Journal of Agricultural Sciences (Tarim Bilimleri Dergisi) 2025, 31 (2) : 538 – 557 DOI: 10.15832/ankutbd.1535974

Journal of Agricultural Sciences
(Tarim Bilimleri Dergisi)

J Agr Sci-Tarim Bili
e-ISSN: 2148-9297
jas.ankara.edu.tr

JAS

Comparison of Logistic Regression, Frequency Ratio, Weight of Evidence and Shannon's Entropy Models in Erosion Susceptibility Analysis in Bingöl (Türkiye) with GIS

Orhan İnik[a*], Mustafa Utlu[b]

[a]*Department of Soil Science and Plant Nutrition, Bingöl University, Bingöl, TÜRKIYE*

[b]*Department of Geography Science and Letter Faculty, Mehmet Akif Ersoy University, Burdur, TÜRKIYE*

ARTICLE INFO

Research Article

Corresponding Author: Orhan İnik, E-mail: oinik@bingol.edu.tr

Received: 20 August 2024 / Revised: 21 December 2024 / Accepted: 23 December 2024 / Online: 25 March 2025

Cite this article

İnik O, Utlu M (2025). Comparison of Logistic Regression, Frequency Ratio, Weight of Evidence and Shannon's Entropy Models in Erosion Susceptibility Analysis in Bingöl (Türkiye) with GIS. *Journal of Agricultural Sciences (Tarim Bilimleri Dergisi),* 31(2):538-557. DOI: 10.15832/ankutbd.1535974

ABSTRACT

Soil erosion is one of the most important and critical processes occurring in Türkiye, as in all parts of the world. It is of great importance to understand the processes that occur as soil erosion continues. The aim of this study is to determine the erosion susceptibility occurring in the Çapakçur Stream basin, one of the important erosion areas of Türkiye. In the study, erosion susceptibility analysis was carried out using 4 different methods Shannon Entropy (SE), Logistic Regression (LR), Frequency Ratio (FR) and Weight of Evidence (WoE) that are effectively used today in erosion susceptibility analysis and determination of critical areas in terms of erosion, and 19 conditioning factors based on these methods. Analysis Results Model performances were evaluated using Receiver Operating Characteristic (ROC) and Area under the Curve (AUC) values based on a dataset consisting of 840 training (70%) and 360 testing (30%) points. According to result of the AUC values show that Logistic regression seems to perform well on both training (AUC= 94.7%) and validating datasets (AUC=93.5%). On the other hand, Weight of Evidence training (AUC= 93.5%) and testing datasets (AUC= 91.4%), Frequency Ratio training (AUC= 93.5%) and testing datasets (AUC=92.4%) of the Weight of Evidence result show that AUC and ROC values similar to Logistic Regression result, but slightly lower than Logistic Regression. Additionally, Shannon Entropy shows that it performs lower than other methods on both training (AUC= 55.7%) and testing datasets (AUC= 56.3%). Conducting analyses based on these methods, especially in erosion susceptibility studies, will facilitate both planning and the accuracy of the results obtained.

Keywords: Erosion susceptibility, Logistic regression (LR), Weight of Evidence (WoE), Frequency Ratio (FR), Shannon's Entrophy (SE).

1. Introduction

Soil plays a crucial role in providing essential nutrients for the nourishment of land covers and types (such as forests, grasslands, and agriculture), controlling the emission rates of greenhouse gases, regulating Earth's temperature, retaining and storing water, preventing droughts, floods, and inundations in basins, and serving as a natural purification environment in terms of pollution. However, since the transition of humanity from the Neolithic period (pre-9000 BCE) to settled societies, various civilizations, from ancient times to the present, have exerted significant pressures on natural resources, especially land. Particularly in modern times, the rapid increase in population has led to a surge in demand for and pressure on natural resources, challenging the environment's self-renewal capacity. Factors such as unhealthy industrialization, unplanned urbanization, improper land use, increased pollution, decrease or extinction of species, and climate change, resulting from this population growth, endanger the sustainability of ecosystems due to biodiversity loss. These pressures, combined with natural factors, lead to one of the biggest problems, soil erosion, and cause land degradation by creating an irreversible risk of desertification (Valentin et al. 2005; Dengiz et al. 2014; Zhuang et al. 2015; Saha et al. 2020). Therefore, soil erosion is a major and critical environmental issue that poses a serious and irreversible threat to agricultural productivity and long-term ecosystem stability wherever it occurs globally, impacting the entire world (Chalise et al. 2019; Mohammed et al. 2020; Bag et al. 2022).

One of the most significant environmental problems encountered when soils and lands are not sustainably managed is the risk of desertification. Land degradation and drought, in short desertification, directly or indirectly affect the lives of approximately 1.2 billion people worldwide. The main reasons for this are extreme changes in climate events and the adverse effects of human activities. About 6% of the world's soils are severely desertified, and approximately 29% are at risk of desertification. Desertification has affected all regions of the globe to some extent, with particularly significant impacts felt in South America, Asia, and Africa. Each year, in addition to approximately 6 million hectares of land that become desertified, an additional 21 million hectares become unusable due to the spread of desertification (UNCCD 2016; Dengiz et al. 2020; İnik 2022; İnik 2023).

Türkiye is among the countries experiencing significant levels of soil erosion mainly due to its topography, climate, soil conditions, and anthropogenic factors, with erosion problems observed in only 13.86% of its land area according to official data (Berberoğlu et al. 2020). Land areas experiencing severe and very severe erosion make up 58.74% of the total landmass. Water erosion, which is the most common type of erosion both in Türkiye and globally, is the primary land degradation problem in our country, affecting 57.15 million hectares of land (Anonymous, 1987; Anonymous, 1998). The susceptibility of Turkish soils to both erosion and desertification is closely related to its geographical location, climate, topography, and soil structure. Soil erosion susceptibility, which we can define as the resistance of soils to erosive forces, varies under changing conditions such as rainfall intensity, slope steepness, changes in soil structure, and hydraulic properties. This situation makes it evident that desertification will have a more pronounced impact in the future under Türkiye conditions (İDEP 2012; Saygın 2013; Karagöz et al. 2015).

In recent years, innovative methods such as Geographic Information Systems (GIS) and machine learning (artificial intelligence algorithms) have been used for the identification of erosion-sensitive areas, resulting in more accurate and successful outcomes (Chakrabortty et al. 2020). Particularly, determining soil erosion susceptibility is of critical importance for implementing measures against erosion. In parallel with these advancements, numerous GIS-based soil erosion models have been developed from the 1990s to the present (Danacıoğlu & Tağıl 2017). In this context, soil erosion susceptibility studies also aim to identify potential areas at risk of soil erosion, particularly within specific sites and river basins. These studies generally focus on areas where erosion is likely to occur, and based on the results of the obtained susceptibility model, potential risk classifications are made. This classification is of great importance for determining measures to be taken for erosion sites within the high-risk category. For example, Bouamrane et al. (2024), examined soil erosion susceptibility maps in the Medjerda basin in North Africa. In this context, they used four different models: Deep Learning Neural Network-AHP (DLNN-AHP), Frequency Ratio-AHP (FR-AHP), Monte Carlo-AHP (MC-AHP), and Fuzzy AHP (F-AHP). They used eight different triggering factors, and the study identified that the distance to the river and rainfall erosivity factor had the greatest impact on soil erosion. Ait Neceur, et al. (2024), examined gully erosion susceptibility mappings in the N'fis River Basin using different machine learning algorithms. A total of 434 inventory data points were used for modeling, with 70% as training data and 30% as test data. The model accuracies were evaluated using the ROC curve. As a result, drainage density, slope, and NDVI were found to be the most influential factors in the field.

Therefore, recent studies on erosion susceptibility have shown a noticeable increase in research where parameter groups and analysis methods are determined by researchers, and independent models are used, alongside ready-made erosion models. Looking at recent studies on erosion susceptibility analysis using GIS technologies in the literature, the majority of them utilize statistical models (Akgün, 2007; Kheir et al. 2008; Conforti et al. 2011; Ogbonna et al. 2011). In this regard, there has been a rapid increase in studies focusing on soil erosion modelling approaches using methods such as frequency ratio (FR), logistic regression (LR), weight of evidence model (WoE), and Shannon entropy (SE), deep learning, machine learning etc.,

This study aims to conduct a soil erosion susceptibility analysis in the Çapakçur Creek basin, which is part of the Murat River Sub-basins and one of the significant soil erosion areas in Türkiye (Avcıoğlu et al. 2022). The purpose is to understand the erosion situation in this basin. For this purpose, four different methods, namely SE, LR, FR, and WoE, were employed to perform susceptibility analyses.

2. Material and Methods

2.1. Description of study site

The Çapakçur Stream basin is located within the borders of Bingöl province, in the Eastern Anatolia Region of Türkiye. It is situated between the latitudes of 38°48' - 38°55' and the longitudes of 40°17' - 40°32', with a basin area of 113.4 km². Generally, it extends in an east-west direction, with a maximum basin length of 21.6 km and a maximum basin width of 9.7 km. The basin has elevations ranging from a minimum of 1044 m to a maximum of 2505 m, with an average elevation of 1735 m, and it is characterized by a highly rugged topography. While the middle and upper parts of the basin have steep relief, only the area towards the city centre of Bingöl represents a relatively flat and plain topography. However, these areas also contain slopes with quite high gradients, and the average slope throughout the basin is 22.5° (Figure 1).

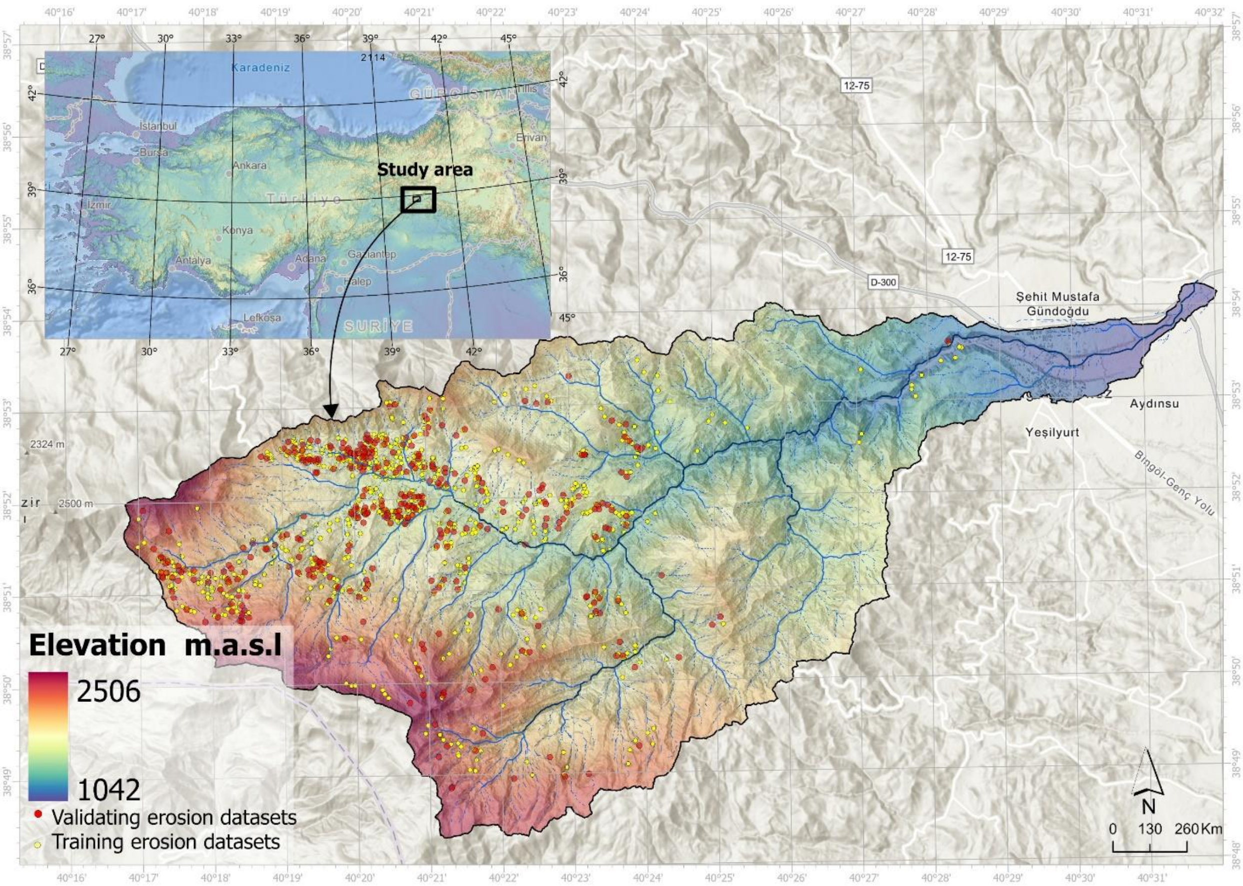

Figure 1- Location of the study area

According to the Koppen-Geiger climate classification, the main climate type is "D", representing a cold and humid temperate climate type (continental climate) in the winter seasons and is the second most common climate type seen in Türkiye (Peel et al. 2007; Öztürk et al. 2017). The sub-climate type is represented as "Dsa". The annual total precipitation in this region is 949 mm, with 117 days of snowfall and a snow cover duration of 76 days (İnik et al. 2022) and the average temperature values fall below 0°C during the winter seasons while exceeding 20 °C during the summer season (Öztürk et al. 2017). Dense erosion observed in the field is strongly controlled not only by climatological characteristics but also by geological and geomorphological features. Therefore, erosion is observed, particularly on steep slopes in the study area (Figure 2). Especially with heavy rainfall, erosion development accelerates in areas where surface runoff occurs, particularly in river valleys with flash floods. According to the general lithological characteristics of the basin, two different basic lithological units are observed in the field. In the lower reaches of the basin, a small portion consists of Quaternary-aged alluviums and volcanic units. One of the intensively observed units consists of sandstone, mudstone, limestone shelf, and sedimentary rocks spreading in areas where erosion events actively occur, while the other unit consists of upper Miocene-Pliocene-aged volcanic rocks, conglomerates, and continental units.

Figure 2- UAV images of the erosion areas in the Çapakçur Stream basin

2.2. Methodology

Geographic information systems and remote sensing technologies were used in this work, and Microsoft Excel and SPSS were used for statistical computations. The general flowchart used in the study is provided in Figure 3. The present study was carried out in the following main steps.

- Firstly, the detection and digitization of erosion areas within the basin boundaries,
- (ii) followed by subsetting the generated inventory data by randomly selecting 70% as training data and 30% as test (validation) data,
- the selection of condition factors and their reclassification and preparation,
- the implementation of LR, FR, WoE, and SE methods,
- the creation of susceptibility models for these methods,
- Testing the performance of erosion susceptibility models using the area under the receiver operating characteristic (AUROC).

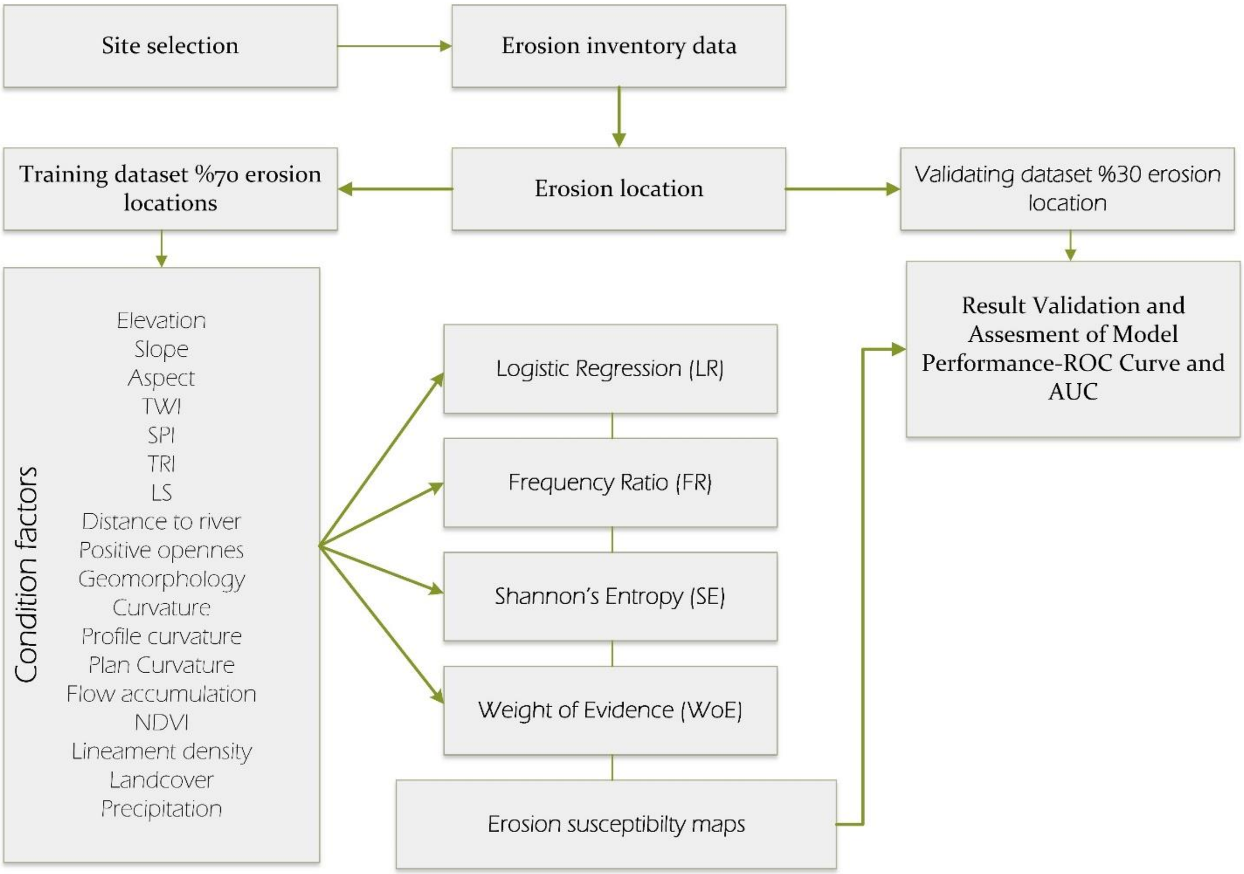

Figure 3- General flowchart of the study

In order to complete this study, 5m resolution DSM data, referred to as "Level-0 DSM5", generated from stereo aerial photographs through automatic matching, have been obtained from the General Directorate of Maps (HGK). From this surface model data, 14 different condition factor datasets have been produced. Additionally, Sentinel-2 satellite imagery with a resolution of 10m has been utilized for NDVI and lineaments analyses, precipitation characteristics from global grid data, and land cover data from global 10m resolution land cover data have been used. In this context, detailed properties of the data used, obtained from different sources and with different resolutions, are presented in Table 1.

Table 1- Details about the data was used in the study

	Data	*Source*	*Type*	*Resolution*	*Product*	*Software*
1	Digital surface model (DSM)	https://www.harita.gov.tr/	Grid	5 m	Slope, Geomorphology, TRI, TWI, LS, Aspect, Curvature (plan, profile), SPI, Positive opennes, Flow acc, Dist to river	ArcGIS Pro SAGA GIS
2	Landcover	https://esa-worldcover.org/en	Grid	10 m	Landcover	ArcGIS Pro
3	Satellite Image	https://scihub.copernicus.eu/Sentinel-2	Grid	10 m	NDVI, Lineaments	ArcGIS Pro
4	Road	https://overpass-turbo.eu/	Vector		Distance to road	ArcGIS Pro
6	Precipitation	Fijk and Hijmans, 2017	Grid	1 km	Precipitation	ArcGIS Pro
7	Geomorphology	https://saga-gis.sourceforge.io/saga_tool_doc/7.3.0/ta_lighting_8.html	Grid	5 m	Geomorphologic unit	SAGA GIS

2.3. Erosion inventory mapping

Susceptibility models are conducted for the purpose of the likelihood of past events recurring in the future. Thus, it is important for making rapid, on-site, and accurate decisions through the identification of potential areas. Therefore, inventory data collected from areas where the erosion have occurred is necessary for susceptibility modelling. Additionally, inventory data plays a critical role in the creation of statistically different susceptibility models and in measuring the performance of the resulting model outcomes. (Choubin et al. 2019). It particularly contributes to the accuracy, reliability, and effectiveness of susceptibility models. This facilitates the comparison of different modelling results for susceptibility models created for study areas. In this study, inventory data were used to determine the influence of each variable on erosion dependent on four different statistical methods for erosion susceptibility modelling (Figure 1). Fieldwork, drone imagery, and Google Earth were utilized to obtain the inventory data. A total of 1200 inventory point data were collected from various points in the study area where erosion occurred. Of these, 70% were used as training data and 30% as test (validating) data (Conforti et al. 2011; Gayen & Saha 2017; Hembram et al. 2019). In the ArcGIS Pro environment, subset features were randomly determined in the study area, resulting in 840 training data points and 360 test data points. The training inventory data were used within the scope of the model creation to learn the relationship between each factor used and the erosion status, while the test data were used to measure the accuracy of the created susceptibility models.

2.4. Erosion conditioning factors

In susceptibility modelling to be conducted in natural disaster research, there are many triggering and controlling factors involved in the occurrence of disasters. Therefore, in the creation of erosion susceptibility models, the most important stage is the selection of these data, as it will significantly impact the quality of the study and the accuracy of the results (Rahmati et al. 2017; Garosi et al. 2018). In erosion susceptibility studies, there are many factors that influence the occurrence, development, and progression of erosion. In this context, the selection of controlling factors has been made taking into account previous studies. However, there is no fixed controlling factor in erosion susceptibility studies (Conoscenti et al. 2013). Therefore, this study has selected 19 factors to improve the identification of erosion susceptibilities with enhanced accuracy and quality, as determined by the researchers (Rahmati et al. 2017; Arabemeri et al. 2020, Lei et al. 2020; Lana et al. 2022). Triggering factors are shaped by climatic, geomorphological, anthropogenic, and geological characteristics. These include elevation, slope, aspect, Normalized Difference Vegetation Index (NDVI), Topographic Wetness Index (TWI), Stream Power Index (SPI), Topographic Ruggedness Index (TRI), slope length (LS), distance to river, distance to road, land cover, lineament density, positive openness, geomorphology, curvature, plan curvature, profile curvature, flow accumulation, and precipitation data (Figure 4). Within the scope of the study, triggering factors were evaluated by classifying them into 5 classes according to the natural breaks method (Jenks, 1967).

Elevation, significantly influences erosion, especially in rill erosion processes. Higher elevations often lead to steeper slopes, which increase surface runoff and erosion potential (Conoscenti et al. 2008, Zhu et al. 2014; Zabihi et al. 2018; Zabihi et al. 2019). Accordingly, the elevation data for the Çapakçur Stream basin ranges from 1042 to 2506 m, with values divided into 5 classes as follows: 1042–1363, 1363–1621, 1621–1680, 1840–2092, and 2092–2506 m (Figure 4).

Slope, plays a key role in determining the extent of surface runoff, which directly affects erosion rates. Steeper slopes increase the potential for soil and sediment transport, making these areas more susceptible to erosion (Dramis and Gentili, 1977; Valentin et al. 2005; Güney, 2018). The slope values range from 0 to 76.1°, classified into 5 categories as follows: 0–15.2, 15.2–30.5, 30.5–45.7, 45.7–60.9, and 60.9–76.1 (Figure 4).

Aspect, one of the most important factors to consider when evaluating erosion susceptibility processes in a specific area or watershed is aspect (Carrara et al. 1991; Maharaj 1993; Guzzetti et al. 1999; Nagarajan et al. 2000; Güney 2018). Aspect is expressed with values ranging from 0 to 360 degrees in a clockwise direction. The aspect of a slope directly or indirectly influences erosion processes as it controls various climatic characteristics and vegetation cover (Dai et al. 2001; Çevik & Topal 2003; Pulice et al. 2009; İnik 2023). Aspect characteristics were evaluated and classified into 9 classes based on intermediate and main directions (Figure 4).

Normalized Difference Vegetation Index (NDVI), is the most commonly used method for analysing information about vegetation using GIS data to measure the amount of vegetation cover in an area. It serves as an indicator of green biomass in the area. NDVI is calculated using Equation 1 (Pettorelli et al. 2005):

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{1}$$

NIR represents the Near-Infrared band reflectance, RED represents the Red band reflectance. The NDVI values, analyzed in 5 classes, have minimum and maximum values ranging from 0.078 to 0.606. Accordingly, the NDVI values are classified into 5 categories as follows: 0.078–0.125, 0.125–0.203, 0.203–0.294, 0.294–0.405, and 0.405–0.606 (Figure 4).

Topographic wetness index (TWI), is generally defined as the influence of topography on the location and extent of areas where surface water will flow, and it is calculated according to Equation 2.

$$TWI = In\left(\frac{A_s}{Tan\beta}\right) \tag{2}$$

Here, A_s represents the contributing area to the cell, and β represents the slope gradient (Beven & Kirkby 1979). It provides clues about the degree of soil saturation or the movement of water within and on the surface of the soil, depending on the topography. This is because soil moisture content directly affects surface runoff, infiltration of water, ponding and other situations. This demonstrates the impact of TWI on erosion processes and it has been generated as a parameter for use in erosion susceptibility (Conforti et al. 2011; Sharma & Pandey, 2022). Accordingly, the TWI values range from 0.293 to 24.117 and are classified into 5 categories as follows: 0.293–3.188, 3.188–4.963, 4.936–7.392, 7.392–11.036, and 11.036–24.117 (Figure 4).

Stream Power Index (SPI), generated by flowing water or surface runoff on a particular slope directly affects erosion. SPI represents the power index in this context and is one of the key factors controlling erosion processes (Güney, 2018). Additionally, areas where high stream power index values are observed indicate a significant erosion potential, as they have the potential energy to transport sediments, in other words, they indicate locations that could serve as sediment sources (Kakembo et al. 2009). SPI is determined according to Equation 3.

$$SPI = In(As \times tan\beta) \tag{3}$$

Here, A_s represents the contributing area to the cell, and β represents the slope gradient (Nikolova 2022).

Topographic Ruggedness Index (TRI) is used to represent the amount of elevation difference between adjacent cells of a DEM. This scanning function template is used to create a visual representation of TRI with your elevation data. For example, it is assumed that 0-80 m represents a flat terrain surface, 81-116 m represents a nearly flat surface, 117-161m represents a slightly rugged surface, or 959-4367m represents an extremely rugged surface (Różycka et al. 2017; Habib, 2021; Trevisani et al. 2023). The SPI values range from -6.907 to 15.398 and are classified into 5 categories as follows: -6.907 to -3.671, -3.671 to 0.178, 0.178 to 2.189, 2.189 to 5.076, and 5.076 to 15.398 (Figure 4).

The LS factor, is a parameter used to measure soil erosion rates in erosion prediction models such as USLE and RUSLE. It controls surface flow velocity and is considered one of the most important factors for sediment transport (Haan et al. 1994). The technique for calculating the LS factor is provided by Moore and Burch (1986) as follows (Equation 4):

$$LS = \left(\text{Flow accumulation} \times \frac{\text{A}}{22.13}\right)^{0.4} \times sin(\frac{Slope}{0.0896})^{1.3} \tag{4}$$

The LS factor ranges from 0 to 807.97. Accordingly, it is classified into 5 categories as follows: 0–6.377, 6.377–25.348, 25.348–57.033, 57.033–129.90, and 129.90–807.97.

Distance to river, one of the important factors used to assess erosion susceptibility is the distance to the river. This factor is crucial for understanding what is more prone to erosion. Distance to river refers to the distance of an area from the river or streambed. (Arabemeri et al. 2020). Accordingly, the Distance to River ranges from a minimum of 1358 m to a maximum of 18607 m, with 5 classes as follows: 1358–6363, 6363–8663, 8664–10490, 10490–12519, and 12519–18607 (Figure 4).

Distance to road, roads have a negative impact on the sustainability of areas where surface runoff may be suitable for channels. Therefore, determining the distance to the road factor is important in identifying erosion susceptibility maps (Nkonge et al. 2023). In the study, in ArcGIS, the distance from each scanning cell to the road section (meters) was calculated using the Euclidean distance tool. Accordingly, the Distance to road ranges from a minimum of 0.82 m to a maximum of 11601 m, with 5 classes as follows: 0.82, 0.82-2275, 2275-3821, 3822-5960 and 5960-11601 (Figure 4).

Landcover, has a significant impact on geomorphological processes on slopes. Bare areas are generally more susceptible to erosion. The presence of vegetation reduces erosion susceptibility due to its ability to reduce the erosive effect of surface runoff (Anabalagan, 1992; Dai et al. 2001; Çevik & Topal, 2003; Conforti et al. 2011). Therefore, the land use and land cover of the basin have been emphasized as one of the geographical factors affecting erosion susceptibility. Accordingly, there are 7 different land cover types in the study area. These are tree cover, shrubland, bare/scarce vegetation, permanent water bodies, grassland, and cropland (Figure 4).

Lineament density, Surface lines represent weak areas with high permeability and low resistance. The distance to linearity is an important influencing factor for erosion development. This is because it represents a weak surface characterized by heavily fractured rocks in an area. Surface lines also promote soil degradation (Foumelis et al. 2004). The lineament density ranges from

0 to 5046. Accordingly, it is classified into 5 categories as follows: 0–0.396, 0.396–1069, 1069–1781, 1781–2652, and 2652–5046 (Figure 4).

Positive opennes (Po), is important for identifying narrow and deep valleys and determining convex units, especially those affecting erosional processes, within these valleys (Doneus 2013). The Po value ranges from 0.424 to 1687 and is classified into 5 categories as follows: 0.424–1127, 1127–1231, 1231–1315, 1315–1404, and 1404–1687 (Figure 4).

Precipitation, one of the fundamental factors that influences soil erosion, affecting surface runoff and causing increases or decreases in soil erosion and loss, is slope (Zhao et al. 2022)Global grid data was used in generating the rainfall dataset (Fick & Hijmans 2017). The study area has rainfall values ranging from 765 to 1379 mm. Accordingly, the precipitation data, classified into 5 categories, has values ranging from 76–888, 888–1011, 1011–1134, 1134–1256, and 1256–1379 mm (Figure 4).

Geomorphology, in erosion susceptibility mapping, it is important to determine the geomorphons representing the classification of land parcels and category forms. These are generated from a digital elevation model and utilized in the study (Stepinski & Jasiewicz, 2011; Jasiewicz & Stepinski, 2013)**.** According to the geomorphological data, the study area has the following morphological units: flat, summit, ridge, shoulder, spur, slope, hollow, footslope, valley, and depression (Figure 4).

Plan curvature, geomorphological and land morphology definitions, such as plan curvature analysis, are determined. Plan curvature refers to the effect on erosion formation, which is related to whether water moves away or towards during downstream flow. Therefore, the plan curvature layer is an important factor in triggering and developing erosion (Rahmati et al. 2016). The plan curvature values range from -113.78 to 108.39, and the classes are as follows: -113.78 to -10.97, -10.97 to -3.12, -3.12 to 2.971, 2.971 to 10.813, and 10.813 to 108.39 (Figure 4).

Profile curvature, which is believed to control the erosive force of the river, is particularly an important parameter for gully erosion (Conoscenti et al. 2013). The profile curvature values range from -98.501 to 119.071, and the classes are as follows: -98.501 to -12.375, -12.375 to -3.839, -3.839 to 2.136, 2.136 to 11,525, and 11.525 to 119.071 (Figure 4).

Curvature, units with concave landforms (negative curvature) generally affect water accumulation, while points with convex landforms (positive curvature) facilitate easy water flow and increase the likelihood of erosion (Ohlmacher, 2007). Curvature values range from -164.86 to 187.766. The first two classes contain negative values, while the last three classes have positive values (2.46 to 187.766) (Figure 4).

Flow accumulation, affects the direction of water flow and the likelihood of water accumulation It particularly influences the velocity of surface runoff. High flow accumulation results in increased surface runoff, causing soil erosion and carrying more sediment (Zhang et al. 2021). The flow accumulation values range from 0.001 to 6370.808 (Figure 4).

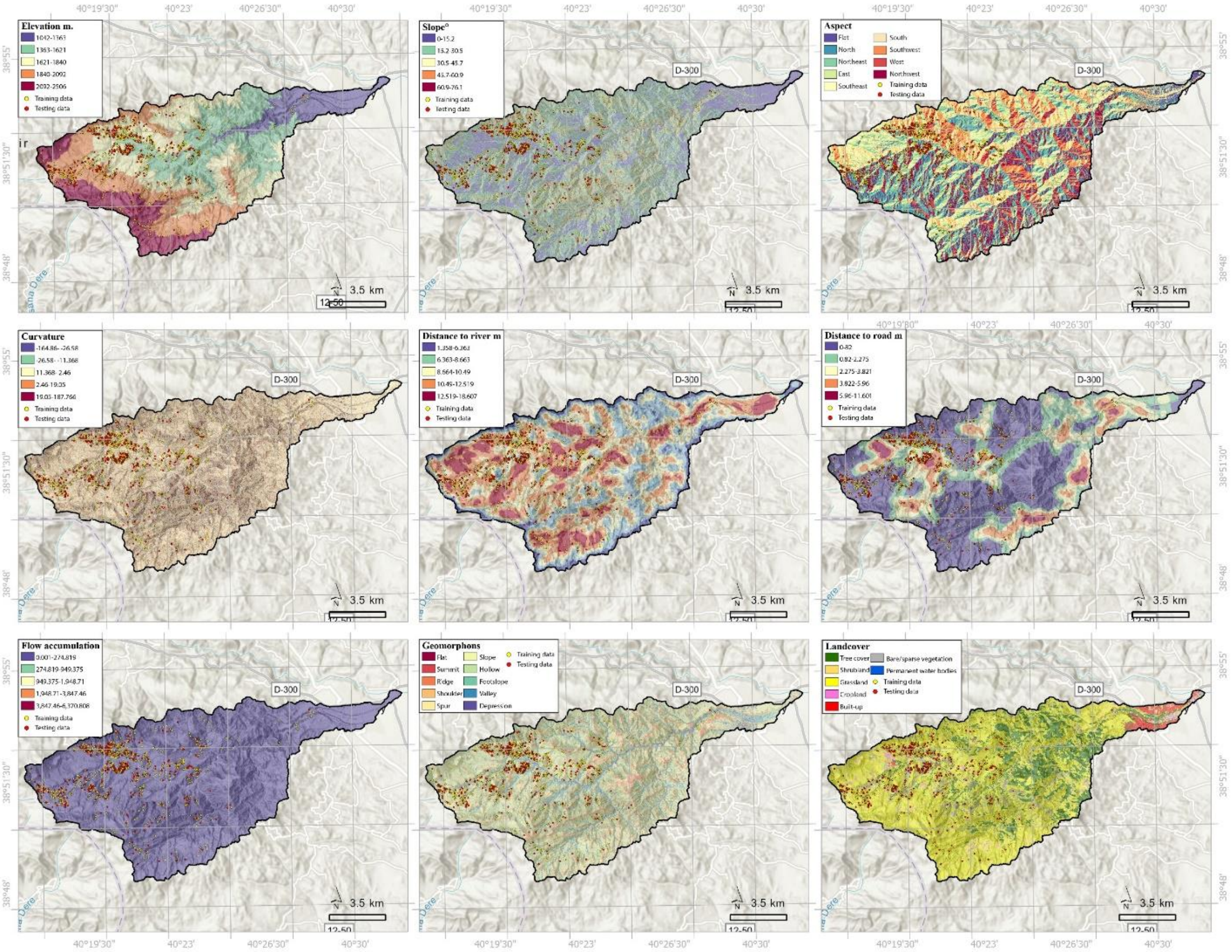

Figure 4- Condition factor of erosion susceptibility analysis

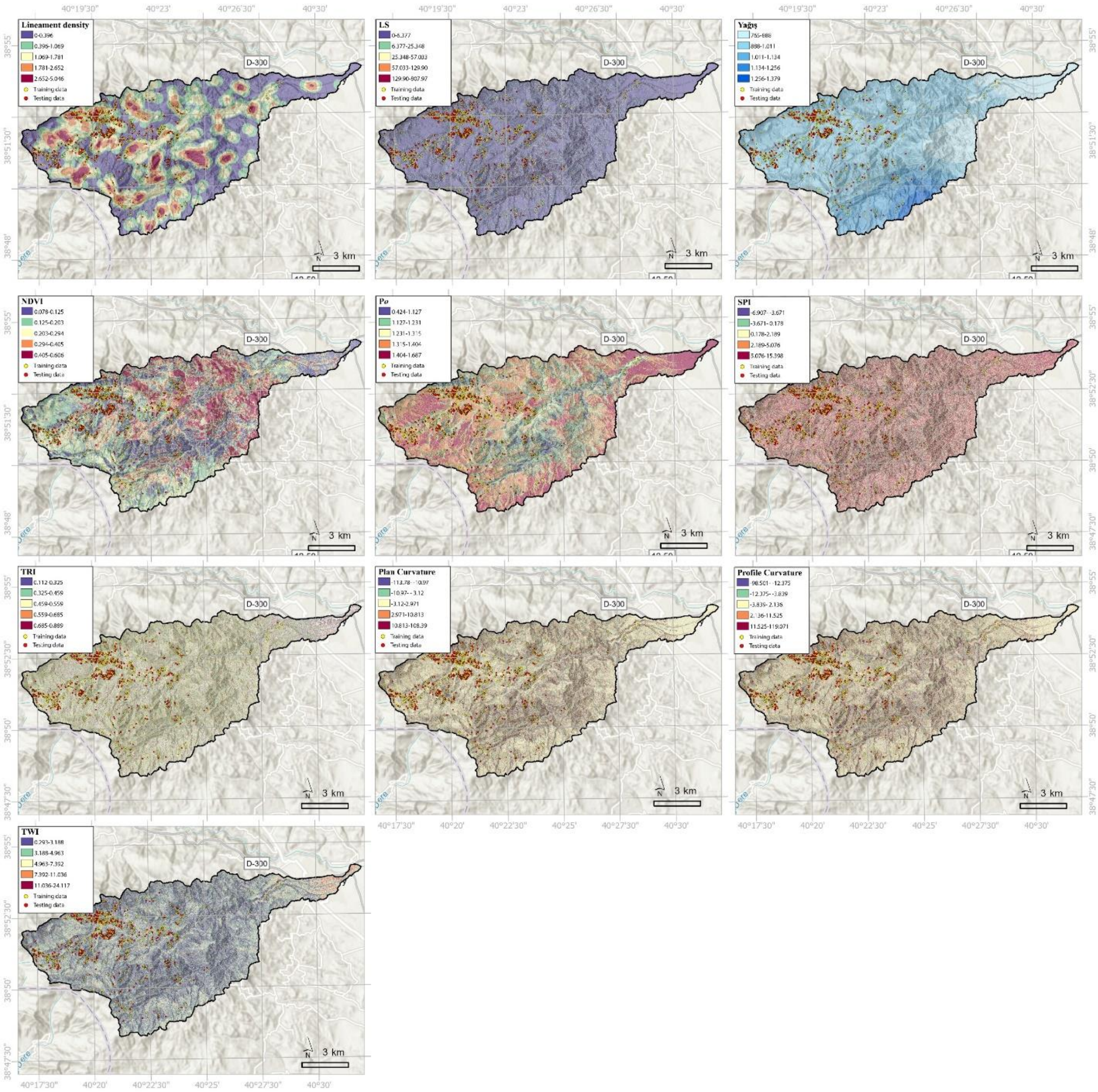

Figure 4- Condition factor of erosion susceptibility analysis (continued)

2.5. Erosion susceptibility mapping

Frequency Ratio (FR)

Frequency ratio (FR) is a commonly used method in erosion susceptibility assessment. FR is defined as the ratio of the probability of an event occurring, using all factors affecting a natural event (such as erosion) that occurred in the past, to the probability of not occurring (Bonham Carter, 1994; Dai and Lee 2002). Parameters used in erosion susceptibility analysis were correlated with erosion surfaces, which are considered evidence of severe erosion in the study area. The frequency ratio allows for the consideration of both the areas where erosion is severe and the extent of the areas covered by the parameters used in the research area. Equation 5 was utilized in the calculation of FO.

$$FO = \frac{X}{Y} \tag{5}$$

In Equation 1, X represents the percentage of erosion surface presence within each subclass of a parameter influencing erosion, while Y represents the percentage of each subclass of a parameter influencing erosion within that parameter. The values for X and Y were calculated based on Equations 6 and 7.

$$X = \frac{A}{B} \times 100 \tag{6}$$

$$Y = \frac{C}{D} \times 100 \tag{7}$$

Weight of Evidence (WoE)

The "Weight of Evidence (WoE)" model is used to calculate the Bayesian probability model more explicitly and which is commonly used in susceptibility modeling for spatial prediction (Akıncı et al. 2017; Kılıçoğlu, 2020). This model is a method within the Bayesian approach where conditional and unconditional probabilities are applied using sufficient data. Through this model, positive weights are assigned to predictions that anticipate erosion occurring in the studied area in the future, while negative weights are assigned to predictions that anticipate no erosion (Akıncı et al. 2017). The WoE model has been mathematically expressed by Van Westen et al. (2003) and Regmi et al. (2010). The determination of weights relies on the following equations.

$$W^{+} = \frac{\dfrac{A_1}{A_1 + A_2}}{\dfrac{A_3}{A_3 + A_4}} \tag{8}$$

$$W^{-} = \frac{\dfrac{A_2}{A_1 + A_2}}{\dfrac{A_4}{A_3 + A_4}} \tag{9}$$

$$C = W^{+} - W^{-} \tag{10}$$

In these equations; A_1: number of erosion cells in the sub-class, A_2: number of erosion cells outside the sub-class, A_3: number of non-erosion cells in the parameter subclass, A_4: number of non-erosion cells outside the sub-class, W+: Positive weight, W−: Negative weight, C: Represents the weight contrast. Positive weight (W+) is used to indicate the importance of the presence of the factor in terms of erosion formation. If this value is positive (+), the presence of the relevant factor is conducive to erosion formation; if negative (-), it is not conducive. Negative weight (W−) is used to evaluate the importance of the absence of the factor in terms of erosion formation. If this value is positive (+), the absence of the relevant factor is conducive to erosion formation; if negative (-), it is not conducive. C (weight contrast) reflects the spatial relationship of the prediction variable with erosion, indicating the difference between positive and negative weights. A positive value indicates a spatial relationship of the variable with erosion, while a negative value indicates no spatial relationship of the variable with erosion. A weight contrast value equal to zero indicates that the subcategory of the factor causing erosion is not significant (meaningful) for analysis (Van Westen et al. 2003; Neuhauser & Terhorst, 2007; Corsini et al. 2009, Akıncı et al. 2015; Akıncı et al. 2017; Kılıçoğlu, 2020).

Logistic Regression (LR)

There are three common types of multivariate statistical analysis methods widely used in the literature: multiple regression, logistic regression (LR), and discriminant analysis. In multivariate statistical analysis methods, factors that could cause erosion for a known land parcel are relatively examined, and the reasons for the occurrence of events are investigated. The land parcel examined through the analysis methods is based on data obtained by examining whether erosion has occurred or not (Akgün, 2007). The most significant limitation in multivariate statistical analysis studies is the long processing time due to the use of grid cells in the studies. However, the most important advantage of the method is that it is largely an objective method. Logistic regression is one of the most commonly used multivariate analysis methods in producing erosion susceptibility maps. LR analysis is based on a multivariate regression relationship between a dependent variable and multiple independent variables. The logistic regression method is defined by the following Equation (Equation 11):

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n \tag{11}$$

Shannon's Entropy (SE)

Shannon's entropy method is one of the techniques used in susceptibility analysis. This method is actively utilized in the implementation of susceptibility models for natural disasters such as floods, landslides, and erosion. (Sharma et al. 2012; Hembram et al. 2020; Islam et al. 2022; Utlu, 2023). It is generally based on the concept of entropy, which measures the level of uncertainty or randomness, and abnormality between causality and consequences (Lin 1991; Yufeng & Fengxiang 2009). It determines the maximum and minimum impact levels, or in other words, the entropy level of factors influencing the occurrence of hazards (Yulianto et al. 2020). The ratio of the high and low values of the measure affects the susceptibility level. A high value indicates a high probability situation, while a low value indicates a low probability situation (Al-Hinai & Abdalla 2021). Shannon entropy is computed using the formula as specified below (Equation 12):

$$Pd_{ij} = \frac{FR_{ij}}{\sum_{i=1}^{m_j} FR_{ij}} \tag{12}$$

Pd_{ij}, Represents the probability density, while, FR_{ij} and denotes the frequency rate in the given parameters. After calculating the probability density, the obtained values are used to calculate the entropy (Equation 13-14).

$$Ev_j = \sum_{i=1}^{m_j} Pd_{ij}\, log_2\, Pd_{ij}, j = 1, \ldots, n \tag{13}$$

$$Ev_{jmax} = log_{2\, m_j} \tag{14}$$

Ev_j ve Ev_{jmax} Entropy values, m_j represent the number of classes (Equation 15) in calculating the weights of factors used in erosion susceptibility assessment.

$$Ic_j = \left(Evj_{max} - \frac{Ev_j}{Ev_{jmax}} \right), I = (0,1)1j = 1, \ldots n \tag{15}$$

The formula is utilized, and in the formula Ic_j: represents the coefficient of the relevant layer (Equation 16),

$$Cw_j = I_j FR \tag{16}$$

Cw_j: It represents the weight value representing the entire relevant layer.

2.6. Model evaluation

In natural disaster studies, the ROC curve and AUC values method is widely used for assessing the model performances of susceptibility models generated to understand flood, landslide, rock fall, and erosion events. ROC and AUC method currently, most of the researchers are actively using this approach (Tehrany et al. 2013; Miao et al. 2023; Utlu, 2023; El Miloudi et al. 2024). Data from the susceptibility model's event inventory is required to measure model performances with accuracy and reliability. As a result, evaluating the results that are produced using the ROC and AUC approaches is simple. Plotting the true positive rate (TPR) and false positive rate (FPR) of binary classification models is carried out for assessing their efficacy using ROC and AUC metrics. Equation 17 displays the y-axis as TPR (True Positive Rate), and Equation 18 shows the x-axis as FPR (False Positive Rate).

$$y\ axis\ TPR = \frac{TP}{TP + FN} \tag{17}$$

$$x\ axis\ FPR = 1 - \frac{FP}{FP + TN} \tag{18}$$

AUC represents the area under the ROC curve and ranges from 0 to 1. Equation 19, (Amiri et al. 2019; Baiddah et al. 2023):

$$AUC = \frac{\sum TP + \sum TN}{P + N} \tag{19}$$

P: the total number of erosion data, N: the total number of data without erosion data (Baiddah et al. 2023).

3. Results

3.1. Multi-collinearity assessment

This evaluation indicates whether there is a linear relationship among multiple independent variables. Especially in susceptibility analyses, the importance of model accuracy and reliability plays a role, and examining and evaluating the correlation between independent variables is crucial (Graham 2003; Rahmati et al. 2017; Roy & Saha 2019; Wang et al. 2021). Thus, variables with high correlation can lead to unstable coefficient estimates and inaccurate predictions, while on the other hand, they can reduce the predictive accuracy of the model. Therefore, the independent variables to be considered should have a VIF (Variance Inflation Factor) value below 10 and a tolerance threshold value above 0.1. Because values that go outside of and beyond these boundaries signify the existence of problems with multicollinearity (Kelava et al. 2008; Arabameri et al. 2020). The VIF values of curvature, plan curvature, and profile curvature among the 19 distinct condition data exhibit strong association among independent variables, suggesting multicollinearity. As a result, 16 factor values were taken into account, and these 3 condition factors were not included in the analysis (Table 2).

Table 2- Multi-collinearity values of conditioning factors

No	*Factors*	*Tolerance*	*VIF*
1	Elevation	0.742	1.347
2	Distance to river	0.315	3.173
3	Distance to road	0.869	1.151
4	TWI	0.226	4.426
5	TRI	0.545	1.837
6	SPI	0.214	4.426
7	Slope	0.161	6.229
8	Geomorphology	0.503	1.989
9	Distance to lineament	0.922	1.085
10	LS	0.442	2.265
11	LU	0.393	2.547
12	NDVI	0.367	2.723
13	Plan curvature	0.020	51.109
14	Profile curvature	0.017	58.948
15	Curvature	0.006	167.909
16	Precipitation	0.279	3.587
17	Positive openness	0.243	4.112
18	Flow accumulation	0.854	1.170
19	Aspect	0.916	1.091

3.1. Erosion susceptibility modelling

Erosion susceptibility analysis has been conducted on the Çapakçur Creek basin using different statistical methods dependent on various algorithms. In this context, FR, LR, WoE, and SE methods were employed, and the results obtained has been classified into 5 classes using the natural breaks classification Jenks algorithm (Jenks, 1967). According to multicollinearity assessment, 16 out of the 19 selected factors were considered. Among the 16 factors used across 4 different methods, those with the highest impact on erosion in the Çapakçur Stream basin are as follows: for the WoE (Weights of Evidence) method, elevation and slope; for the LR (Logistic Regression) method, NDVI, land use, slope, and TWI (Topographic Wetness Index); for the FR (Frequency Ratio) method, NDVI, precipitation, flow accumulation, and elevation; and for the SE (Statistical Evaluation) method, the significant factors are elevation and precipitation data. These classes are very low, low, moderate, high, and very high. The results of the susceptibility model obtained by different methods are presented in Figure 5. According to result of the four methods, the areas with high and very high erosion susceptibility, show quite similar distributions except for the SE method, direct correlation with the geomorphological features of the relief. In the LR, FR, and WoE methods, areas with high and very high susceptibility are found in steep slopes and ridges with narrow and deep valleys, while in the SE method, areas with high susceptibility are observed in river valley bottoms and areas with low slope values close to flat relief. In addition, areas with very low, low, and moderate erosion susceptibility correspond to similar terrain characteristics in the LR, FR, and WoE methods, generally comprising low-lying flat areas and gently sloping terrains. However, in the SE method, these areas appear to be high-slope and steep terrain, contrasting with the high and very high erosion susceptibility areas identified by the three other methods (Figure 5).

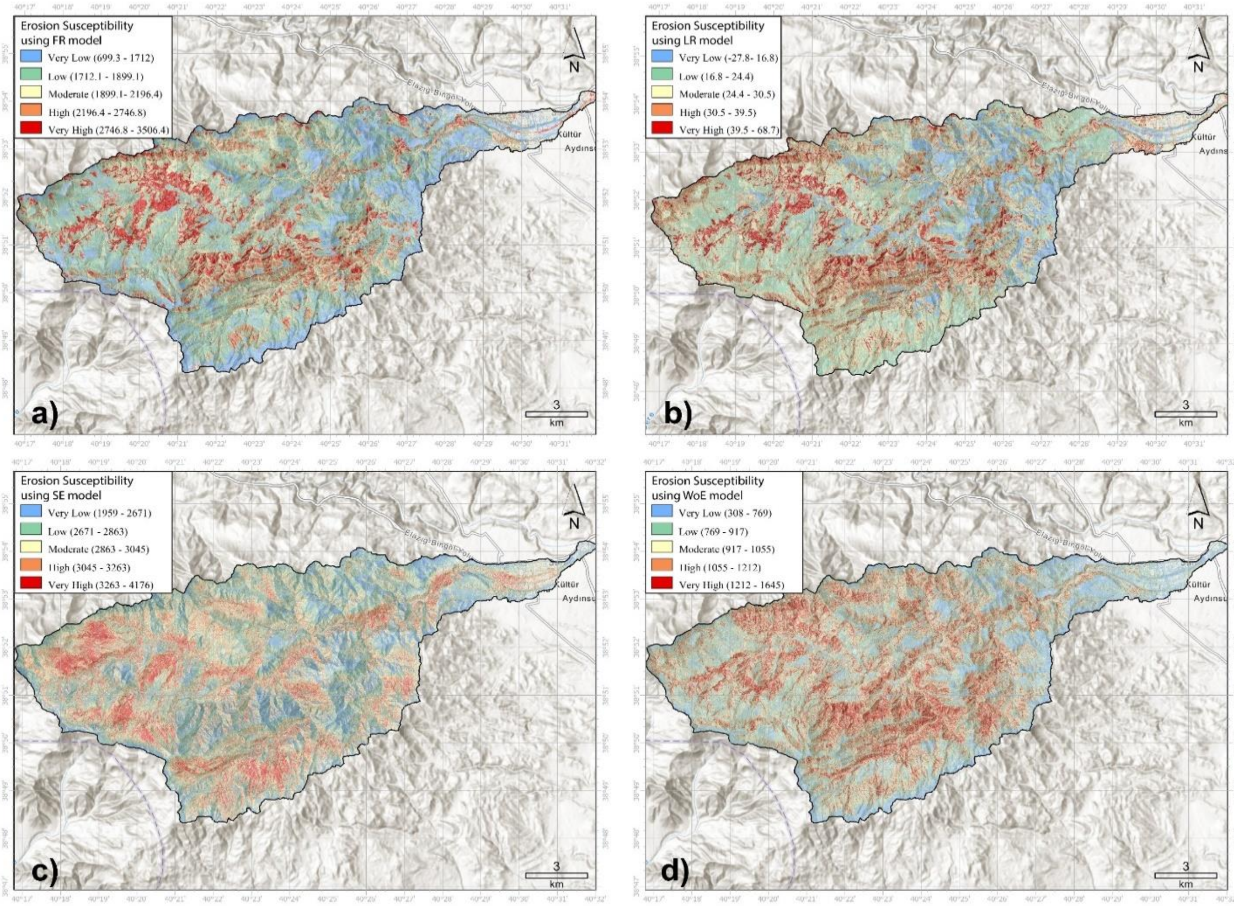

Figure 5- Erosion susceptibility models using different statistical methods a) FR b) LR c) SE d) WoE

Each method shows variations in the distribution of susceptibility classes, which could indicate differences in the susceptibility of the methods or the underlying assumptions they rely on. According to the spatial distributions of each susceptibility classes for different statistical model results are as follows for the LR method: very low class covers 12.2 km² (11%), low class covers 31 km² (27.9%), moderate class covers 37.2 km² (33.5%), high class covers 21.1 km² (19%), and very high class covers 9.5 km² (8.5%) (Figure 6). FR method: very low class covers 20.7 km² (18.7%), low class covers 39.6 km² (37.7%), moderate class covers 25.7 km² (23.1%), high class covers 15 km² (13.5%), and very high class covers 10 km² (9%) (Figure 6). WoE method: very low class covers 14.6 km² (13.1%), low class covers 28.3 km² (25.5%), moderate class covers 30.9 km² (27.8%), high class covers 25.4 km² (22.9%), and very high class covers 11.8 km² (10.7%) (Figure 6). SE method: very low class covers 16.1 km² (14.5%), low class covers 30.2 km² (27.2%), moderate class covers 32.9 km² (29.6%), high class covers 23.6 km² (21.2%), and very high class covers 8.3 km² (7.5%) (Figure 6).

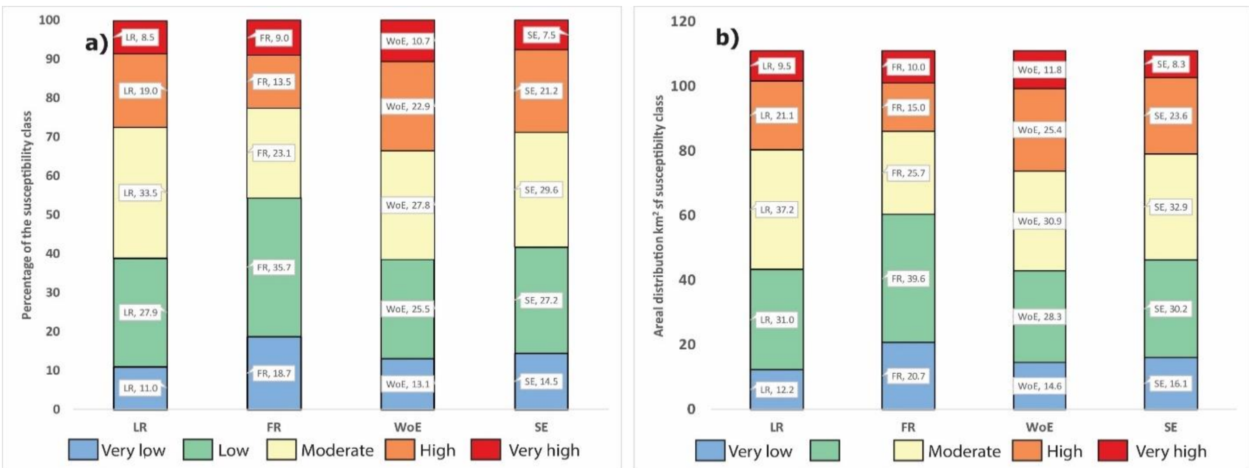

Figure 6- a) The percentage soil erosion susceptibility classes in different models (%), b) The area erosion susceptibility classes in different models (km^2)

3.2. Susceptibility model evaluation and comparison

The performance of the resulting erosion susceptibility models was assessed using the ROC curve and AUC (Du et al. 2017). The AUC value is commonly categorized as follows, and the resulting numbers offer a crucial indication of the model's validity and accuracy. A result falls into one of five categories: poor (0.5–0.6), fair (0.6–0.7), good (0.7–0.8), very good (0.8–0.9), and excellent (0.9–1.0) (Rahmati et al. 2016). As the ROC curve approaches 1, it indicates the presence of a prediction model with perfect accuracy, while moving away from 1 generally signifies a decrease in overall model accuracy (Nkonge et al. 2023). In this study, the performances of susceptibility models created based on different statistical methods were evaluated using both training and test datasets (Figure 7). According to result of the ROC and AUC values, Logistic regression seems to perform well on both training (AUC: 94.7) and validating datasets (AUC: 93.5), with slightly higher performance on the training set compared to the testing set. This indicates that the model might be slightly overfitting the training data, but the drop in performance from training to testing is not substantial and LR result correspond to “excellent” range, demonstrating high predictive accuracy and validity. Weight of Evidence performs consistently on both training (AUC: 93.5) and testing datasets (AUC 91.4), but it shows slightly lower performance compared to Logistic Regression, especially on the validating set. Frequency Ratio performs well, similar to Weight of Evidence and Logistic Regression, but slightly lower than Logistic Regression. It also exhibits consistent performance on both training (AUC: 93.5) and testing datasets (AUC: 92.4) also shows that “excellent” category. These values indicating that it is an effective model for erosion susceptibility prediction. Frequency Ration (FR) shows similar and close results to the WoE method. These values are high for both training and test data. Although it is lower compared to the LR method, it falls within the "excellent" category. In this context, the FR method Training AUC (93.5) and testing AUC (91.4) has yielded good results for the applicability of the soil erosion susceptibility model. Shannon's Entropy shows significantly lower performance compared to the other methods on both training (AUC: 55.7), testing datasets (AUC: 56.3) also indicated that “poor” category. These low values SE is not effectively capturing the patterns needed to predict erosion susceptibility. The AUC values close to 0.5 suggest that this model’s prediction is only slightly better than random guessing, confirming that it is not possible to use for this basin and areas. This suggests that the model built using Shannon's Entropy might not be capturing the underlying patterns effectively. Because, Shannon's Entropy may not be the most appropriate method for modeling erosion susceptibility in your specific study area. The entropy-based model might be too simplistic or fail to capture complex interactions among environmental factors like topography, land use, or soil properties that contribute to erosion. If the data does not exhibit enough "randomness" or variability in relation to erosion patterns, entropy may not be able to generate meaningful insights.

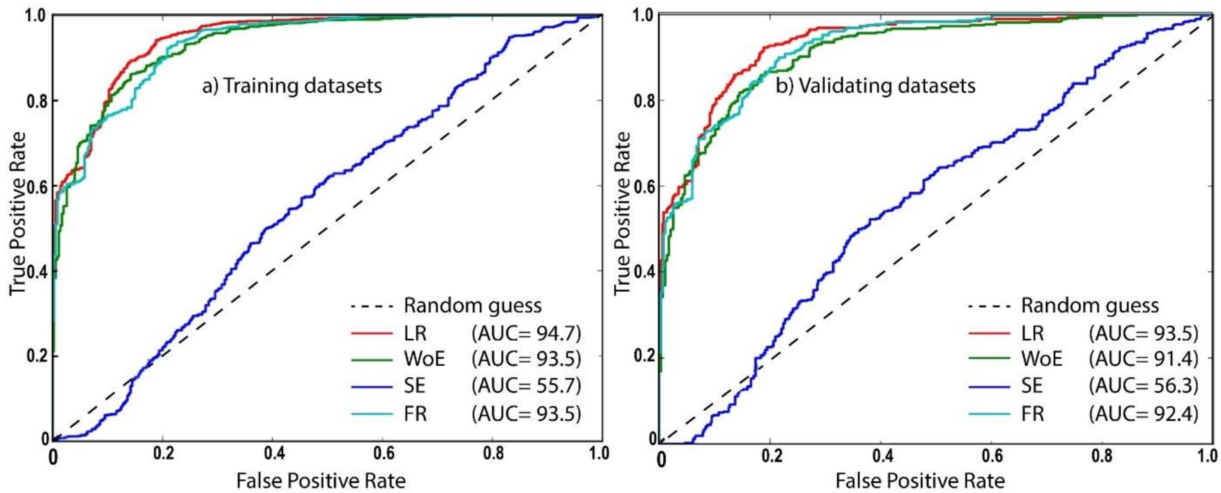

Figure 7- Prediction and success rates of the different susceptibility models based on a) training datasets b) validating datasets

4. Discussion

Due to environmental factors and human activities, the frequency of natural disasters in recent years has increased, and this can lead to loss of life, resources and property (He et al. 2012; Chen et al. 2017). The most common of these is soil erosion. Erosion events have directly and indirectly caused many negative problems, and they are quite diverse. These effects include increased sediment transport, damage to vegetation, soil degradation and the failure of surface water to flow into groundwater. As sediment transport increases and riverbeds fill, inevitable events such as floods and inundations occur. In addition, it results in water pollution and habitat destruction. Soil loss due to erosion restricts the habitat and root system of the vegetation (Sterk, 2003; Podhrazska et al. 2015; Duniway et al. 2019; Saxena, 2021). Therefore, in natural disaster studies, it is very important to take the necessary precautions and make critical planning in case past events recur in the future. Despite efforts to prevent these events, it is important to make inferences about possible future situations or to model them in order to prevent them effectively. Therefore, sensitivity models are developed to take necessary precautions and minimize possible hazards and risks and to reduce erosion and especially desertification risks (Morgan & Nearing, 2016; Batista et al. 2019). However, due to the various lithological, climatic and geomorphological characteristics of different regions, the suitability of these models varies. Therefore, it is important to test different methods and approaches in a specific area to determine the most suitable model.

In this study, sensitivity modelling was carried out to understand the advanced erosion status in the Çapakçur River basin. According to the literature and multicollinearity analysis, nineteen conditioning factors (elevation, distance to river, distance to road, TWI, TRI, SPI, slope, geomorphology, distance to lineament, LS, LU, NDVI, plan curvature, profile curvature, curvature, positive openness, precipitation, flow accumulation and aspect) (Table 2) and four different methods (LR, FR, WoE and SE) were used to determine the sensitivity and model performances were measured using ROC curves and AUC values to determine the most suitable model (Xu et al. 2012; Demir et al. 2022; Jaafari et al. 2014; Ding et al. 2017; Bhandari et al. 2024).

Logistic Regression shows the highest performance among the methods, both on training and testing datasets, indicating its robustness in capturing the relationship between the predictors and the target variable. LR is a widely used model for classification, especially for binary classification problems (Kleinbaum et al. 2002). Therefore, to obtain a better classifier, WoE and FR models were also included to determine the sensitivity (Chen et al. 2019). Weight of Evidence and Frequency Ratio also perform reasonably well, with consistent performance across training and testing datasets. Shannon's Entropy performs noticeably worse than the other approaches, indicating that it may not be appropriate for this specific purpose or that it may need to be implemented more precisely. In summary, Shannon's Entropy performs noticeably worse than Logistic Regression, which looks to be the best approach for this erosion susceptibility mapping problem. Other methods that perform well include Frequency Ratio and Weight of Evidence. The efficacy of Logistic Regression in mapping erosion susceptibility is highlighted by its good performance in capturing the connection between predictors and the target variable. Furthermore, the trustworthiness of Frequency Ratio and Weight of Evidence as alternative modeling methodologies is highlighted by their constant performance.

5. Conclusions

Many statistical methods are used in earth science studies, especially in performing sensitivity analyses using computer technologies and modern techniques. In addition to these, there are actively preferred and extremely popular methods. These methods include deep learning, machine learning, artificial intelligence and different techniques produced depending on them. Thus, it plays an important role in evaluating the results obtained from many methods used and taking the necessary planning and precautions (Baiddah et al. 2023).

In this study, erosion sensitivity analysis was conducted in the Çapakçur River basin, which is critical in terms of erosion. Logistic regression (LR), frequency ratio (FR), weight of evidence (WoE), and Shannon's entropy methods were used in conducting sensitivity analysis in the study area. Within the scope of sensitivity analysis, 19 methods were utilized, and it was decided to use 16 conditioning factors based on multicollinearity assessment method based on VIF and Tolerance values. These factors include topographic, climatic, anthropogenic, and environmental factors. Based on the results obtained in this study, it has been determined that the most efficient methods are Logistic Regression and Frequency Ratio, as well as the Weight of Evidence method. Dependent on these methods, erosion susceptibility models can be developed in river basins with similar lithological, geomorphological, and climatic characteristics. Consequently, measures and plans can be made to prevent the progression of existing erosion in both large and small-scale basins, based on the results of erosion susceptibility models.

Also, these erosion susceptibility models provide valuable tools for authorities and policymakers to effectively address erosion issues and mitigate associated negative impacts such as sediment transport, soil degradation, and habitat destruction. By identifying areas sensitive to erosion, authorities can prioritize targeted interventions and implement appropriate land management practices to minimize erosion risks. In the future, research could explore both the integration of advanced machine learning techniques and remote sensing data to improve the accuracy and predictive capabilities of erosion susceptibility models using high resolution topographic, climatic datasets. Furthermore, incorporating real-time monitoring data and climate projections can offer insights into evolving erosion patterns and support proactive erosion prevention strategies.

Ethics Statement: There is no study in this study that requires permission from the ethics committee.

Availability of data and materials: Data will be available on request.

Ethics approval: All authors have read, understood, and have complied as applicable with the statement on “Ethical responsibilities of Authors” as found in the Instructions for Authors. Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Competing interests: The authors declare no competing interests.

References

Ait Neceur H, Abdo H G, Igmoullan B, Namous M, Alshehri F & A Albanai J (2024). Implementation of random forest, adaptive boosting, and gradient boosting decision trees algorithms for gully erosion susceptibility mapping using remote sensing and GIS. Environmental Earth Sciences 83(3): 121 https://doi.org/10.1007/s12665-024-11424-5

Akgün A (2007). Ayvalık ve yakın çevresinin erozyon ve heyelan duyarlılığının Coğrafi Bilgi Sistemleri tabanlı incelenmesi. *Doktora Tezi, Dokuz Eylül Üni. Fen Bilimleri Ens*. İzmir.

Akıncı H, Özalp A Y & Kılıçer S T (2015). Assessment of landslide susceptibility in planned areas using geographic information systems and AHP method: Artvin Example. Journal of Natural Hazards and Environment 1(1-2): 40-53 (In Turkish)

Akıncı H, Doğan S & Kılıçoğlu C (2017). Landslide susceptibility mapping of Canik (Samsun) district using bayesian probability and frequency ratio models. *Selcuk University Journal of Engineering Science and Technology* 5(3): 283- 299 https://doi.org/10.15317/scitech.2017.89

Al-Hinai H & Abdalla R (2021). Mapping Coastal Flood Susceptible Areas Using Shannon's Entropy Model: The Case of Muscat Governorate, Oman. *ISPRS Int J Geo-Information* 10: 252. https://doi.org/10.3390/ijgi10040252

Amiri M, Pourghasemi H R, Ghanbarian G A & Afzali S F (2019). Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma* 340: 55–69. https://doi.org/10.1016/j.geoderma.2018.12.042

Anabalagan R (1992). Landslide hazard evaluation and zonation mapping in mountainous terrain. Eng Geol 32: 269–277. https://doi.org/10.1016/0013-7952(92)90053-2

Anonymous (1998). Management of Agricultural and Pasture Lands. National Environmental Action Plan. DTP, Ankara. (In Turkish)

Anonymous (1987). Türkiye General Soil Management Planning. Ministry of Agriculture, Forestry and Rural Affairs, General Directorate of Rural Services. Ankara. (In Turkish)

Arabameri A, Chen W, Loche M, Zhao X, Li Y, Lombardo L & Bui DT (2020). Comparison of machine learning models for gully erosion susceptibility mapping. Geoscience Frontiers, 11(5): 1609-1620. https://doi.org/10.1016/j.gsf.2019.11.009

Avcıoğlu A, Görüm T, Akbaş A, Moreno-de las Heras M, Yıldırım C & Yetemen Ö (2022). Regional distribution and characteristics of major badland landscapes in Turkey. Catena, 218, 106562. https://doi.org/10.1016/j.catena.2022.106562

Bag R, Mondal I, Dehbozorgi M, Bank S P, Das D N, Bandyopadhyay J & Nguyen X C (2022). Modelling and mapping of soil erosion susceptibility using machine learning in a tropical hot sub-humid environment. *Journal of Cleaner Production* 364: 132428. https://doi.org/10.1016/j.jclepro.2022.132428

Baiddah A, Krimissa S, Hajji S, Ismaili M, Abdelrahman K & El Bouzekraoui M (2023). Head-cut gully erosion susceptibility mapping in semi-arid region using machine learning methods: insight from the high atlas, Morocco. *Frontiers in Earth Science*, *11*. https://doi.org/10.3389/feart.2023.1184038

Batista P V, Davies J, Silva M L & Quinton J N (2019). On the evaluation of soil erosion models: Are we doing enough?. *Earth-Science Reviews* 197: 102898. https://doi.org/10.1016/j.earscirev.2019.102898

Beven K J & Kirkby M J (1979) A physically based, variable contributing area model ofbasin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrol Sci J 24(1):43–69. https://doi.org/10.1080/02626667909491834

Bhandari B P, Dhakal S & Tsou C Y (2024). Assessing the Prediction Accuracy of Frequency Ratio, Weight of Evidence, Shannon Entropy, and Information Value Methods for Landslide Susceptibility in the Siwalik Hills of Nepal. *Sustainability*, *16*(5), 2092. https://doi.org/10.3390/su16052092

Bonham Carter G F (1994). Geographic Information Systems for geoscientists, Modeling with GIS. *Pergamon Press*, Oxford.

Bouamrane A, Boutaghane H, Bouamrane A, Dahri N, Abida H, Saber M, Kantoush S A & Sumi T (2024). Soil erosion susceptibility prediction using ensemble hybrid models with multicriteria decision-making analysis: Case study of the Medjerda basin, northern Africa. International Journal of Sediment Research. https://doi.org/10.1016/j.ijsrc.2024.08.003

Carrara A, Cardinali M, Detti R, Guzzetti F, Pasqui V & Reichenbach P (1991). GIS techniques and statistical models in evaluating landslide hazard. Earth Surf Processes and Landforms. 16(5): 427-445. https://doi.org/10.1002/esp.3290160505

Çevik E & Topal T (2003). GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). *Environmental Geology* 44 (8): 949-962 https://doi.org/10.1007/s00254-003-0838-6

Chakrabortty R, Pal S C, Sahana M, Mondal A, Dou J, Pham B T & Yunus A P (2020). Soil erosion potential hotspot zone identification using machine learning and statistical approaches in eastern India. Natural hazards. https://doi.org/10.1007/s11069-020-04213-3

Chalise D, Kumar L, Spalevic V & Skataric G (2019). Estimation of sediment yield and maximum outflow using the IntErO model in the sarada river basin of Nepal. Water 11: 952. https://doi.org/10.3390/w11050952.

Chen W, Panahi M & Pourghasemi H R (2017). Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *Catena* 157: 310-324

Chen W, Sun Z & Han J (2019). Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. Applied sciences 9(1): 171. https://doi.org/10.3390/app9010171

Choubin B, Rahmati O, Tahmasebipour N, Feizizadeh B & Pourghasemi H R (2019). Application of fuzzy analytical network process model for analyzing the gully erosion susceptibility. In *Advances in Natural and Technological Hazards Research* (Vol. 48, pp. 105–125). Springer Netherlands. https://doi.org/10.1007/978-3-319-73383-8_5

Conforti M, Aucelli P P C, Robustelli G & Scarciglia F (2011). Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). *Natural Hazards* 56(3): 881–898. https://doi.org/10.1007/s11069-010-9598-2

Conoscenti C, Agnesi V, Angileri S, Cappadonia C, Rotigliano E & Märker M (2013). A GIS-based approach for gully erosion susceptibility modelling: a test in Sicily, Italy. *Environ Earth* 70: 1179–1195. https://doi.org/10.1007/s12665-012-2205-y

Conoscenti C, Di Maggio C & Rotigliano E (2008). Soil erosion susceptibility assessment and validation using a geostatistical multivariate approach: a test in Southern Sicily. Natural hazards 46: 287-305. https://doi.org/10.1007/s11069-007-9188-0

Corsini A, Cervi F & Ronchetti F (2009). Weight of evidence and artificial neural networks for potential groundwater spring mapping: an application to the Mt. Modino area (Northern Apennines, Italy). *Geomorphology*, 111(1-2), 79- 87). doi: 10.1016/j.geomorph.2008.03.015

Dai F C & Lee C F (2002). Landslide characteristics and slope instablity modelling using GIS, Lantau Island, Hong Kong. *Geomorphology* 42: 213-228

Dai F C, Lee C F, Li J & Xu Z W (2001). Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environmetal Geology* 40(3): 381-391

Danacıoğlu Ş & Tağıl Ş (2017). Assessment of Erosion Risk in Bakırçay Basin Using Rusle Model. *Journal of Balıkesir University Social Sciences Institute* 20(37) (in Turkish)

Demir Y, Meral A & Doğan Demir A (2022). Estimation of Soil Losses in Çapakcur Watershed (Bingol, Turkey) Using RUSLE Method and Comparison of Predicted Soil Losses with Sediment Yield. *Kahramanmaraş Sütçü İmam University Journal of Agriculture and Nature*, 25(Supplement Issue 2) 523–537. https://doi.org/10.18016/ksutarimdoga.vi.1059631 (In Turkish)

Dengiz O, İmamoğlu A, Saygın F, Göl C, Ediş S & Doğan A (2014). Soil erosion risk assessment of Inebolu basin with icona model. *Anatolian Journal of Agricultural Sciences* 29(2): 136-142. doi: 10.7161/anajas.2014.29.2.136-14 (In Turkish)

Dengiz O, Öztaş T, Haliloğlu M & Şahin K (2020). Balancing of Land Degradation. Turkish Agricultural Engineering IX.Technical Congress. 81-104 (In Turkish)

Ding Q, Chen W & Hong H (2017). Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. Geocarto international, 32(6): 619-639. https://doi.org/10.1080/10106049.2016.1165294

Doneus M (2013). Openness as visualization technique for interpretative mapping of airborne lidar derived digital terrain models. *Remote Sens* 5: 6427–6442. https://doi.org/10.3390/rs5126427

Dramis F & Gentili B (1977). Contributo allo studio delle acclività dei versanti nell'appennino umbro-marchigiano. *Studi Geologici Camerti* 3: 153-164

Du G L, Zhang Y S, Iqbal J, Yang Z H & Yao X (2017). Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China. *Journal of Mountain Science* 14: 249-268. https://doi.org/10.1007/s11629-016-4126-9

Duniway M C, Pfennigwerth A A, Fick S E, Nauman T W, Belnap J & Barger N N (2019). Wind erosion and dust from US drylands: a review of causes, consequences, and solutions in a changing world. Ecosphere, 10(3), e02650. https://doi.org/10.1002/ecs2.2650

El Miloudi Y, El Kharim Y, Bounab A & El Hamdouni R (2024). Effect of Rockfall Spatial Representation on the Accuracy and Reliability of Susceptibility Models (The Case of the Haouz Dorsale Calcaire, Morocco). *Land*, *13*(2). https://doi.org/10.3390/land13020176

Fick S E & Hijmans R J (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 37:4302–4315. https://doi.org/10.1002/joc.5086

Foumelis M, Lekkas E & Parcharidis I (2004). Landslide susceptibility mapping by GIS-based qualitative weighting procedure in Corinth area. Bulletin of the Geological Society of Greece XXXVI, 904– 912. In: Proceedings of the 10th international congress, Thes- saloniki, April 2004

Garosi Y, Sheklabadi M, Pourghasemi H R, Besalatpour A A, Conoscenti C & Van Oost K (2018). Comparison of differences in resolution and sources of controlling factors for gully erosion susceptibility mapping. *Geoderma* 330: 65–78. https://doi.org/10.1016/j.geoderma.2018.05.027

Gayen A & Saha S (2017). Application of weights-of-evidence (WoE) and evidential belief function (EBF) models for the delineation of soil erosion vulnerable zones: a study on Pathro river basin, Jharkhand, India. *Modeling Earth Systems and Environment*, *3*(3), 1123–1139. https://doi.org/10.1007/s40808-017-0362-4

Graham M H (2003). Confronting Multicollinearity In Ecological Multiple Regression. Ecology 84:2809–2815. https://doi.org/10.1890/02-3114

Güney Y (2018). Use of frequency ratio method in erosion susceptibility analysis: Selendi Stream Basin (Manisa) example. *Journal of Soil Science and Plant Nutrition* 6(2): 73-85. (In Turkish)

Guzzetti F, Carrara A, Cardinali M & Reichenbach P (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31 (1-4): 181-216. https://doi.org/10.1016/s0169-555x(99)00078-1

Haan C T, Barfield B J & Hayes J C (1994). Design Hydrology and Sedimentology For Small Catchments. Academic Press An Imprint of Elseiver New York 38-101

Habib M (2021). Quantifying topographic ruggedness using principal component analysis. Advances in Civil Engineering, 1-20. https://doi.org/10.1155/2021/3311912

He S, Pan P, Dai L, Wang H & Liu J (2012). Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. Geomorphology, 171: 30-41. https://doi.org/10.1016/j.geomorph.2012.04.024

Hembram T, Paul G C & Saha S (2019). Spatial prediction of susceptibility to gully erosion in Jainti River basin, Eastern India: a comparison of information value and logistic regression models. *Modeling Earth Systems and Environment* 5(2): 689–708. https://doi.org/10.1007/s40808-018-0560-8

Hembram T K, Paul G C & Saha S (2020). Modelling of gully erosion risk using new ensemble of conditional probability and index of entropy in Jainti River basin of Chotanagpur Plateau Fringe Area, India. Appl Geomatics 12:337–360. https://doi.org/10.1007/s12518-020-00301-y

İDEP (2012). Climate Change National Action Plan 2011-2023, Ministry of Environment and Urbanization, Ankara. (In Turkish)

İnik O (2022). Negative Effects of Land Degradation on Human Life. "Innovative Methods, Theories and Applications in Health Sciences". Iksad Publications.ISBN: 978-625-8213-40-9. Ankara. (In Turkish)

İnik O (2023). Investigation of Land Degradation Balancing Studies in Çapakçur Microcatchment of Bingöl Province. Atatürk University, Institute of Science. (Doctoral Thesis) (In Turkish)

İnik O, İnik Ö, Öztaş T & Yüksel A (2022). Soil Temperature Prediction with Long Short Term Memory (LSTM). Turkish Journal of Agriculture and Natural Sciences , 9(3): 779-785. https://doi.org/10.30910/turkjans.1101753

Islam S, Tahir M & Parveen S (2022) GIS-based flood susceptibility mapping of the lower Bagmati basin in Bihar, using Shannon's entropy model. Model Earth Syst Environ 8:3005–3019. https://doi.org/10.1007/s40808-021-01283-5

Jaafari A, Najafi A, Pourghasemi H R, Rezaeian J & Sattarian A (2014). GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *International Journal of Environmental Science and Technology* 11: 909-926. https://doi.org/10.1007/s13762-013-0464-0

Jasiewicz J & Stepinski T F (2013). Geomorphology 182: 147-156. ScienceDirect

Jenks G (1967). The Data Model Concept in Statistical Mapping. In *International Yearbook of Cartography* (pp. 7:186-190)

Jenks G F (1967). The data model concept in statistical mapping. International Yearbook of Cartography 7: 186–190

Kakembo V, Xanga W W & Rowntree K (2009). Topographic thresholds in gully development on the hillslopes of communal areas in Ngqushwa Local Municipality, Eastern Cape, South Africa. Geomorphology 110:188–194. https://doi.org/10.1016/j.geomorph.2009.04.006

Karagöz A, Doğan O, Erpul G, Dengiz O, Sönmez B, Tekeli İ, Saygın S D & Madenoğlu S (2015). Evaluation of the Possible Effects of Desertification, Drought and Erosion in Turkey. Proceedings Book of the 8th Turkish Agricultural Engineering Technical Congress-1, 118. (In Turkish)

Kelava A, Moosbrugger H, Dimitruk P & Schermelleh-Engel K (2008). Multicollinearity and Missing Constraints. Methodology 4:51–66. https://doi.org/10.1027/1614-2241.4.2.51

Kheir R B, Abdallah C & Khawlie M (2008). Assessing soil erosion in Mediterranean karst landscapes of Lebanon using remote sensing and GIS. Engineering Geology, 99(3-4): 239-254. https://doi.org/10.1016/j.enggeo.2007.11.012

Kılıçoğlu C (2020). Production of landslide susceptibility map of Vezirköprü district of Samsun province using frequency ratio method and bayesian probability model. Afyon Kocatepe University Journal of Science and Engineering Sciences, 20(1), 138-154 (In Turkish)

Kleinbaum D G, Dietz K, Gail M, Klein M & Klein M (2002). Logistic regression (p. 536). New York: Springer-Verlag.

Lana J C, Castro P D T A & Lana C E (2022). Assessing gully erosion susceptibility and its conditioning factors in southeastern Brazil using machine learning algorithms and bivariate statistical methods: A regional approach. Geomorphology, 402, 108159. https://doi.org/10.1016/j.geomorph.2022.108159

Lei X, Chen W, Avand M, Janizadeh S, Kariminejad N, Shahabi H, Costache R, Shahabi H, Shirzadi A & Mosavi A (2020). GIS-based machine learning algorithms for gully erosion susceptibility mapping in a semi-arid region of Iran. Remote Sensing, 12(15): 2478. https://doi.org/10.3390/rs12152478

Lin J (1991). Divergence measures based on the Shannon entropy. IEEE Trans Inf Theory 37:145–151. https://doi.org/10.1109/18.61115

Maharaj R J (1993). Landslide processes and landslide susceptibility analysis from an Upland Watershed: a case study from St Andrew, Jamaica, West Indies. Engineening Geology 34 (1-2): 53-79. https://doi.org/10.1016/0013-7952(93)90043-c

Miao F, Zhao F, Wu Y, Li L & Török Á (2023). Landslide susceptibility mapping in Three Gorges Reservoir area based on GIS and boosting decision tree model. Stochastic Environmental Research and Risk Assessment, 37(6): 2283-2303. https://doi.org/10.1007/s00477-023-02394-4

Mohammed S, Al-Ebraheem A, Holb I J, Alsafadi K, Dikkeh M, Pham Q B, Linh N T T & Szabo S (2020). Soil management effects on soil water erosion and runoff in Central Syria—a comparative evaluation of general linear model and random forest regression. Water 12, 2529. https://doi.org/10.3390/w12092529.

Moore I & Burch G (1986). Physical Basis of the Length – slope Factor in the Universal Soil Loss Equation. *Soil Science Society of America Journal* 50: 1294-1298

Morgan R P C & Nearing M (2016). Handbook of erosion modelling. John Wiley & Sons.

Nagarajan R, Roy A, Vinod Kumar R, Mukherjee A & Khıre M V (2000). Landslide hazard suspectibility mapping based on terrain and climatic factors for tropical monsoon regions. Bulletin Engineering Geology and the Environment 58(4): 275-287. https://doi.org/10.1007/s100649900032

Neuhauser B & Terhorst B (2007). Landslide susceptibility assessment using “weights-of-evidence” applied to a study area at the Jurassic escarpment (SW-Germany). Geomorphology 86(1-2): 12-24. https://doi.org/10.1016/j.geomorph.2006.08.002

Nikolova V, Mitova M & Dimitrov E (2022). Topographic factor of water erosion–analysis of watershed morphometry and RUSLE LS factor in GIS environment. *Review of the Bulgarian Geological Society* 83(1): 3-14. https://doi.org/10.52215/rev.bgs.2022.83.1.3

Nkonge L K, Gathenya J M, Kiptala J K, Cheruiyot C K & Petroselli A (2023). An Ensemble of Weight of Evidence and Logistic Regression for Gully Erosion Susceptibility Mapping in the Kakia-Esamburmbur Catchment, Kenya. Water (Switzerland), 15(7). https://doi.org/10.3390/w15071292

Nkonge L K, Gathenya J M, Kiptala J K, Cheruiyot C K & Petroselli A (2023). An Ensemble of Weight of Evidence and Logistic Regression for Gully Erosion Susceptibility Mapping in the Kakia-Esamburmbur Catchment, Kenya. *Water (Switzerland)* 15(7): https://doi.org/10.3390/w15071292

Ogbonna J U, Alozie M, Nkemdırım V & Eze M U (2011). GIS analysis for mapping gully erosion impacts on the geo-formation of the Old Imo State, Nigeria. *ABSU Journal of Environment Science and Technology* 1: 48-61

Ohlmacher G C (2007) Plan curvature and landslide probability in regions dominated by earth flows and earth slides. Eng Geol 91:117–134. https://doi.org/10.1016/j.enggeo.2007.01.005

Öztürk D (2017). Investigation of Urban Sprawl with Shannon Entropy and Fractal Analysis: Samsun Example. 16th Turkey Scientific and Technical Congress on Mapping pp. 3-6. (In Turkish)

Öztürk M Z, Çetinkaya G & Aydın S (2017). Climate Types of Turkey According to Köppen-Geiger Climate Classification. *Journal of Geography*, 35: 17–27. https://doi.org/10.26650/jgeog295515

Peel M C, Finlayson B L & McMahon T A (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences* 11(5): 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Podhrazska J, Kučera J, Karasek P & Konečná J (2015). Land Degradation by Erosion and Its Economic Consequences for the Region of South Moravia (Czech Republic). Soil & Water Research, 10(2). https://doi.org/10.17221/143/2014-swr

Pulice I, Scarcıglia, F, Leonardi L, Robustelli G, Conforti M, Cuscıno M, Lupiano V & Critelli S (2009). Studio multidisciplinare di forme e processi denudazionali Nell’area di Vrica (Calabria Orientale). Bollettino della Società Geografica Italiana. 87(I–II): 399-414.

Rahmati O, Haghizadeh A, Pourghasemi H R & Noormohamadi F (2016). Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison. *Natural Hazards* 82(2): 1231–1258. https://doi.org/10.1007/s11069-016-2239-7

Rahmati O, Tahmasebipour N, Haghizadeh A, Pourghasemi H R & Feizizadeh B (2017). Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* 298: 118–137. https://doi.org/10.1016/j.geomorph.2017.09.006

Regmi N R, Giardino J R & Vitek J D (2010). Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. Geomorphology, 115(1-2): 172-187. https://doi.org/10.1016/j.geomorph.2009.10.002

Roy J & Saha DS (2019). GIS-based Gully Erosion Susceptibility Evaluation Using Frequency Ratio, Cosine Amplitude and Logistic Regression Ensembled with fuzzy logic in Hinglo River Basin, India. Remote Sens Appl Soc Environ 15:100247. https://doi.org/10.1016/j.rsase.2019.100247

Różycka M, Migoń P & Michniewicz A (2017). Topographic Wetness Index and Terrain Ruggedness Index in geomorphic characterisation of landslide terrains, on examples from the Sudetes, SW Poland. Zeitschrift für geomorphologie, Supplementary issues 61(2): 61-80. https://doi.org/10.1127/zfg_suppl/2016/0328

Saxena S (2021). A study on causes and consequences of soil erosion. Asian Journal of Research in Business Economics and Management, 11(10): 100-105. https://doi.org/10.5958/2249-7307.2021.00036.0

Saygın S D (2013). Climate Change and Global Warming: What Awaits Us? Köy-Koop News, April, 2013 Page: 17. (In Turkish)
Sharma B & Pandey A (2022). Mapping of Erosion Hazard in and around Kharagpur Hills, Bihar using hydrological indices. In MOL2NET'22, Conference on Molecular, Biomed., Comput. & Network Science and Engineering. Basel, Switzerland: MDPI. https://doi.org/10.3390/mol2net-08-12638
Sharma L P, Patel N, Ghose M K & Debnath P (2012). Influence of shannon's entropy on landslide-causing parameters for vulnerability study and zonation-a case study in sikkim, india. *Arab J Geosci* 5: 421–431. https://doi.org/10.1007/s12517-010-0205-3
Stepinski T F & Jasiewicz J (2011). Geomorphometry Papers, Redlands pp. 109-112
Sterk G (2003). Causes, consequences and control of wind erosion in Sahelian Africa: a review. Land Degradation & Development, 14(1): 95-108. https://doi.org/10.1002/ldr.526
Tehrany M S, Pradhan B & Jebur M N (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. Journal of Hydrology, 504: 69–79. https://doi.org/10.1016/j.jhydrol.2013.09.034
Trevisani S, Teza G & Guth P L (2023). Hacking the topographic ruggedness index. Geomorphology, 439: 108838. https://doi.org/10.1016/j.geomorph.2023.108838
UNCCD (2016). Science-Policy Notes. Balanced Lands-Balancing Land Degradation Scientific Concept Framework. 02 September 2016.).
Utlu M (2023). Flood Susceptibility Analysis of Ezine Stream Basin (Kastamonu-Bozkurt) Using Frequency Ratio and Shannon Entropy Method. *Journal of Geomorphological Research* pp. 160–178. https://doi.org/10.46453/jader.1358845 (In Turkish)
Valentin C, Poesen J & Li Y (2005). Gully erosion: Impacts, factors and control. Catena 63: 132–153. https://doi.org/10.1016/j.catena.2005.06.001
Van Westen C J, Rengers N & Soeters R (2003). Use of geomorphological information in indirect landslide susceptibility assessment, Natural Hazards 30(2003): 399-419. https://doi.org/10.1023/b:nhaz.0000007097.42735.9e
Wang F, Sahana M, Pahlevanzadeh B, Pal S C, Shit P K, Piran M J & Mosavi A (2021). Applying different resampling strategies in machine learning models to predict head-cut gully erosion susceptibility. *Alexandria Engineering Journal* 60(6): 5813-5829. https://doi.org/10.1016/j.aej.2021.04.026
Xu C, Xu X, Lee Y H, Tan X, Yu G & Dai F (2012). The 2010 Yushu earthquake triggered landslide hazard mapping using GIS and weight of evidence modeling. Environmental Earth Sciences 66: 1603-1616. https://doi.org/10.1007/s12665-012-1624-0
Yufeng S & Fengxiang J (2009). Landslide Stability Analysis Based on Generalized Information Entropy. 2009 International Conference on Environmental Science and Information Application Technology. IEEE, pp. 83–85
Yulianto F, Fitriana H L & Sukowati K A D (2020). Integration of remote sensing, GIS, and Shannon's entropy approach to conduct trend analysis of the dynamics change in urban/built-up areas in the Upper Citarum River Basin, West Java, Indonesia. Model Earth Syst Environ 6:383–395. https://doi.org/10.1007/s40808-019-00686-9
Zabihi M, Mirchooli F, Motevalli A, Darvishan A K, Pourghasemi H R, Zakeri M A & Sadighi F (2018). Spatial modelling of gully erosion in Mazandaran Province, northern Iran. Catena, 161: 1-13. https://doi.org/10.1016/j.catena.2017.10.010
Zabihi M, Pourghasemi H R, Motevalli A & Zakeri M A (2019). Gully erosion modeling using GIS-based data mining techniques in Northern Iran: a comparison between boosted regression tree and multivariate adaptive regression spline. *Advances in Natural and Technological Hazards Research*, 1-26. https://doi.org/10.1007/978-3-319-73383-8_1
Zhang Z, Xu W, Li L, Huang J, Deng, L & Wang Q (2021). Effects of temporal conservation measures on water erosion processes of disturbed soil accumulation in construction projects. *Journal of Cleaner Production*, 319, 128612. https://doi.org/10.1016/j.jclepro.2021.128612
Zhao J, Wang Z, Dong Y, Yang Z & Govers G (2022). How soil erosion and runoff are related to land use, topography and annual precipitation: Insights from a meta-analysis of erosion plots in China. Science of The Total Environment, 802, 149665. https://doi.org/10.1016/j.scitotenv.2021.149665
Zhu H, Tang G, Qian K & Liu H (2014). Extraction and analysis of gully head of Loess Plateau in China based on digital elevation model. *Chinese geographical science* 24: 328-338. https://doi.org/10.1007/s11769-014-0663-8
Zhuang Y, Du C, Zhang L, Du Y & Li S (2015). Research trends and hotspots in soil erosion from 1932 to 2013: A literature review. Scientometrics 10: 743–758. doi: 10.1007/s11192-015-1706-3