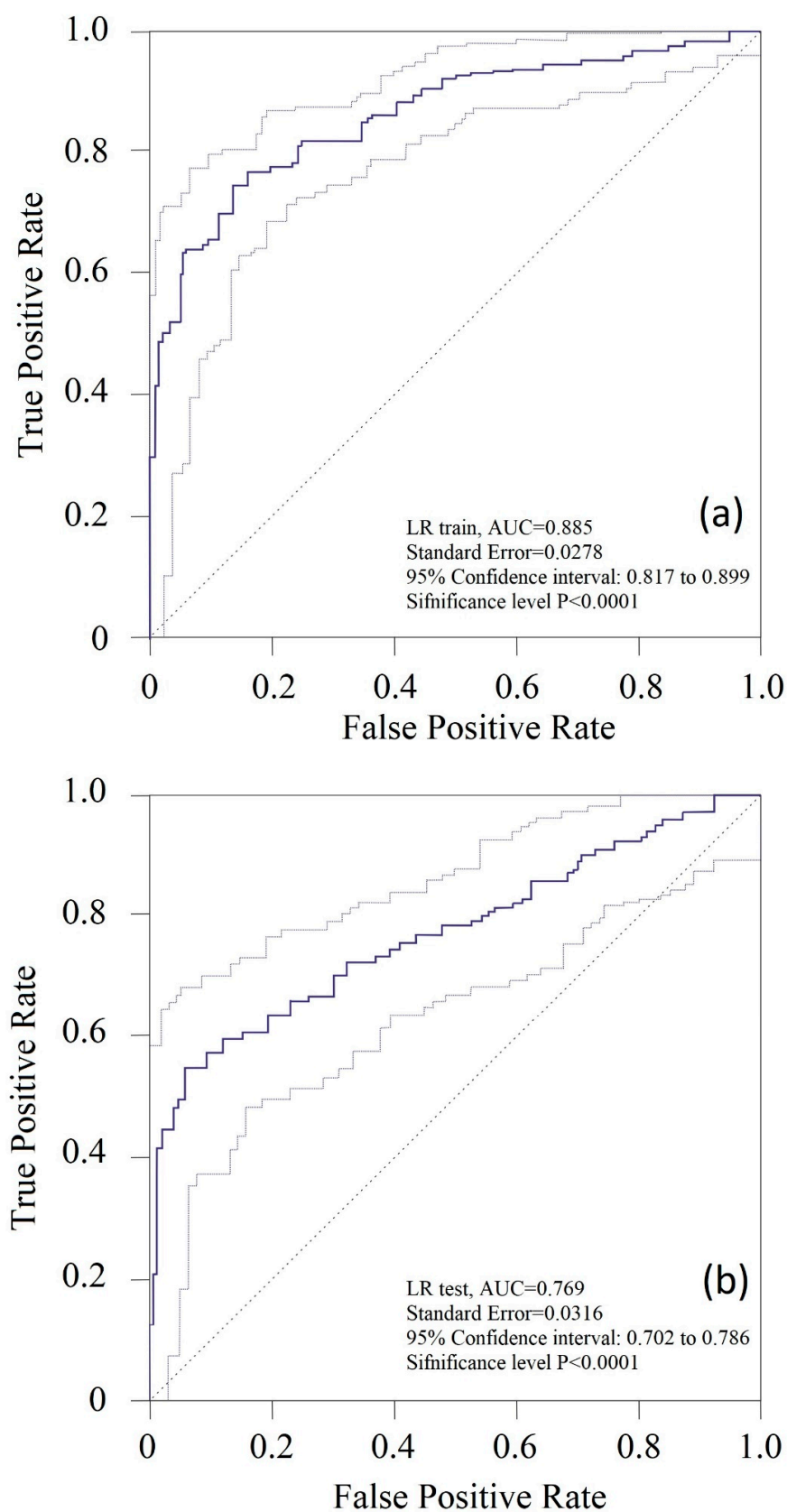

Figure 8. ROC curves are estimated for (**a**) training and (**b**) testing datasets.

To establish a stronger relationship between AUC and the susceptibility map results, it is necessary to consider additional factors. These may include conducting field validation to compare the map's predictions with observed landslide occurrences and incorporating expert knowledge to evaluate the map's consistency with known landslide-prone areas. In this research, both methods were utilized to verify the model capability.

Firstly, by comparing the high and very high susceptible areas with the topographical data and recorded occurred landslides, it has been determined that these areas have a good overlap. Also, using the expert check, the model has been checked concerning geological, topographical, and other predisposing factors. The results are indicated the model has good agreements with real data.

In this article, a comparative verification was conducted to justify the performance of different predictive models. The evaluation involved a ROC curve analysis, and Table 2 presents the results, specifically focusing on the AUC curve values. Based on the comparative analysis, it is evident that LR achieved satisfactory overall accuracy and quality in predicting the models. This observation serves as a justification for the effectiveness of LR as a viable approach in the predictive modeling of the specific context being studied.

Table 2. A comparative model justification results based on the ROC analysis curve.

Model	AUC	Standard Error	Reliability	Expert Opinion
SVM	0.862	0.0291	Reliable	Reliable
NB	0.655	0.0445	Need attention	Need attention
DT	0.591	0.0639	Need attention	Need attention
RF	0.730	0.0308	Reliable	Reliable
LR	0.885	0.0278	Reliable	Reliable

From the evaluation results, it can seem that the spatial distribution of the recorded historical landslides is located in high susceptibility level, and a very-high susceptibility level is not showing such consistency. This may be why the low data quality and model used logic learning to cover the input-database limitations using feature engineering. Feature engineering involves extracting relevant features or creating new ones based on domain knowledge to enhance the model's predictive power. In this task, we applied this technique to mitigate the limitations and overcome to region database scale.

4. Discussion

The application of LR in landslide susceptibility analysis for Maragheh County is highlighted as an effective approach. The study incorporates field surveys and remote sensing to identify several predisposing factors that contribute to landslides in the region. These factors include elevation, slope aspect, slope angle, rainfall, land use, lithology, weathering, distance from faults, distance from rivers, distance from roads, and distance from cities. These factors represent the morphologic, climatologic, geologic, and human elements influencing landslide occurrences. To create the dataset for analysis, historical landslide events (20 in total) are recorded and verified using satellite images. These events serve as a reference to correct the susceptible areas prone to land sliding. The dataset includes information about the predisposing factors and the location of historical landslides. It is then randomly divided into training and testing sets, with the training set comprising 70% of the dataset and the remaining 30% used for testing purposes.

LR is implemented as the chosen method to develop the landslide susceptibility assessment for Maragheh County. The results are then converted into a GIS environment, allowing for visualization and spatial analysis. The performance of the LR model is evaluated using ROC curves and AUC for both the testing and training sets, providing an overall measure of accuracy for the modeling process. The outcomes of the analysis confirm the reliability of the LR method for landslide susceptibility mapping. The assessment reveals that the northern and northeastern parts of Maragheh County are classified as highly susceptible areas. The general view of the County indicates moderate to high

susceptibility classes for landslides. These findings have important implications for the region's urban planning and sustainable development, emphasizing the need to consider landslide susceptibility in decision-making processes.

The application of LR in landslide susceptibility analysis for Maragheh County demonstrates its flexibility as a machine learning-based classification model. By considering multiple predisposing factors and incorporating historical landslide data, the study provides a reliable assessment of landslide susceptibility. The results highlight specific high-risk areas and emphasize the importance of considering landslide susceptibility in planning and development initiatives. Mainly, LR offers several advantages when applied to landslide susceptibility analysis in Maragheh County. One of the key benefits of using LR is the interpretability of its results. The coefficients assigned to each input factor in LR provide a clear understanding of their influence on landslide susceptibility. This transparency allows stakeholders and decision-makers to comprehend the relative importance of various predisposing factors in the study area.

Additionally, LR is well-suited for binary classification tasks, making it suitable for determining whether an area is susceptible or not susceptible to landslides. This ability to model the relationship between input factors and binary outcomes facilitates the identification of the landslide-prone regions within Maragheh County. The flexibility of LR is another advantage. It can handle both continuous and categorical input variables, including diverse predisposing factors such as elevation, slope angle, land use, and distance from geological features. These factors play a crucial role in assessing landslide susceptibility in the region. Statistical significance assessment is also possible with LR, enabling the identification of input variables that significantly impact landslide susceptibility. This information helps prioritize mitigation efforts and allocate resources accordingly.

LR is applicable even when working with small datasets, which is the case in this analysis with only 20 historical landslide events. Despite the limited data, LR can still provide meaningful insights and predictions for susceptibility mapping. Techniques like data augmentation and feature engineering can enhance the model's performance in such scenarios [52,53]. Furthermore, LR results can seamlessly integrate into a Geographic Information System (GIS) environment. This integration facilitates the visualization and spatial analysis of landslide susceptibility patterns in Maragheh County. Such spatial information aids decision-makers in urban planning and sustainable development initiatives by clearly understanding the areas at higher risk. Its interpretability, binary classification capabilities, flexibility, applicability to small datasets, statistical significance assessment, and compatibility with GIS contribute to a comprehensive understanding of landslide susceptibility patterns. This knowledge supports effective mitigation strategies and informs sustainable development practices in the region.

5. Conclusions

This study aimed to conduct a comprehensive landslide susceptibility analysis for Maragheh County in the East Azerbaijan province, located in the northwest of Iran. Through extensive field surveys and remote sensing observations, key predisposing factors were identified and extracted for the studied region. These factors and 20 records of historical landslides served as the fundamental data for the Logistic Regression (LR) model utilized to prepare landslide susceptibility maps. The LR model was evaluated using a rigorous validation process, employing ROC curves, and calculating the Area Under the Curve (AUC) for the testing and training sets. The obtained AUC values were 0.885 for the training set and 0.769 for the testing set, indicating high accuracy in the model's predictions. These results validate the reliability and effectiveness of the LR model in assessing landslide susceptibility for the region. The generated landslide susceptibility maps provided valuable insights into the distribution of landslide-prone areas within Maragheh County.

The obtained susceptibility maps provided valuable insights into the county's spatial distribution of landslide-prone areas. The majority of the region, covering approximately 42.88% of the total area, exhibited moderate landslide suitability. This finding emphasizes

the importance of incorporating landslide mitigation and management strategies into various regional planning initiatives to minimize potential risks. Moreover, the northern part of the county demonstrated a significantly higher susceptibility to landslides, particularly in the mountain foothill areas. This information is crucial for land use planning and sustainable development initiatives, highlighting the importance of implementing appropriate measures to minimize landslides' risks and potential impacts in these vulnerable regions. Overall, this study contributes to a better understanding of landslide susceptibility in Maragheh County by utilizing the LR model and integrating field surveys, remote sensing, and historical landslide data.

Author Contributions: A.C., L.Z. and A.B.M.: conceptualization, methodology, software, validation, formal analysis, investigation, resources, writing—original draft preparation, Y.A.N. and M.A.: writing—review and editing, visualization, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Nature Sciences Foundation of China (Grant No. 42250410321).

Data Availability Statement: All required data is within the article.

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

References

1. Lombardo, L.; Mai, P.M. Presenting logistic regression-based landslide susceptibility results. *Eng. Geol.* **2018**, *244*, 14–24. [CrossRef]
2. Aditian, A.; Kubota, T.; Shinohara, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* **2018**, *318*, 101–111. [CrossRef]
3. Sevgen, E.; Kocaman, S.; Nefeslioglu, H.A.; Gokceoglu, C. A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random forest. *Sensors* **2019**, *19*, 3940. [CrossRef]
4. Nhu, V.H.; Shirzadi, A.; Shahabi, H.; Singh, S.K.; Al-Ansari, N.; Clague, J.J.; Ahmad, B.B. Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2749. [CrossRef] [PubMed]
5. Carrara, A. Multivariate methods for landslide hazard evaluation. *Math. Geol.* **1983**, *15*, 403–426. [CrossRef]
6. Abeysiriwardana, H.D.; Gomes, P.I. Integrating vegetation indices and geo-environmental factors in GIS-based landslide-susceptibility mapping: Using logistic regression. *J. Mount. Sci.* **2022**, *19*, 477–492. [CrossRef]
7. Oehorst, B.A.N.; Kjekstad, O.; Patel, D.; Lubkowski, Z.; Knoeff, J.G.; Akkerman, G.J. Workpackage, Determination of Socio-Economic Impact of Natural Disasters. *J. Assess. Socio-Econ. Impact Eur.* **2005**, *173*, 1–14.
8. Guzzetti, F.; Carrarra, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [CrossRef]
9. Azarafza, M.; Azarafza, M.; Akgün, H.; Atkinson, P.M.; Derakhshani, R. Deep learning-based landslide susceptibility mapping. *Sci. Rep.* **2021**, *11*, 24112. [CrossRef]
10. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [CrossRef]
11. Lee, S.; Pradhan, B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* **2007**, *4*, 33–41. [CrossRef]
12. Budimir, M.E.A.; Atkinson, P.M.; Lewis, H.G. A systematic review of landslide probability mapping using logistic regression. *Landslides* **2015**, *12*, 419–436. [CrossRef]
13. Nanehkaran, Y.A.; Mao, Y.; Azarafza, M.; Kockar, M.K.; Zhu, H.H. Fuzzy-based multiple decision method for landslide susceptibility and hazard assessment: A case study of Tabriz, Iran. *Geomech. Eng.* **2021**, *24*, 407–418.
14. Nikoobakht, S.; Azarafza, M.; Akgün, H.; Derakhshani, R. Landslide Susceptibility Assessment by Using Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 5992. [CrossRef]
15. Nefeslioglu, H.A.; Duman, T.Y.; Durmaz, S. Landslide susceptibility mapping for a part of tectonic Kelkit Valley (Easten Black Sea Region of Turkey). *Geomorphology* **2008**, *94*, 401–418. [CrossRef]
16. Bai, S.B.; Wang, J.; Lü, G.N.; Zhou, P.G.; Hou, S.S.; Xu, S.N. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology* **2010**, *115*, 23–31. [CrossRef]
17. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378.

18. Bui, D.T.; Lofman, O.; Revhaug, I.; Dick, O. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat. Hazards* **2011**, *59*, 1413–1444. [CrossRef]

19. Ebrahimi, Z. Investigating the causes of landslides in the west of Mazandaran province. In Proceedings of the First International Congress of Earth, Space and Clean Energy, Ardabil, Iran, 5 November 2015.

20. Raja, N.B.; Çiçek, I.; Türkoğlu, N.; Aydin, O.; Kawasaki, A. Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Nat. Hazards* **2017**, *85*, 1323–1346. [CrossRef]

21. Lin, L.; Lin, Q.; Wang, Y. Landslide susceptibility mapping on a global scale using the method of logistic regression. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 1411–1424. [CrossRef]

22. Leoni, G.; Barchiesi, F.; Catallo, F.; Dramis, F.; Fubelli, G.; Lucifora, S.; Mattei, M.; Pezzo, G.; Puglisi, C. GIS methodology to assess landslide susceptibility: Application to a river catchment of Central Italy. *J. Maps* **2009**, *5*, 87–93. [CrossRef]

23. Ercanoğlu, M.; Gökçeoğlu, C. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). *Eng. Geol.* **2004**, *75*, 229–250. [CrossRef]

24. Othman, A.N.; Naim, W.M.; Noraini, S. GIS based multi-criteria decision making for landslide hazard zonation. *Proc. Soc. Behav. Sci.* **2012**, *35*, 595–602. [CrossRef]

25. Vahidnia, M.H.; Alesheikh, A.A.; Alimohammadi, A.; Hosseinali, F. A GIS-based neurofuzzy procedure for integrating knowledge and data in landslide susceptibility mapping. *Comput. Geosci.* **2010**, *36*, 1101–1114. [CrossRef]

26. Okalp, K.; Akgün, H. National level landslide susceptibility assessment of Turkey utilizing public domain dataset. *Environ. Earth Sci.* **2016**, *75*, 847. [CrossRef]

27. Zêzere, J.L.; Pereira, S.; Melo, R.; Oliveira, S.C.; Garcia, R.A.C. Mapping landslide susceptibility using data-driven methods. *Sci. Total Environ.* **2017**, *589*, 250–267. [CrossRef]

28. Azarafza, M.; Ghazifard, A.; Akgün, H.; Asghari-Kaljahi, E. Landslide susceptibility assessment of South Pars Special Zone, southwest Iran. *Environ. Earth Sci.* **2018**, *77*, 805. [CrossRef]

29. Trigila, A.; Iadanza, C.; Esposito, C.; Scarascia-Mugnozza, G. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology* **2015**, *249*, 119–136. [CrossRef]

30. Chen, W.; Shahabi, H.; Zhang, S.; Khosravi, K.; Shirzadi, A.; Chapi, K.; Pham, B.T.; Zhang, T.; Zhang, L.; Chai, H.; et al. Landslide Susceptibility Modeling Based on GIS and Novel Bagging-Based Kernel Logistic Regression. *Appl. Sci.* **2018**, *8*, 2540. [CrossRef]

31. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Pham, B.T.; Bui, D.T. A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment. *Geocarto Int.* **2019**, *34*, 1427–1457. [CrossRef]

32. Sujatha, E.R.; Sridhar, V. Landslide susceptibility analysis: A logistic regression model case study in Coonoor, India. *Hydrology* **2021**, *8*, 41. [CrossRef]

33. Shan, Y.; Chen, S.; Zhong, Q. Rapid prediction of landslide dam stability using the logistic regression method. *Landslides* **2020**, *17*, 2931–2956. [CrossRef]

34. Hosmer, D.W.; Lemeshow, J.S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013.

35. Lee, S.; Choi, J.; Min, K. Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. *Int. J. Rem. Sens.* **2004**, *25*, 2037–2052. [CrossRef]

36. Ayalew, L.; Ymagishi, H.; Marui, H.; Kanno, T. GIS-based susceptibility mapping with comparisons of result from methods and verifications. *J. Eng. Geol.* **2005**, *81*, 432–445. [CrossRef]

37. Lee, S.; Sambath, T. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *J. Environ. Geol.* **2006**, *50*, 847–855. [CrossRef]

38. Yalcin, A. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena* **2011**, *85*, 274–287. [CrossRef]

39. Ozdemir, A. GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *J. Hydrol.* **2011**, *411*, 290–308. [CrossRef]

40. Shahabi, H.; Hashim, M.; Ahmad, B.B. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central Zab basin, Iran. *Environ. Earth Sci.* **2015**, *73*, 8647–8668. [CrossRef]

41. Tekin, S. Completeness of landslide inventory and landslide susceptibility mapping using logistic regression method in Ceyhan Watershed (southern Turkey). *Arab. J. Geosci.* **2021**, *14*, 1706. [CrossRef]

42. Nwazelibe, V.E.; Unigwe, C.O.; Egbueri, J.C. Integration and comparison of algorithmic weight of evidence and logistic regression in landslide susceptibility mapping of the Orumba North erosion-prone region, Nigeria. *Model. Earth Syst. Environ.* **2023**, *9*, 967–986. [CrossRef]

43. Gu, T.; Li, J.; Wang, M.; Duan, P. Landslide susceptibility assessment in Zhenxiong County of China based on geographically weighted logistic regression model. *Geocarto Int.* **2022**, *37*, 4952–4973. [CrossRef]

44. Azarafza, M.; Mokhtari, M.H. Evaluation of drought effect on Urmia Lake salinity changes using remote sensing techniques. *Arid Biome Sci. Res. J.* **2013**, *3*, 1–14.

45. Iran Meteorological Organization (IMO). *Climatological Data from Maragheh Station*; The Iran Meteorological Organization: Tehran, Iran, 2022; Available online: http://www.irimo.ir/ (accessed on 5 October 2022).

46. Geological Survey of Iran (GSI). *Geological Map of Maragheh City in Scale: 1:100000 (Map Sheet)*; Geological Survey of Iran Press, Maps Unit: Tehran, Iran, 2009.

47. 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]

48. Fox, J. *An R Companion to Applied Regression*; SAGE Publications: Thousand Oaks, CA, USA, 2011.

49. Adnan, M.S.G.; Rahman, M.S.; Ahmed, N.; Ahmed, B.; Rabbi, M.F.; Rahman, R.M. Improving spatial agreement in machine learning-based landslide susceptibility mapping. *Rem. Sens.* **2020**, *12*, 3347. [CrossRef]

50. Huang, F.; Teng, Z.; Guo, Z.; Catani, F.; Huang, J. Uncertainties of landslide susceptibility prediction: Influences of different spatial resolutions, machine learning models and proportions of training and testing dataset. *Rock Mech. Bull.* **2023**, *2*, 100028. [CrossRef]

51. Discacciati, A.; Orsini, N.; Greenland, S. Approximate Bayesian logistic regression via penalized likelihood by data augmentation. *Stata J.* **2015**, *15*, 712–736. [CrossRef]

52. O'brien, S.M.; Dunson, D.B. Bayesian multivariate logistic regression. *Biometrics* **2004**, *60*, 739–746. [CrossRef]

53. 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. *Rem. Sens.* **2019**, *11*, 196. [CrossRef]

54. Gao, H.; Fam, P.S.; Tay, L.T.; Low, H.C. Three oversampling methods applied in a comparative landslide spatial research in Penang Island, Malaysia. *SN Appl. Sci.* **2020**, *2*, 1512. [CrossRef]

55. Soma, A.S.; Kubota, T.; Mizuno, H. Optimization of causative factors using logistic regression and artificial neural network models for landslide susceptibility assessment in Ujung Loe Watershed, South Sulawesi Indonesia. *J. Mount. Sci.* **2019**, *16*, 383–401. [CrossRef]

56. Feby, B.; Achu, A.L.; Jimnisha, K.; Ayisha, V.A.; Reghunath, P. Landslide susceptibility modelling using integrated evidential belief function based logistic regression method: A study from Southern Western Ghats, India. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100411. [CrossRef]

57. Dai, F.C.; Lee, C.F. Landslide characteristics and slope instability modeling using GIS, Lantau, Hong Kong. *Geomorphology* **2002**, *42*, 213–228. [CrossRef]

58. Aggarwal, C.C. *Neural Networks and Deep Learning: A Textbook*; Springer: Berlin/Heidelberg, Germany, 2018.

59. Müller, A.C.; Guido, S. *Introduction to Machine Learning with Python: A Guide for Data Scientists*; O'Reilly Media: Sevastopol, CA, USA, 2016.

60. Mondal, B.; Koner, C.; Chakraborty, M.; Gupta, S. Detection and investigation of DDoS attacks in network traffic using machine learning algorithms. *Int. J. Innov. Technol. Explor. Eng.* **2022**, *11*, 1–6. [CrossRef]

61. Luo, X.; Lin, F.; Chen, Y.; Zhu, S.; Xu, Z.; Huo, Z.; Peng, J. Coupling logistic model tree and random subspace to predict the landslide susceptibility areas with considering the uncertainty of environmental features. *Sci. Rep.* **2019**, *9*, 15369. [CrossRef] [PubMed]