International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.2 (2025) pp. 3255-3270
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Production of Highway Landslide Susceptibility Map with Machine Learning Techniques: A Local Study from Türkiye, Artvin-Ardanuç Road Line

Selim Taşkaya[1*], Oktay Aksu[2], Samet Doğan[3], Mustafa Kurt[4]

[1] (Land Management and Use Doctoral Program, Department of Geomatics Engineering, Graduate Education Institute, Istanbul Okan University, İstanbul 34959, Türkiye)
Land Registry and Cadastre Program, Department of Architecture and Urban Planning, Artvin Coruh University, Artvin 08100, Türkiye
*** Corresponding Author Email:** selim_taskaya@artvin.edu.tr- **ORCID:** 0000-0002-4290-3684

[2] Department of Geomatics Engineering, Graduate Education Institute, Istanbul Okan University, İstanbul 34959, Türkiye
Email: oktay.aksu@okan.edu.tr - **ORCID:** 0009-0001-5584-6079

[3] (Land Management and Use Doctoral Program, Department of Geomatics Engineering, Graduate Education Institute, Istanbul Okan University, İstanbul 34959, Türkiye)
Land Map and Cadastre Program, Department of Architecture and Urban Planning, Kastamonu University, Kastamonu 37250, Türkiye
Email: sametdogan@kastamonu.edu.tr – (samedogan@stu.okan.edu.tr) **ORCID:** 0000-0002-3103-1593

[4] Department of Geomatics Engineering, Graduate Education Institute, Istanbul Okan University, İstanbul 34959, Türkiye
Email: mustafa.kurt@okan.edu.tr - **ORCID:** 0000-0001-8470-2832

Article Info:

DOI: 10.22399/ijcesen.1792
Received : 21 February 2025
Accepted : 09 May 2025

Keywords :

Landslide
Machine Learning
Highway
Model
Analysis

Abstract:

Landslide (landslide) is a natural event that occurs when the upper layer of the soil slips away when certain parameters are met. This natural event occurs in many places in the world. In Turkey, landslides are observed especially in the Eastern Black Sea Region. Therefore, a landslide susceptibility map was tried to be produced in order to investigate the question of how sensitive a piece of land can be to landslides as a region. In particular, it was tried to investigate how important a landslide susceptibility map can be in determining a highway line. In our study, the taxonomy of the 35 km road line between the Ardanuç District of Artvin Province, 65.36 km2 soil region area was determined by considering 11 elements such as altitude, aspect, soil moisture index, precipitation, curvature, curvature angle, land cover, lithology, distance to drainage networks, distance to fault lines, and slope. The landslide susceptibility maps produced were divided into five susceptibility classes as very high, high, medium, low and very low. The predictive skills of the susceptibility models were examined by supervised algorithms of machine learning such as linear regression, logistic regression, support vector machine, decision tree and random forest and XG Boost (extreme gradient boosting) which would be the most suitable model.

1. Introduction

The surface of the earth is constantly changing due to natural disasters and human activities. Mass movements are one of the important factors that cause these changes. The factors that are effective in the occurrence of mass movements are the geology of the land, geomorphological features, slope, different vibrations, vegetation and disintegration [1,2]. Landslide, which is a mass movement, is the movement along the slope under the influence of gravity as a result of the destabilization of the rock, soil or similar material forming the slope and its shape and location change until it reaches equilibrium [1,3]. Landslide, which is generally referred to as a mass movement, has many different definitions in the literature, but in general, it is the movement of soil, rubble or rock along the slope under the influence of gravity. Landslides are a natural disaster that has a significant effect on the change on the earth with the destructive effect it creates on society. Landslides are a natural process that erodes and wears out the earth as long as they do not affect the vital activities of people [1,4].

1.1 Purpose of the Study

In the implementation phase of any structure model in a landslide region, it is essential to compare the factors that cause landslides in that region with various methods in a controlled manner. Projects that will be carried out without a good analysis of the region may not reach the desired result in a short time. Especially, determining and implementing road engineering structures such as highways, both land elevation stages and soil filling or cutting routes such as leveling with a model that will minimize the possibility of landslides will help to create a long-lasting and solid road. Especially, learning possible situations in a controlled manner with machine learning techniques will be a good reference source in implementing projects such as highways.

1.2 Literature

Machine learning is increasingly central to the modern economy. All industries, businesses, and consumer experiences are somehow affected by the rapid rise of this technology. Economically significant applications of machine learning include facial recognition, language translation, credit score and loan default predictions, medical diagnoses, product recommendations, driving routes, fraud alerts, etc. [5].
Biomass, especially woodchips, plays a critical role in the transition from fossil fuels, which are the primary contributors to global carbon emissions. Traditional methods of assessing woodchip quality, such as laboratory analysis for moisture content, ash content, nitrogen levels, and heating value, face limitations due to time constraints and variability in material composition. Machine Learning offers a solution by providing real-time, accurate predictions that can optimize combustion efficiency and reduce environmental impact [6].
To ensure the safe and reliable operation of power systems, grid operators must stably solve the non-convex nonlinear OPF problem in (near) real-time for massive power grids, posing enormous computational challenges. The vast amount of available data generated by the digitization of power systems and recent advances in machine learning have opened new opportunities for grid operators to create shortcuts to predict or solve the OPF problem in near real-time [7].
Machine learning methods have proven to be a useful tool to solve complex problems based on historical data in both scientific and engineering applications. These features make them a great candidate to provide a better insight into the operating characteristics of hydrogen electrochemical devices such as electrolyzers and fuel cells. In another study current research status on applying machine learning methods for predicting operating parameters, failure detection, and fault detection in hydrogen electrochemical devices, emphasizing diagnosis and prognosis have been analyzed [8].
Pipeline companies face challenges in maintaining the integrity and reliability of their pipelines. They are working towards predictive maintenance using machine learning-based approaches to predict anomalies. Training machine learning models requires sufficient data. Therefore, data quality becomes important because inaccurate data will lead to incorrect or erroneous decision regarding pipeline condition assessment and subsequent management [9].
With the increasing use of machine learning models in critical applications such as image classification, natural language processing, and cyber security, there is a growing concern about the vulnerability of these models to adversarial attacks. Evasion attacks, in particular, pose a significant threat by manipulating input data to mislead the model's predictions. It presents an overview of evasion attacks against machine learning models and performs an adaptive white-box evasion attack to highlight how defensive measures can be replaced with more powerful evasion attack algorithms [10].
As a core component of artificial intelligence, machine learning has gained significant importance in the laser cladding field in recent years. By using algorithms to analyze data, distinguish patterns and regularities, and make predictions and decisions, machine learning has significantly impacted various aspects of laser cladding processes. The emergence of defects during the cladding procedure poses significant challenges to the quality and performance of the cladding layers. Addressing the reliability and reproducibility of the cladding quality is a major concern in laser cladding technology. Leveraging data-driven machine learning algorithms enables the monitoring and detection of defects throughout the laser cladding process [11].

2. Materials and Methods

Various methods can be used to produce landslide maps. The most commonly used of these is the Analytical Hierarchy Method, which is used in comparing studies based on geographic information systems. With the AHP comparison matrix, mapping about values and the most needed elements can be determined. With this determination, the parameters needed in a complete landslide mapping are analyzed graphically and statistically by shedding light on the machine learning supervised analysis sections. In the study, the values of the elements that cause landslides to be examined with five techniques of machine learning together with AHP were tried to be examined. These machine learning techniques are linear regression, logistic regression, support vector machines, detection from decision trees and extreme gradient boosting, which are varieties of supervised learning.

2.1 AHP (Analytical Hierarchy Proces)

It was developed by Saaty in 1977 as a usable model for calculating multi-criteria decision-making problems [12.13]. It is one of the most widely used methods today. The reason for this is that AHP allows the use of qualitative and quantitative criteria together in the evaluation and selection processes of decision options [12,14]. This method is used in multi-purpose choices where there are many decision makers when choosing from a large number of alternatives [12].
In the hierarchy design, levels and elements are determined, concepts are defined and questions are formulated. Thus, a hierarchy is created with the element representing the general purpose at the top. The second stage is the evaluation of the hierarchy. At this stage, the relationships between the two Elements in the hierarchy are compared. In this evaluation, a pairwise comparison matrix, which is in the form of a square matrix, is used. In each comparison, it is decided which of the two elements is important depending on the element at the upper level, which expresses the degree of importance. Numbers are used in the expression of the degree of importance [12,15].
This analysis consists of three main steps: determining the parameter weights, assigning the calculated weights, and summing these parameters. The Analytical Hierarchy Process Method (AHP) can be used to express the relative contribution of each parameter to the overall sensitivity in terms of parameter-specific weight [12,16,17]. In the AHP method, the inconsistency index, known as the consistency ratio (CR), indicates the probability that the matrix evaluations are performed randomly [12,16,17].

$$CR = CI / RI \qquad (1) \qquad [12].$$

RI is the mean of the resulting consistency index and is based on the matrix order given by [16], while CI is the consistency index [12].

CI can also be expressed as;

$$CI = (\lambda max - n)/(n - 1) \qquad (2) \qquad [12].$$

λmax is the largest or principal eigenvalue of the matrix and can be easily calculated from the matrix, and n is the rank of the matrix. A CR of 0.1 or less is a reasonable level of sensitivity [12,17,18]. A CR above 0.1 requires reconsideration of the matrix evaluation due to inconsistent operation of the specific factor ratings [12].

2.2. Linear Regression

Expressing the relationship between two or more variables with a straight line is called linear regression [19].

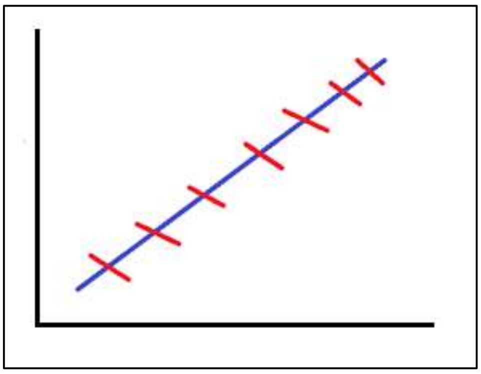

Figure 1. *Linear Regression Schematic Representation*

In Figure 1,everything can be seen much more easily thanks to the drawn line. As for what is included; y is the formula of the line, and R^2 is the score of the independent variable's ability to explain the dependent variable [19].

$Y=Ax+B$ (3) Estimated value equation [19].

2.3. Logistic Regression

Logistic regression method, which is the most popular of the multivariate statistical methods, is preferred for producing susceptibility maps in landslide susceptibility studies [20,21].

The purpose of logistic regression is to explain the relationship between the dependent variable and the independent variables and to express the result of the dependent variable as a probability according to the independent variables [20,22]. In addition, the logistic regression method is a regression method that allows assignment and classification.

$\log\left(\frac{p}{1-p}\right) = Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$ (4) [20].

In this model, Y represents the dependent variable between 0 and 1, β_0 represents the dependent variable or constant when the independent variables take the value of zero, $\beta_1,\ldots,\beta_n$ represents the regression coefficients of the independent variables and $X1,\ldots,Xn$ represents the independent variables. Detailed explanations of the logistic regression technique can be found in various textbooks in the literature [20,23,24].

2.4. Support Vector Machine

In Figure 2, support vector machine (SVM), also known in the literature as Support vector machine (SVM), is one of the machine learning algorithms and is used in applications such as classification, regression and discrimination. The SVM algorithm is based on two basic operations [25,26]. One of these is the creation of a hyperplane and the other is the separation of data sets according to this plane [25]. However, data sets may not always be perfectly separated by a hyperplane. For this reason, using kernel functions, the input data is transformed in a way that it can be separated on the hyperplane. The algorithm created for this operation uses this hyperplane to classify the data in n-dimensional space and to maximize the separation between the classes [25].

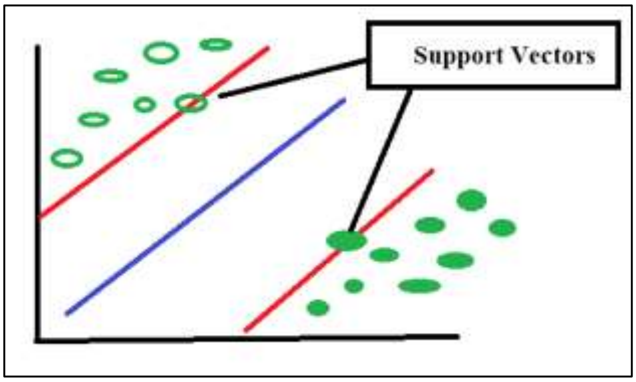

Figure 2. *Support Vector Machine Representation [25,27,28].*

2.5. Random Forest Algorithm

The Random Forest algorithm is a learning algorithm in which combinations of many randomly selected decision trees are trained [29,30]. The working principle of the algorithm is based on the principle that in an example where the estimation process is performed, the final decision is made by combining the predictions made by each of the decision trees that come together to form a forest [29,31]. Here, the selected decision trees, each of which performs a separate classification, come together to form decision forests. The higher the number of decision trees determined by the user, the higher the accuracy that can be obtained from the result [29,30]. The Random Forest algorithm is a frequently used method among machine learning techniques because it exhibits high performance in the classifications made [29].

2.6. Extreme Gradient Boosting Algorithm

The algorithm is a widely used, scalable, effective tree boosting technique. It produces results ten times faster than existing similar popular systems and can be scaled to billions of samples [29,32]. They stated that the most important factor in the success of the XGBoost algorithm is its scalability in all cases [29]. XGBoost algorithm is used to solve regression and classification problems. It is based on decision trees and is considered the best among these algorithms. It is a method consisting of combinations of predictions and achieves high predictive power. The algorithm creates a single strong model by combining multiple weak prediction models. Instead of examining all the values in the data, XGBoost divides the data into pieces and works according to the parts it separates. Here, the more the number of pieces, the smaller the intervals will be and as a result, the algorithm will make higher predictions [29,33].

3. Findings

In Figure 3 and 4, the image of the 35-40 km highway route connecting Artvin City Center and Ardanuç District, extracted from the Sentinel 2 [34] satellite image, along approximately 100 meters of the north and south lines of the road route, was taken as the study area. The 40 km starting and ending point of the study area between Artvin center and Ardanuç center, the stream branch of the Deriner Dam located on the immediate south axis along the road line, the road, and the landslide areas continuing along the northern 100-meter axis of the road were determined as the legend of the study area.

The highway line is approximately between $41^0 20$ E-$41^0 30$ E-$41^0 40$ E and $41^0 70$ N-$41^0 80$ N-$41^0 90$ N, latitude and longitude. While the Artvin city center is approximately 345 meters high, the Ardanuç District center is approximately 558 meters.

Along the road route determined in the study area, since the south of the line is water, it was reduced to 100 meters north and the question of the extent to which this line could be exposed to landslides was tried to be answered. For this, 11 parameters used especially in the landslide inventory were applied to this road line.

Maps were obtained with the help of QGIS and ARCGIS by creating legends generally in 7 or above, based on 11 elements including elevation, aspect, soil moisture, precipitation, curvature, curvature angle, land cover, lithology, distance to drainage networks, distance to fault lines, slope. The sensitivity analyses of the model created with the 11 landslide inventory were tried to be evaluated with comparison matrices with a total of 6 methods using Analytical Hierarchy (AHP), linear, logistic regression, support vector model,

random forest and XG Boost machine learning and Python.

3.1 Area Based Landslide Parameter Displays

The inventory base data of 11 landslide parameters were created with QGIS in the light of the landslide inventory map obtained from Artvin Provincial Disaster and Emergency Directorate [35], 1/25,000 scale topographic maps obtained from General Directorate of Mapping [36], 1/100,000 scale Corine land cover data obtained from European Union Copernicus Land Monitoring Service [37], 1/25,000 scale digital forest inventory data obtained from Artvin Forest Regional Directorate, 1/25,000 scale digital geological map obtained from General Directorate of Mineral Research and Exploration [38], and last 20 years data from General Directorate of Meteorology [39].

The legend formations of 11 parameters affecting the road line are determined between 7 and 9 values. The reason for this is the need contained in the parameter.

In Figure 5, it is revealed that the slope degrees in the landslide inventory increase from 2^0 to 37^0 in the north of the 35 km road line with 5 degrees increase. The height of the line is based on the average of 345 meters from the sea level starting point of the city entrance, and it is found that the road rises to 585 meters along the route

with a 30-meter increase difference and then descends to 400 meters again. In the rainfall profile, it is revealed that the average rainfall amount is 255 mm in 15-20 km of the road line and there is a rainfall profile that rises and falls to 750 mm on an annual basis.In Figure 6, the aspect representation, which is the shaped form of the road line according to the directions, the straight, concave and convex point representations of the road route as curved, and the drainage network, which shows the drainage of the road during rainfall and its function as drying, are presented at 20-meter intervals.

In Figure 7, it is determined that the scale of the soil moisture index expressed as TWI is approximately between 23 and 1, and the curvature angle is passed at 100-meter intervals. Since Artvin has a geography that is generally not centrally located to the fault line, the distance amount of the route was determined at 300-meter

intervals in the small-scale fault diagram.

Soil moisture index is TWI=$In(\frac{A}{\tan(B)})$ (5) [40,41].

It is stated that the section between the flow area and the slope will be determined according to the value obtained by taking In.

In Figure 8, the rock types from the city center to the district and the rock formation determined by the district name at the peak of the district are expressed with the ground map, and how the vegetation spreads on the land along the route is shown.

4. Results and Analysis (Modeling)

4.1. AHP Analysis

With the Analytical Hierarchy Method, the importance parameter is determined by giving an odd number from 1 to 9 or an even number of values according to the equality importance between the odd numbers according to the comparison matrix consistency rate of 11 criteria of the landslide with very low, low, medium, high and very high importance [42].

In Tables 1 and 2, the hypothetical pairwise comparison matrix of 11 parameters and the criteria, weight criteria and weight classes of the study area are shown in 5 classes, revealing that altitude, slope, lithology, land cover, distance to fault lines and precipitation rate are more important than other inventory elements.

AHP period degrees are determined by taking the parameters from the starting point to the 5th level in parallel with the legend and remaining with the 5th importance level. Importance levels are generally selected in this way.

4.2. Landslide Susceptibility Mapping and Model Analysis

The line examined was determined as a surface area of 65.36 km^2 between 35-40 km. The study area contains a total of 95 landslide polygon points and 14,822 landslide pixels.

The pixels with landslides are numerically defined as 1, and the pixels without are defined as 0. The number of non-landslide pixels is added as much as the number of landslide pixels, which indicates that the model performance can be seriously affected if the number of landslide and non-landslide cells in the training data set is significantly unbalanced. For this reason, researchers use as many non-landslide samples as the number of landslide samples in their studies [43,44,45,46,47]. Generally, the entire data group is used as a training and test group at a certain percentage. Although there is no accepted rule in the creation of two subsets, the majority of researchers in landslide susceptibility mapping studies use a ratio of 70:30, especially in the selection of "landslide" samples. In this approach, 70% of the randomly selected landslides in the inventory map are used for training the models and the remaining 30% are used for model validation [48,49,50,51,52].

The study, as a result of the correlation of 11 landslide sets in a total number of 29,644 pixels, when the training and model test were separated by 70 percent and 30 percent, the prediction rate was not examined since the linear and support data machine model could reach 0.60 and 0.65 success rates. These two machine supervised learning were eliminated.The success rate reached around 92-95-99 percent with logistic regression, random forest and XG Boost learning. Therefore, the model was established on 3 machine learning.

Logistic regression

As seen in Table 3 and Table 4, the study area was tested with a total of 29,644 pixel numbers, with a success rate of 70% and 30% for training and testing within the framework of the estimated sensitivity model, and the success rate was 0.920, and the estimated result was 0.894 frequency rate, with a pixel count of 20750 and a pixel count of 8894.

Random Forest and XG Boost Modeling

In Table 5, the landslide percentage sub-percentages corresponding to the legend content of the parameters that have a direct effect on 6 landslides of the 14,822 landslide area pixel count of the landslide set and the frequency ratios they create are shown. Thus, it has been revealed how important the slope degree and height are.

4.3. Landslide Susceptibility Mapping

In Figure 9, with Random Forest, the spatial distribution was found to be 45.6% very low, 24.3% low, 15.6% medium, 5% high, and 9.5% very high.

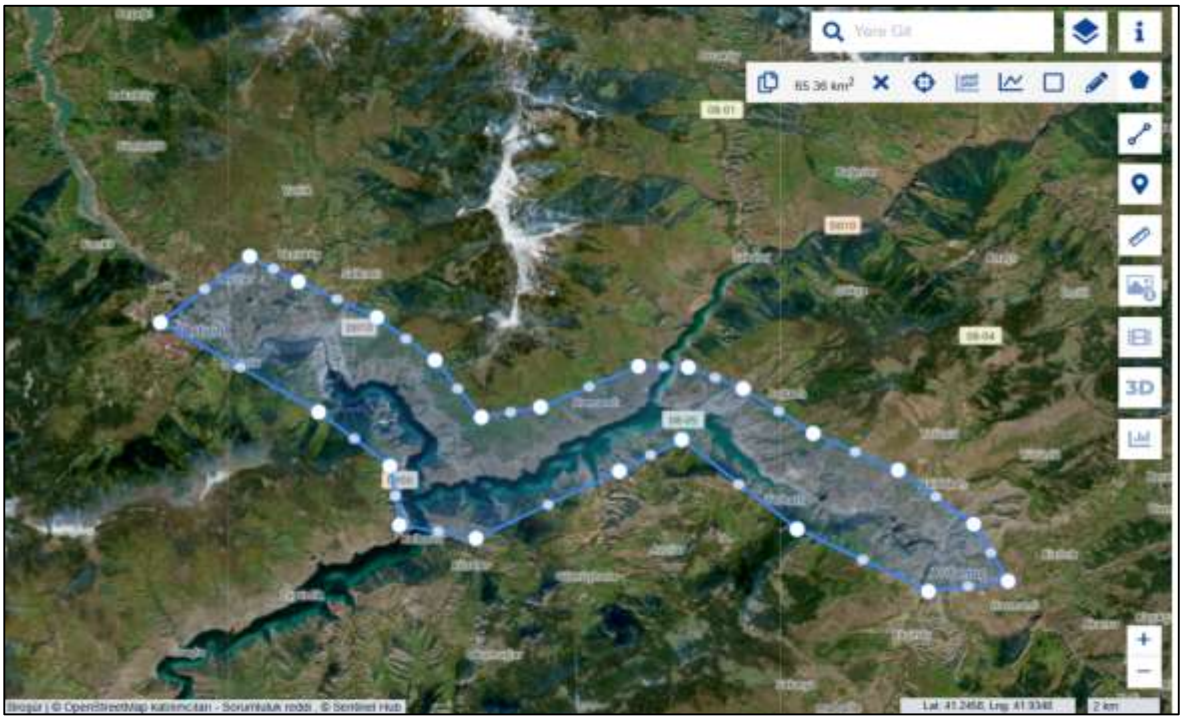

Figure 3. *Satellite image with 10 meter resolution [34].*

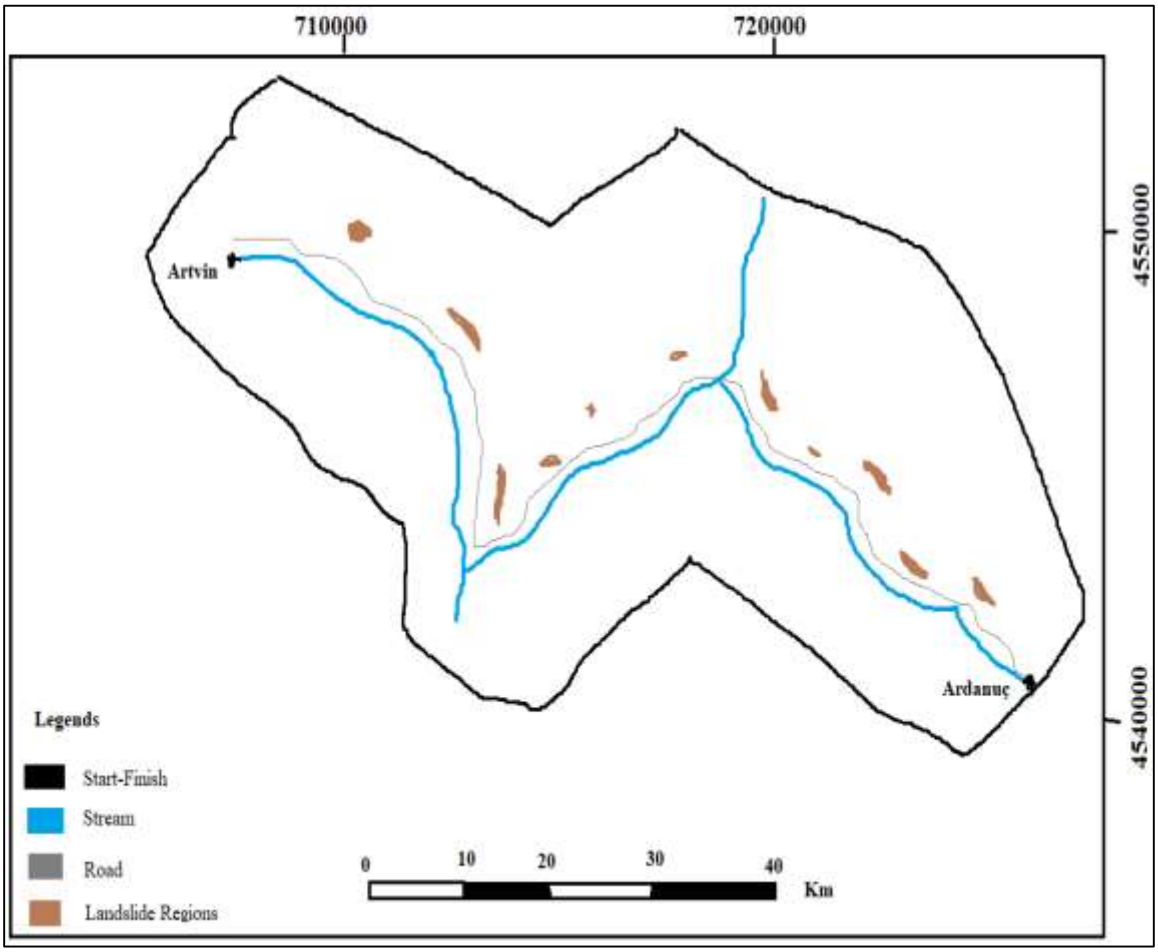

Figure 4. *Representation of the study area*

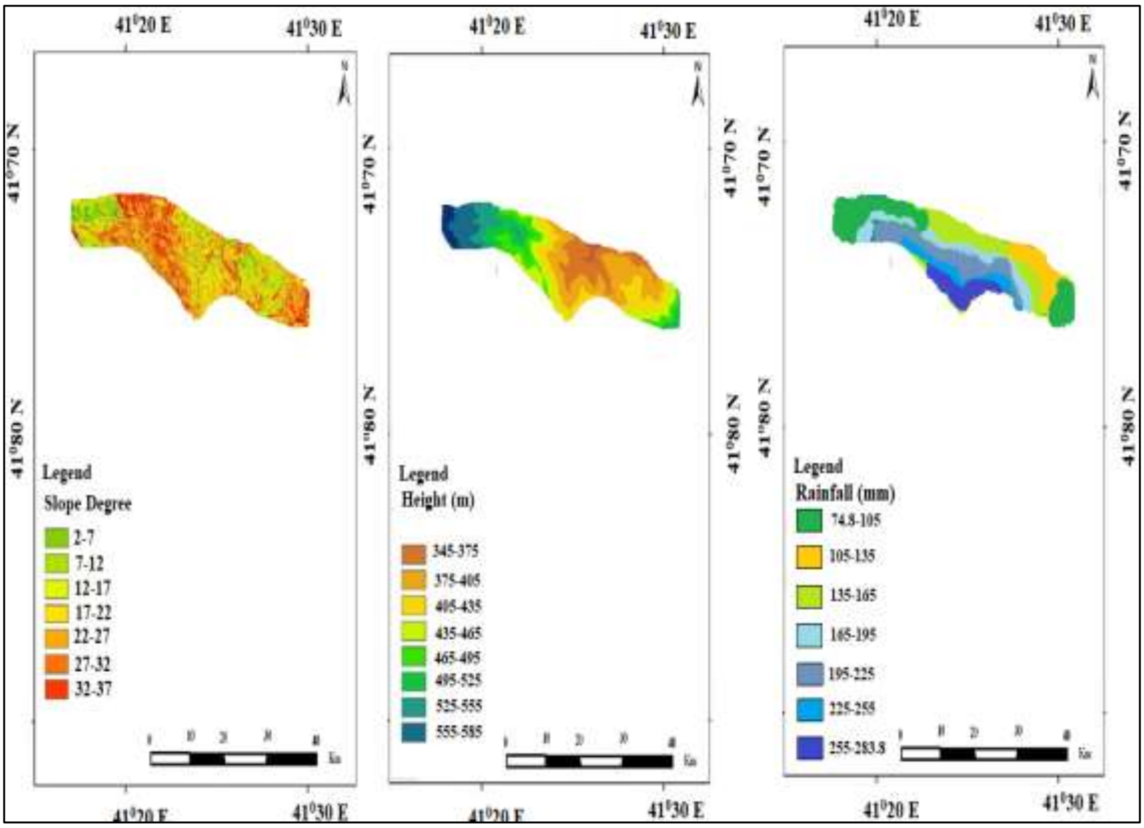

Figure 5*. Slope, Elevation and Rainfall Map Output*

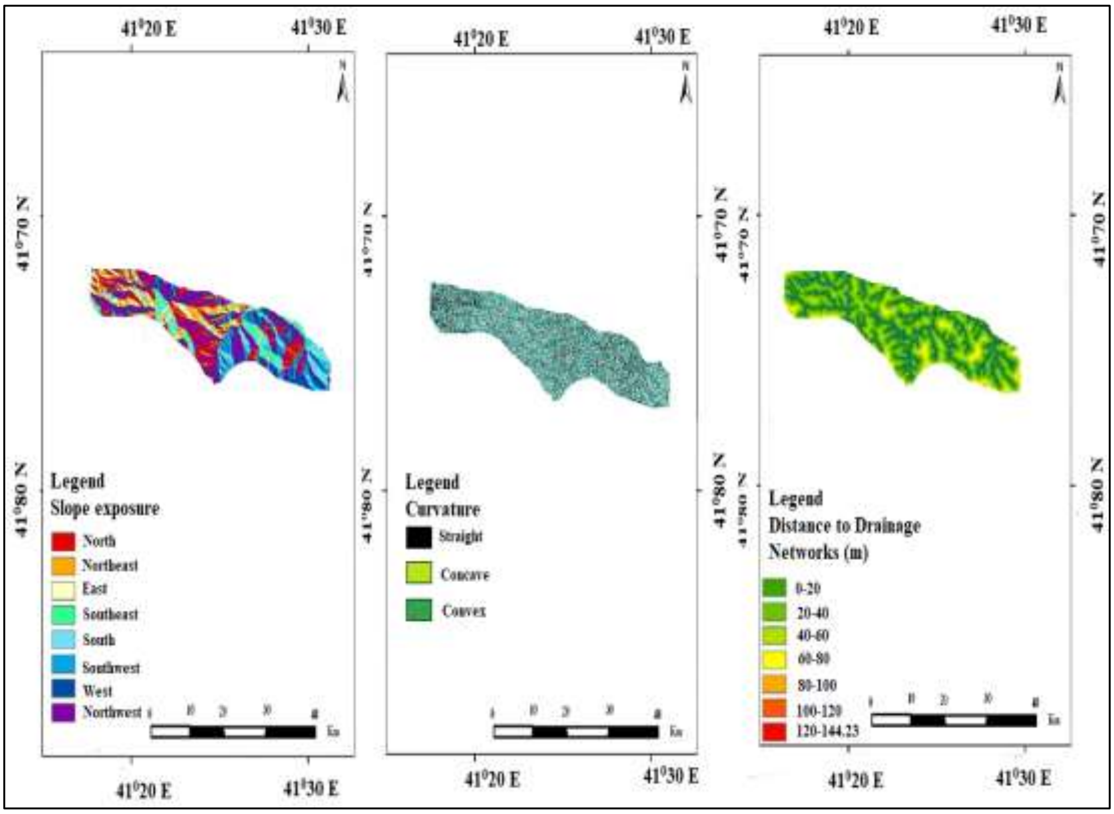

Figure 6. *Aspect, Curvature and Distance to Drainage Networks Map Output*

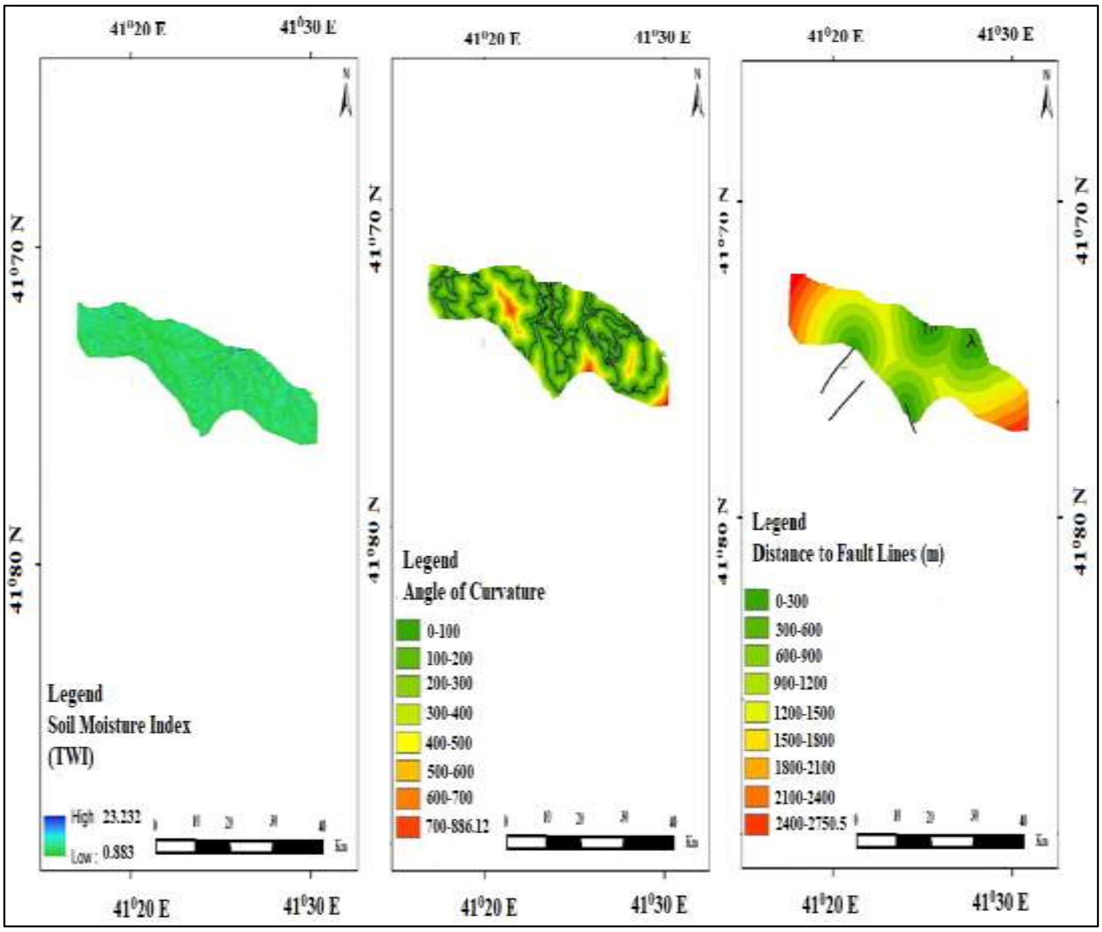

Figure 7. *Soil Moisture Index, Curvature Angle and Distance to Fault Lines Map Output*

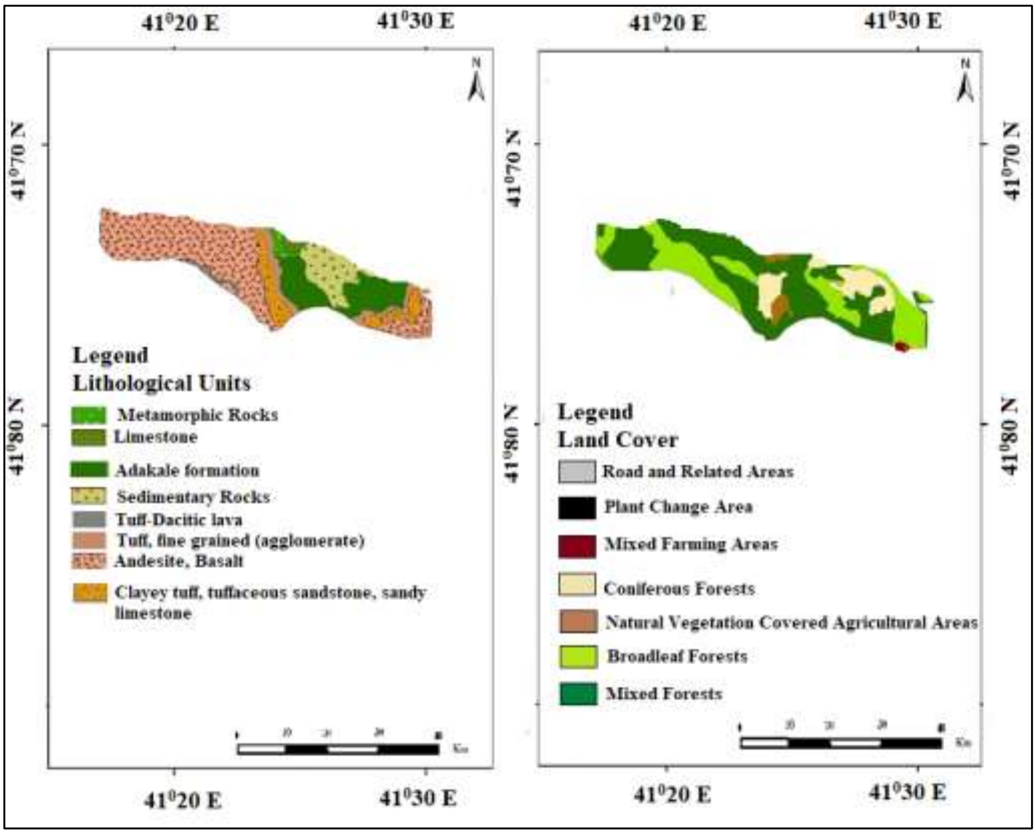

Figure 8. *Lithology and Land Cover Information Map Output*

Table 1. *Binary Landslide Parameters Comparison Matrix*

	A	B	C	D	E	F	G	H	I	J	K
A (Slope)	1	5	7	3	2	5	5	3	4	8	6
B (Height)	1/5	1	8	2	3	5	6	2	4	5	4
C (Rainfall)	1/7	1/8	1	2	3	6	5	3	3	7	4
D (Slope Exposure)	1/3	1/2	1/2	1	3	4	2	3	3	2	3
E (Curvature)	1/2	1/3	1/3	1/3	1	2	2	3	3	3	2
F (Drainage)	1/5	1/5	1/6	1/4	1/2	1	5	4	2	6	5
G (TWI)	1/5	1/6	1/5	1/2	1/2	1/5	1	2	2	4	3
H (Angle of Curvature)	1/3	1/2	1/3	1/3	1/3	1/4	1/2	1	2	3	4
I (Fault Lines)	1/4	1/4	1/3	1/3	1/3	1/2	1/2	1/2	1	7	4
J (Lithology)	1/8	1/5	1/7	1/2	1/3	1/6	1/4	1/3	1/7	1	5
K (Land Cover)	1/6	1/4	1/4	1/3	1/2	1/5	1/3	1/4	1/4	1/5	1

Table 2. *Research Criteria, Weight Degrees and Weight Classes*

Criterion	Criterion Factors	Weight Degree	Weight Class
A (Slope)	22+	5	Very High
	17-22	4	High
	12-17	3	Middle
	7-12	2	Low
	2-7	1	Very Low
B (Height)	465+	5	Very High
	435-465	4	High
	405-435	3	Middle
	375-405	2	Low
	345-375	1	Very Low
C (Rainfall)	195+	5	Very High
	165-195	4	High
	135-165	3	Middle
	105-135	2	Low

	75-105	1	Very Low
D (Slope Exposure)	North Northeast East Southeast South	5 4 3 2 1	Very High High Middle Low Very Low
E (Curvature)	Convex Convex Concave Straight Straight	5 4 3 2 1	Very High High Middle Low Very Low
F (Drainage)	80+ 60-80 40-60 20-40 0-20	5 4 3 2 1	Very High High Middle Low Very Low
G (TWI)	23+ 18-23 13-18 8-13 0.8-8	5 4 3 2 1	Very High High Middle Low Very Low
H (Angle of Curvature)	400+ 300-400 200-300 100-200 0-100	5 4 3 2 1	Very High High Middle Low Very Low
I (Fault Lines)	1200+ 900-1200 600-900 300-600 0-300	1 2 3 4 5	Very High High Middle Low Very Low
J (Lithology)	Andesite-Basalt Clayey tuff Adakale formation Sedimentary rocks Limestone	1 2 3 4 5	Very High High Middle Low Very Low
K (Land Cover)	Mixed forests Broadleaf forests Coniferous forests Plant change area Road and Related areas	1 2 3 4 5	Very High High Middle Low Very Low

Table 3. *Logistic Regression Error Matrix*

		Education		
Test		0	1	Total
	0	979	7915	8894
	1	1660	19090	20750

Table 4. *Performance Metric Calculation Display*

	Exact Value	**Sensitivity Value**	**F ratio**	**Pixel Count (Landslide + Landslide-free)**
0	0.894	0.894	0.894	8894
1	0.920	0.920	0.920	20750

Table 5. *Model Comparison Landslide Percentage and Frequency Rates*

Parametres	**Legend**	**Landslide Sub-Percentage (A)**	**Landslide Percentage (B)**	**Frekans Ration (A/B)**
Slope Degree	2-7	25.9	25.8	1.003
	7-12	19.6	20.3	0.965
	12-17	14.3	14.5	0.986
	17-22	14.2	14.5	0.979
	22-27	14.1	14.4	0.979
	27-32	4.9	5.2	0.942
	32-37	4.8	5.3	0.905

Height	345-375	22.4	20.8	1.076
	375-405	25.9	24.9	1.040
	405-435	18.3	20.6	0.888
	435-465	11.2	10.0	1.120
	465-495	11.8	10.7	1.102
	495-525	7.9	8.6	0.918
	525-555	4.2	4.4	0.954
	555-585			
Rainfall	74.8-105	4.8	5.3	0.905
	105-135	11.2	12.3	0.910
	135-165	14.6	16.8	0.869
	165-195	11.0	12.6	0.873
	195-225	14.3	15.4	0.928
	225-255	17.2	18.6	0.924
	255-283.8	20.3	19.0	1.068
Distance to Faults Lines	0-300	12.6	10.6	1.188
	300-600	14.6	12.5	1.168
	600-900	22.4	13.4	1.671
	900-1200	17.8	15.4	1.155
	1200-1500	9.6	10.9	0.880
	1500-1800	14.5	16.5	0.878
	1800-2100	22.3	20.7	1.077
	2100-2400	0	0	0
	2400-2750.5			
Lithogical	Metamorphic	18.2	17.3	1.052
	Limestone	14.0	12.5	1.120
	Adakale	9.3	10.9	0.853
	Sedimentary	15.4	22.6	0.681
	Dacitic Lava	19.2	17.3	1.109
	Tuff	4.2	5.8	0.724
	Andesite, Basalt	5.3	6.4	0.828
	Clayey tuff	8.1	7.2	1.125

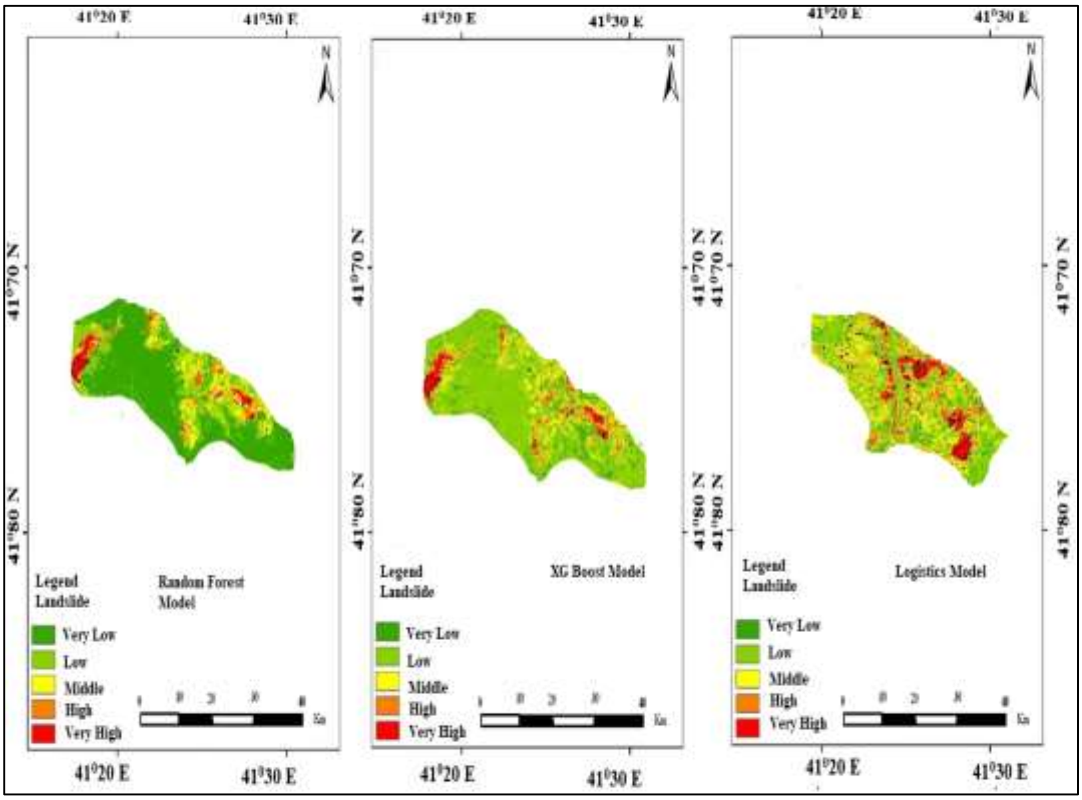

***Figure 9.** Landslide Susceptibility Maps*

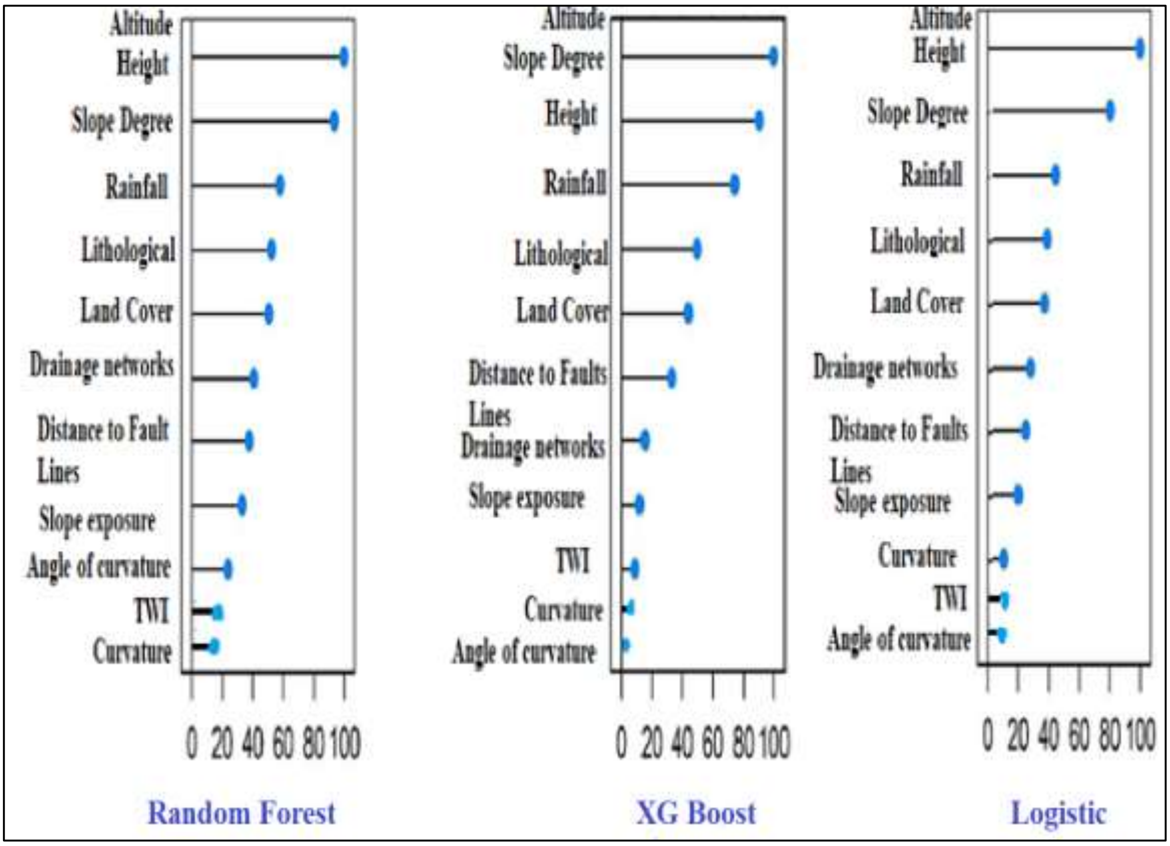

Figure 10. *Importance levels of landslide conditioning factors*

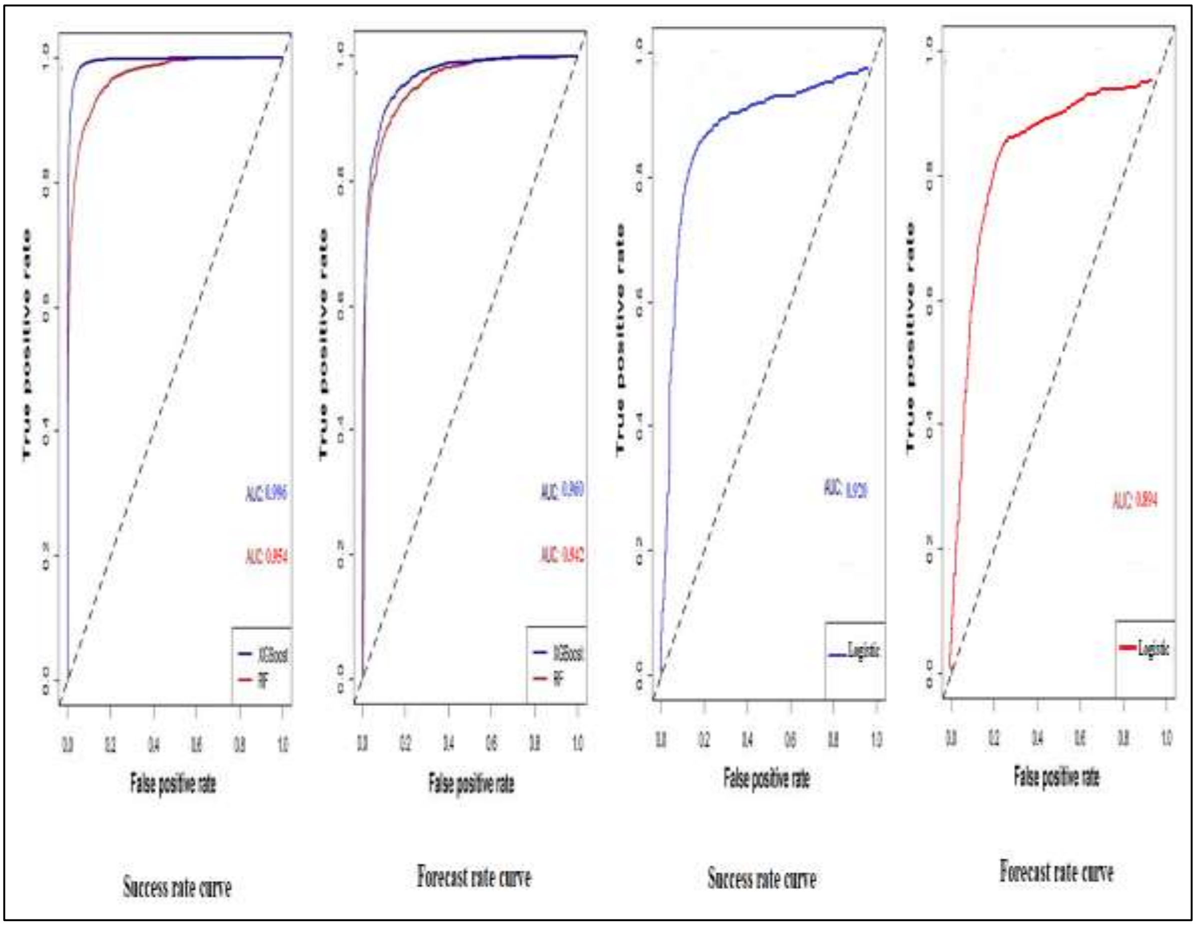

Figure 11. *Model Performance Comparisons (ROC Curves)*

With XG boost, it was found to be 9.6% very low, 60.3% low, 13.6% medium, 5% high, and 11.5% very high. With logistic regression, it was found to be 5% very low and 14.7% very high.

In Figure 10, it is concluded that the slope degree, altitude and precipitation are very important in the taxonomy of 11 landslide parameters in the random forest, Xg boost and logistic regression models, while the soil moisture index, curvature and curvature angle are the least important.

In Figure 11, the Receiver Operating Characteristic (ROC) curve and the Area Under Curve (AUC) method were used to evaluate the models and compare their performances. Since the ROC curve will give the best result vertically towards 1, when the success rate and landslide prediction rates are examined;
Random forest showed success in the field as AUC value in the success rate as 0.954, Xg boost as 0.996, and logistic model as 0.920. In the prediction value, the landslide prediction value of the study area was revealed as random forest 0.942, Xg boost as 0.960, and logistic as 0.894. In this case, it was seen that Xg boost gave the best result and the logistic model gave the worst result. The most advantageous model was determined as xg boost.

5. Discussion and Recommendation

Machine learning models are widely used for various types of problems in natural disaster susceptibility mapping, including classification, clustering, and regression. In developing susceptibility maps for natural disasters using machine learning models, data on previous events (inventory data) plays an important role in training and testing the model [53]. Acknowledges limitations, including the potential inadequacy of the chosen conditioning factors, the impact of resampling on data accuracy, the representativeness of the sampling rate, and the limited number of training examples [54].

In the study, if any structure is to be built in a region exposed to landslides, it has been concluded that the most appropriate project can be achieved with machine learning as a result of investigating how sensitive the area is to landslides. When the factors causing a suitable number of landslides are compared with more than one method, if the same result is achieved, the most effective parameters are evaluated accordingly. In general, it has been concluded that precipitation, slope, height, fault lines, drainage networks, ground structure, land cover are direct factors, while soil moisture index, curvature and curvature angles are the least effective in a landslide parameter base. Our suggestion is that when making a plan, it should be taken into consideration how much this inventory formation can be in urbanization or road network design.

As technology develops, needs change, and the need and desire to carry out more comprehensive and complex projects emerge, new management models are being developed with innovative approaches in addition to traditional construction and nature-based management processes [55].

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Altürk, G. (2019). Production of Landslide Susceptibility Maps Using Machine Learning and Statistical Methods in Geographic Information Systems Environment: Rize Taşlıdere Basin Example, Gebze Technical University, Institute of Science, Department of Surveying Engineering, Geodesy and Geographic Information Technologies Program, Master's Thesis, Gebze. (Yök Thesis No:604517) https://tez.yok.gov.tr/UlusalTezMerkezi/

[2] Özdemir, A. (2008). Introduction to Soil Mechanics and Soil Engineering, Konya: İnci Press.

[3] Cruden, D. (1991). A simple definition of a landslide, Bulletin of the International Association Engineering Geology, 27-29.

[4] Akgün, A. (2007). Investigation of Erosion and Landslide Susceptibility of Ayvalık and Its Surroundings Based on Geographic Information Systems, PhD Thesis, Dokuz Eylül University. (Yök Thesis No:213159)

https://tez.yok.gov.tr/UlusalTezMerkezi/
[5] Caplin, A., Martin, D., and Marx, P. (2025). Modeling machine learning: A cognitive economic approach, *Journal of Economic Theory, J. Econ. Theory* 224 (2025) 105970, page:1-14. https://doi.org/10.1016/j.jet.2025.105970
[6] Gasperini, T., Yeşil, V., and Toscano, G. (2025). Machine learning and woody biomasses: Assessing wood chip quality for sustainable energy production, *Biomass and Bioenergy* 193 (2025) 107527, page:1-12. https://doi.org/10.1016/j.biombioe.2024.107527
[7] Khaloie, H., Dolanyi, M., Toubeau,J.F., and Vallee,F. (2025). Review of machine learning techniques for optimal power flow, *Applied Energy* 388 (2025) 125637, page:1-23. https://doi.org/10.1016/j.apenergy.2025.125637
[8] Franic, N., Pivac, I., and Barbir, F. (2025). A review of machine learning applications in hydrogen electrochemical devices, *International Journal of Hydrogen Energy* 102 (2025) 523–544, page: 523-544. https://doi.org/10.1016/j.ijhydene.2025.01.070
[9] Hussain, M., and Zhang, T. (2025). Machine learning-based outlier detection for pipeline in-line inspection data, *Reliability Engineering and System Safety* 254 (2025) 110553, page:1-12. https://doi.org/10.1016/j.ress.2024.110553
[10] Muthalagu, R., Malik, J., and Pawar, P.M. (2025). Detection and prevention of evasion attacksson machine learning models, *Expert Systems With Applications* 266 (2025) 126044, page:1-23. https://doi.org/10.1016/j.eswa.2024.126044
[11] Ji, X. C., Chen, R. S., Lu, C. Recent advances in machine learning for defects detection and prediction in laser cladding process, *Next Materials* 7 (2025) 100404, page:1-21. https://doi.org/10.1016/j.nxmate.2024.100404
[12] Pepper, T.R. (2019). Comparison of Methods Used in Producing Landslide Susceptibility Maps: Şebinkarahisar District Example, Karadeniz Technical University, Institute of Science, Department of Surveying Engineering, Master Thesis, Trabzon. (Yök Thesis No:595869) https://tez.yok.gov.tr/UlusalTezMerkezi/
[13] Kavas, E. (2009). Geographic Information Systems Based Investigation of Landslide Susceptibility in İzmir Province with Analytical Hierarchical Process Method, TMMOB Geographic Information Systems Congress (CBS2009), Kasım, İzmir.
[14] Yalçın, A. (2008). GIS-Based Landslide Susceptibility Mapping Using Analytical Hierarchy Process And Bivariate Statistics in Ardesen (Turkey): *Comparisons Of Results And Confirmations*, Catena, 72, 1–12.
[15] Çellek, S. (2013). Landslide Susceptibility Analysis of Sinop-Gerze Region, PhD Thesis, Karadeniz Technical University, Institute of Science, Department of Geological Engineering, Trabzon.
[16] Saaty, T.L. (1980). The analytical hierarchy process, McGraw Hill, New York.
[17] Akgün, A. (2012). A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey, Springer-*Verlag, Landslides*, 9, 1, 93-106.
[18] Malczewski, J. (1999). GIS and multicriteria decision analysis. Wiley, New York.
[19] Url 1: https://medium.com/deep-learning-turkiye/her-%C5%9Feyiyle-lineer-regresyon-makine-%C3%B6%C4%9Frenmesi-serisi-1-1ee2aec10b74 (Date of Access:10.03.2025).
[20] Cezayirlioğlu, C. (2023). Assessment of Coastal Landslide Susceptibility with Geographic Information Systems and Logistic Regression Method: Between Ayancık-Gerze, (SİNOP), Eskişehir Technical University Graduate Education Institute, Department of Remote Sensing and Geographic Information Systems, Master Thesis, Eskişehir. (Yök Thesis No:783776) https://tez.yok.gov.tr/UlusalTezMerkezi/
[21] Yilmaz, I. (2010). Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine, *Environmental Earth Sciences*, 61(4), 821 836.
[22] Elmacı, H. (2016). GIS based landslide susceptibility analysis between Çubuk and Kalecik districts of Ankara province and Şabanözü districts of Çankırı province, M.S. Thesis. Department of Environmental Engineering, Gazi University, Institute of Science.
[23] Hosmer, D.W., Lemeshow, S. (2000). Applied Logistic Regression. John Wiley & Sons, New York. 375 pp.
[24] Kleinbaum, D.G., Klein, M. (2002). Logistic regression: A self learning text, 2nd Ed. Springer Verlag, New York. 513 pp.
[25] Horzum, A.I. (2024). Landslide Susceptibility Mapping Using Machine Learning Algorithms, Akdeniz University Institute of Science, Department of Geological Engineering, Master Thesis, Antalya. (Yök Tez No:850586) https://tez.yok.gov.tr/UlusalTezMerkezi/
[26] Vapnik, V. (1995). Nature of Statistical Learning Theory. John Wiley and Sons, Inc., New York.
[27] Burges, C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition, in: Fayyad, U. (Ed.), Data Mining and Knowledge Discovery. Kluwer Academic Publishers, Boston, The Netherlands, pp. 121-167.
[28] San, B.T. (2014). An evaluation of SVM using polygon-based random sampling in landslide susceptibility mapping: The Candir catchment area (western Antalya, Turkey) International Journal of Applied Earth Observation and Geoinformation 26 , 399-412.
[29] Mutlu, Ö. (2023). Generation of Landslide Susceptibility Maps with Machine Learning Techniques: Hopa (Artvin) Example, Artvin

Coruh University Graduate Education Institute, Department of Surveying Engineering, Master Thesis, Artvin. (Yök Thesis No:833051) https://tez.yok.gov.tr/UlusalTezMerkezi
30] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1): 5-32.
[31] Çölkesen, İ. and Şahin, E. K. (2020). Investigating the effects of training set size on map accuracy in landslide susceptibility analysis, *Harita Dergisi*, 164: 43-60.
[32] Chen, T. and Guestrin, C., 2016. XGBoost: A scalable tree boosting system, KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco (pp. 785-794), doi: 10.1145/2939672.2939785.
[33] Muratlar, E. R. (2020). How Does XGBoost Work? Why Does It Perform Well? https://www.veribilimiokulu.com/xgboost-nasil-calisir/ (June 29, 2023, 15:37).
[34] Url 2: https://dataspace.copernicus.eu/explore-data (Date of Access:10.03.2025).
[35] Url 3: https://www.afad.gov.tr/planlar (Date of Access:15.01.2025).
[36] Url 4: https://www.harita.gov.tr/ (Date of Access:12.01.2025).
[37] Url 5: https://corine.tarimorman.gov.tr/corineportal/araziortususiniflari.html (Date of Access:11.02.2025).
[38] Url 6: https://www.mta.gov.tr/ (Date of Access:02.01.2025).
[39] Url 7: https://www.mgm.gov.tr/ (Date of Access:22.01.2025).
[40] Quinn, P. F., Beven, K. J., and Lamb, R. (1995). The In (A/Tanβ) Index: How to calculate it and how to use it within the topmodel framework, hydrological processes 9, 161 182.
[41] Taşoğlu, E. (2020). Landslide susceptibility assessment of Devrek district (Zonguldak) using Artificial Neural Networks, Master's Thesis, Karabük University Graduate Education Institute, Karabük. https://tez.yok.gov.tr/UlusalTezMerkezi/
[42] Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European journal of operational research*, 48(1), 9-26.
[43] Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, *Computers & Geosciences*, 51: 350-365.
[44] Kumar, D., Thakur, M., Dubey, C. S. and Shukla, D. P. (2017). Landslide susceptibility mapping & prediction using support vector machine for Mandakini River Basin, Garhwal Himalaya, India, Geomorphology, 295: 115-125.
[45] Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., Zhu, AX., Pei, X., and Duan, Z. (2018). Landslide susceptibility modeling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China, *Science of the Total Environment*, 626: 1121-1135.
[46] He, Q., Shahabi, H., Shirzadi, A., Li, S., Chen, W., Wang, N., Chai, H., Bian, H., Ma, J., Chen, Y., Wang, X., Chapi, K., and Ahmad, B. B. (2019). Landslide spatial modeling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms, *Science of the Total Environment*, 663: 1-15.
[47] Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H. and Hong, H. (2020). Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping, Catena, 187: 104396.
[48] Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Tien Bui, D., Duan, Z. and Ma, J. (2017). A comparative study of logistic model tree, randomforest, and classification and regression tree models for spatial prediction of landslide susceptibility, Catena, 151: 147-160.
[49] Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., and Pourghasemi, H. R. (2018). Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China, *Computers and Geosciences*, 112: 23-37.
[50] Dou, J., Yunus, A. P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.W., Khosravi, K., Yang, Y., and Pham, B. T. (2019). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan, *Science of the Total Environment*, 662: 332-346.
[51] Park, S. and Kim, J. (2019). Landslide Susceptibility Mapping Based on Random Forest and Boosted Regression Tree Models, and a Comparison of Their Performance, *Applied Sciences*, 9(5), 942.
[52] Soma, A. S., Kubota, T. and Mizuno, H. (2019). Optimization of causative factors using logistic regression and artificial neural network models for landslide susceptibility assessment in Ujung Loe Watershed, South Sulawesi Indonesia, *Journal of Mauntain Science*, 16(2): 383–401.

[53] Bilgilioglu, S.S., Gezgin, C., Iban, M.C., Bilgilioglu, H., Gündüz, H.I., Arslan, S. (2025). Explainable Sinkhole Susceptibility Mapping Using Machine-Learning-Based SHAP: Quantifying and Comparing the Effects of Contributing Factors in Konya, Türkiye. *Appl. Sci.* 2025, 15,3139. https://doi.org/10.3390/app15063139

[54] Iban, M.C., Aksu, O. (2024). SHAP-Driven Explainable Artificial Intelligence Framework for Wildfire Susceptibility Mapping Using MODIS Active Fire Pixels: An In-Depth Interpretation of Contributing Factors in Izmir, Türkiye. *Remote Sens.*, 16, 2842. https://doi.org/10.3390/rs16152842

[55] Taşkaya, S., Wu, D., Kurt, M., Liao, Y., Xu, J., Liao, W. (2024). Exploring the Application of Building Information Modeling (BIM) in Town Planning: Key Roles in the Relationship Between Buildings and Parcels, *International Journal of Computational and Experimental Science and Engineering*, Vol. 10-No.4(2024) pp. -701-717. https://doi.org/10.22399/ijcesen.459