Mercator, Fortaleza, v. 22, e22010, 2023. ISSN:1984-2201

PREDICTIVE MODELING APPLIED TO POTENTIAL SOIL EROSION RISK MAPPING IN THE WESTERN AMAZON

https://doi.org/10.4215/rm2023.e22010

Fonseca, E.L. [a*] - Filho, E.P.S. [b]

(a) PhD in Geography.
ORCID: https://orcid.org/0000-0002-6872-8204. **LATTES:** http://lattes.cnpq.br/4686984012455015.
(b) PhD in Geosciences and Environment
ORCID: https://orcid.org/0000-0002-6271-8435. **LATTES:** http://lattes.cnpq.br/2651173141421798.

MERCATOR

ARTICLE

Article history:
Received 21 November, 2022
Accepted 24 April, 2023
Published 30 April, 2023

(*) CORRESPONDING AUTHOR
Address: IFRO. BR 435, km 63, CP:51, CEP: 76993000, Colorado do Oeste (RO), Brazil. Phone: (+5569) 992221787.
E-mail: elaine.fonseca@ifro.edu.br

Abstract

The modeling of areas susceptible to soil loss due to hydro-erosive processes consists of methods that simplify reality to predict future behavior based on the observation and interaction of a set of geoenvironmental factors. Thus, the objective of the current analysis is to predict susceptibility to soil loss and map areas with the potential risk of erosion using the principles of Binary Logistic Regression (BLR) and Artificial Neural Networks (ANN). The hydrographic sub-basin of the Sete Voltas River (330 km2), Rondônia, Brazil, was defined as the experimental area. Models were obtained using 100 sample units and 14 predictor parameters. Susceptibility was mapped based on five reference classes: very low, low, moderate, high, and very high. ANN obtained an area under the curve (AUC) of 0.808 and global precision of 79.2%, and the BLR model showed an AUC of 0.888 and global precision of 77%. Potentially susceptible areas represent 57.71% and 54.80% of the area for BLR and ANN models, respectively. The greatest potential risks are verified in places with no vegetation cover associated with agricultural practices. The technique proved to be effective, with adequate precision and the advantage of being less time-consuming and expensive than other methods.

Keywords: Binary Logistic Regression; Artificial Neural Network; Erosion Susceptibility.

Resumo / Resumen

MODELAGEM PREDITIVA APLICADA AO MAPEAMENTO DE RISCO POTENCIAL DE EROSÃO DE SOLOS NA AMAZÔNIA OCIDENTAL

A modelagem de áreas suscetíveis à perda de solo por processos hidroerosivos consiste em um instrumento simplificado da realidade com a finalidade de prever comportamentos futuros a partir da observação e interação de um conjunto de fatores geoambientais. À face do exposto, a corrente análise tem como objetivo prever a suscetibilidade à perda de solo por evento hídrico e mapear as áreas com risco potencial de erosão, utilizando os princípios de Regressão Logística Binária (RLB) e Redes Neurais Artificiais (RNA). Para tanto, definiu-se a sub-bacia hidrográfica do rio Sete Voltas (330 km2) como área experimental no município de Colorado do Oeste/RO, sul da Amazônia brasileira. Inicialmente, foi concebido o mapa de inventário de erosão de solo com 100 unidades amostrais e 14 parâmetros preditores que englobasse aspectos ambientais, topográficos e geológicos. A suscetibilidade foi mapeada com base em cinco classes de referência: muito baixa, baixa, moderada, alta e muito alta. A RNA obteve área sob a curva (AUC) de 0,808 e precisão global de 79,2%; o modelo RLB apresentou AUC de 0,888 e precisão global de 77%. As áreas potencialmente susceptíveis representam 57,71% e 54,80% da área para os modelos RLB e RNA, respectivamente. Os maiores riscos potenciais são verificados em locais sem cobertura vegetal associada às práticas agrícolas. A técnica mostrou-se eficaz, com precisão adequada e com a vantagem de ser menos demorada e onerosa em relação a outros métodos.

Palavras-chave: Regressão Logística Binária; Rede Neural Artificial; Suscetibilidade à Erosão.

MODELACIÓN PREDICTIVA APLICADA AL MAPEO DE RIESGO POTENCIAL DE EROSIÓN DEL SUELO EN LA AMAZONÍA OCCIDENTAL

La modelación de áreas susceptibles a la pérdida de suelo por procesos hidroerosivos consiste en un instrumento simplificado de la realidad con el propósito de predecir comportamientos futuros a partir de la observación e interacción de un conjunto de factores geoambientales. En vista de lo anterior, el presente análisis tiene como objetivo predecir la susceptibilidad a la pérdida de suelo por evento hídrico y mapear las áreas con riesgo potencial de erosión, utilizando los principios de Regresión Logística Binaria (RLB) y Redes Neuronales Artificiales (RNA). Para ello, se definió la subcuenca del río Sete Voltas (330 km2) como área experimental en el municipio de Colorado do Oeste/RO, sur de la Amazonia brasileña. Inicialmente, se diseñó el mapa de inventario de erosión del suelo con 100 unidades de muestreo y 14 parámetros predictores que abarcaban aspectos ambientales, topográficos y geológicos. La susceptibilidad fue mapeada con base en cinco clases de referencia: muy baja, baja, moderada, alta y muy alta. La RNA obtuvo un área bajo la curva (AUC) de 0,808 y una precisión global del 79,2%; el modelo RLB presentó un AUC de 0,888 y una precisión global del 77%. Las áreas potencialmente susceptibles representan el 57,71% y el 54,80% del área para los modelos RLB y RNA, respectivamente. Los mayores riesgos potenciales se verifican en sitios sin cobertura vegetal asociada a prácticas agrícolas. La técnica demostró ser eficaz, con una precisión adecuada y con la ventaja de requerir menos tiempo y ser menos costosa en comparación con otros métodos.

Palabras-clave: Regresión Logística Binaria; Red Neuronal Artificial; Susceptibilidad a la Erosión

INTRODUCTION

The mapping of areas with the potential for erosion is of fundamental importance considering the current environmental dynamics in the southern region of the Brazilian Amazon. Unplanned human settlement, as a consequence of government policies and initiatives, has caused significant changes in the surface characteristics of the soil, favoring the development and evolution of erosive processes (FONSECA, 2017).

The unique context of the Amazon is the result of a combination of factors, including geological, geomorphological, vegetation, and the predominant climate since the Mesozoic. For example, in terms of geology, the study area in Rondônia State, Brazil, is a sedimentary basin primarily made up of crystalline and sedimentary formations originating from a range of periods, from the Archean to the Holocene. The geomorphology is formed by a flattened relief with depressions, and the vegetation is quite diverse with current phytophysiognomies including savanna, Cerrado, and forest domains (CPRM, 1999; PLANOFLORO, 1998).

In the study area, there is a predominance of anthropogenic landscapes, represented by cultivated pastures, agriculture, and secondary vegetation at various stages of forest succession. Forest remnants are made up of small, isolated fragments across the region. In the Amazon, these and other factors are responsible for substantial soil loss every year due to erosion (TAKAKI, 2002; ARRUDA et al. 2004; FONSECA, 2017), with significant consequences for the maintenance of agricultural productivity and the value of rural properties.

In general, erosion that occurs in the Amazon is accelerated due to human interference, which has been particularly acute since the 1970s, when a series of negative environmental events such as deforestation, fire, logging, and encroachment of the agricultural frontier resulted in an imbalance in the soil-vegetation equilibrium (ALBUQUERQUE; VIEIRA, 2014).

Thus, the problems resulting from soil loss are diverse and are concerning due to deterioration in soil quality and function (MOSAVI et al. 2020, ROY; SAHA, 2021). Predicting which areas are more sensitive to environmental problems enables us to recommend the best ways to implement development projects and promote greater environmental sustainability (XING et al. 2021).

The mapping of areas with potential risk of soil loss offers a representation of current knowledge about land use in relation to responses to erosive processes. From this, it is possible to identify areas susceptible to diverse aspects of environmental impacts caused by land use (ARABAMERI et al. 2021).

The identification and understanding of the triggering factors of erosive processes have been facilitated by advances in techniques such as artificial intelligence, algorithms, GIS, and remote sensing (AL-NAJJAR; PRADHAN, 2021). These techniques allow for the straightforward assessment of the state of soil degradation, both over time and at specific points, and in areas that are difficult to access, thus contributing significantly to the study of soil erosion.

There are several techniques for modeling susceptibility to soil loss, including: logistic regression (RAJA et al. 2016); information value (SARKAR et al. 2013); conditional probability (RAHMATI et al. 2017); frequency ratio (MELIHO et al. 2018); entropy index (JAAFARI et al. 2014); certainty factor (SOMA; KUBOTA, 2018); Frequency radio (WANG et al. 2016); weights of evidence (WOE) (GOYES-PEÑAFIEL; HERNANDEZ-ROJAS, 2021); fuzzy logic and neuro-fuzzy (YAVARI et al. 2018); artificial neural networks (ANN) (SHAHRI et al. 2019); support vector machine (LEE et al. 2017); adversary generative network (AL-NAJJAR, PRADHAN, 2021); and convolutional neural network (MEENA et al. 2021), among others.

For this study, we chose to use binary logistic regression (BLR) and artificial neural networks (ANN). The choice of a BLR model is due to the high degree of reliability and ease of dealing with categorical independent variables. According to Bissacot (2015), logistic regression models are a standard technique that are well established as tools to aid in decision making.

On the other hand, the ANN model was chosen because of its simplicity and efficiency. ANNs can deal with complex data, identify subtle patterns present in the training input, and solve problems with unidentified patterns occurring in the input data (CONFORTI et al. 2014).

Logistic regression is a statistical technique that produces a model to predict values. Therefore, it is a well-recognized and widely applied nonlinear system used to predict the probability of presence or absence of a dependent variable outcome for a set of predictor variables, that can be continuous, discrete, or a mixture of both (SARKAR; MISHRA, 2018).

A regression model can be defined as a mathematical equation that expresses the relationship between variables. It assesses the likelihood of an observation belonging to each group, estimating the probability that an observation belongs to a certain group (MALHOTRA, 2019).

ANNs, in turn, are applications of artificial intelligence (VAEZI et al. 2020), that contain an interrelated set of artificial neurons that process information using a connectionist computation formula (AL-SHAWWA; ABU-NASER, 2019). They are naturally more versatile, powerful, and scalable, making them ideal for handling high-complexity machine learning tasks (GERON, 2017). ANNs simulate the functioning of a neuron using mathematical equations and consist of a network of artificial neurons in which logical information or numerical values can be processed to generate an output or an answer.

Considering this context, and based on an informed selection of geoenvironmental predictors, this study used BLR statistical treatments and ANN machine learning to build models of susceptibility to soil erosion due to hydric events in the southern region of the Western Amazon, Brazil.

MATERIAL AND METHODS

LOCATION AND CHARACTERISTICS OF THE EXPERIMENTAL AREA

The research was undertaken in the western Amazon, Rondônia State, Northern Brazil. The experimental area is a hydrographic sub-basin (Sete Voltas River) with an area of 330.49 km2 (13°04'45.329” S / 60°30'42.942” W) in the municipality of Colorado do Oeste (Figure 1).

The average annual rainfall is 1,900 mm per year-1 (FONSECA, 2017). Based on the Koppen classification, the predominant climate is Tropical Rainy (Aw), with an average air temperature during the coldest month greater than 18 °C (megathermal) (SEDAM, 2010). The average annual air temperature is high and uniform, with variation of the average between 24 and 26 °C (VIEIRA et al. 2014). The dry period is characterized by three months with rainfall of less than 50 mm (June, July, and August), a limited range of annual thermal amplitude, and notable daily thermal amplitude (FONSECA, 2017).

The drainage network includes tributaries of the Guaporé River and springs along the edge of the Chapada dos Parecis mountain range (RADAMBRASIL, 1979). According to Fonseca and Silva Filho (2017), the sub-basins in the region are dendritic in an exorheic system, composed mostly of first-order channels that flow directly into the main river.

The vegetation cover includes Semi-deciduous Ombrophilous Forest, Cerrado, and areas of transition between the two biomes, occurring as natural forest fragments or those that have regenerated after anthropogenic disturbance (FONSECA et al. 2018).

Among the economic activities in the region, agriculture is particularly important and is based on the production of annual crops (soybeans and corn) and livestock for milk and beef. Small-scale rural properties are predominant in the region, most of which use intensive production systems with poor zootechnical herd indices.

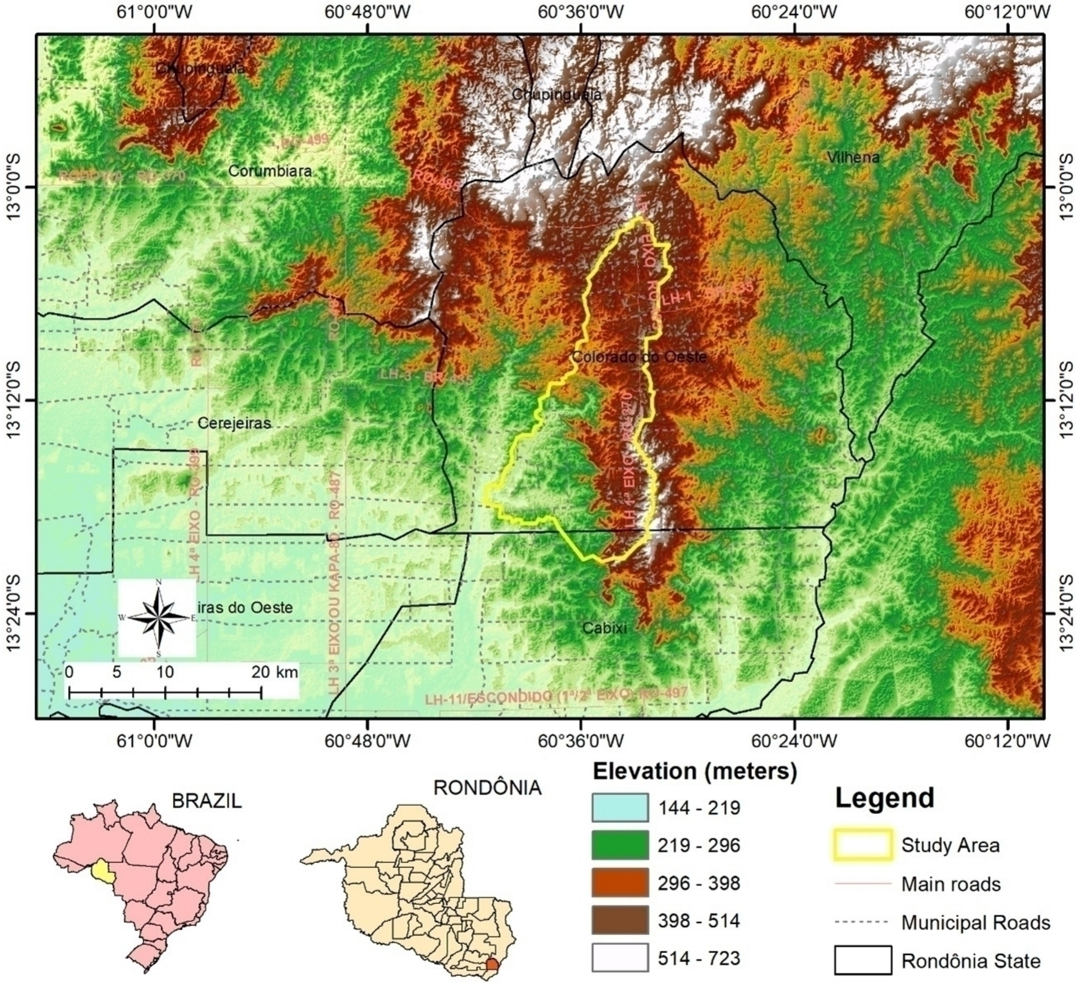

Figure 1 - Location of the experimental area.

METHODOLOGICAL PROCEDURES

The mapping of soil erosion susceptibility used a BLR statistical model and ANN machine learning. The erosion inventory map was obtained from 100 sampling units: 50 units with the presence of erosion and 50 units with an absence of erosion.

The sample units were mapped in July 2020 with the aid of a PHANTOM 4 (DJI) Unmanned Aircraft System (UAS), using a regular sampling grid of 2 km x 2 km. The aerial photographs were used only as observation data - visual interpretation. The maps were supplemented with information from PLANAFLORO/RO and orbital images from the Sentinel 2B satellite with a spatial resolution of 10 m on the date of July 17, 2020. The images were geometrically corrected, and an enhancement technique was applied.

The sample selection criteria were visual and based on the following indicators: presence of exposed soil; removal of organic surface horizons; laminar erosion; presence of terraces; and formation of ravines and gullies. Sites that showed any or all of these indicators were considered susceptible to erosion (coded as 1 or success); otherwise, susceptibility was considered minimal or nil (coded as 0 or failure).

The selection of the parameters was based on field observations, the researchers' experience, current literature, the scale of analysis, data availability, and the purpose of the research. The databases used were: Brazilian Geomorphometric Database – TOPODATA (IBGE); Socioeconomic-Ecological

Zoning of the State of Rondônia - PLANAFLORO (SEDAM/RO); Geological Survey of Brazil – CPRM; LANDSAT 8 Satellite (Scene 230/69, May 22, 2020).

The input layers were constructed in a 14 x 100 matrix (100 samples of 14 target elements). For modeling, 70% of the samples were used for training the model and 30% for validating the results in terms of specificity, sensitivity, and precision. The statistical calculations and algorithms of the BLR and ANN models were developed and simulated using SPSS Statistics 26.0.

INPUT PARAMETERS

The input parameters included three main aspects: 1. Topographic parameters derived from the digital elevation model (Elevation, Slope, Aspect, Slope Curvature, Composite Topographic Index (CTI), and Stream Power Index (SPI)); 2. Geological Parameters (Lithology, Drainage Density, And Lineament Density); 3. Environmental Parameters (Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), Land Use and Occupation, Erosivity, and Soil Type), as described in Table 1.

Elevation influences the flow of water in different paths over the terrain and acts to structure the landscape (SOUZA et al. 2003). Slope is related to an increase in the slope gradient and flow shear that stimulates the detachment of soil particles (GUERRA et al. 2014).

The Slope orientation in mountainous terrain strongly influences the amount of solar radiation received on the ground (PERREAULT et al. 2016). The curvature of the slope refers to the divergent or convergent flows of matter and energy on the slopes in relation to a horizontal plane (GUADAGNIN; TRENTIM, 2014).

The Composite Topographic Index (CTI) directly reflects the present or past conditions of moisture and water flow dynamics. CTI is used to characterize the spatial distribution of saturation zones and soil water content (PINHEIRO, 2015). The Stream Power Index (SPI) describes the potential of the topography to aggregate large amounts of surface water (CAPOANE, 2015).

The Lineament Density can indicate the fragilities, faults, and fractures of the terrain (CARMO et al. 2016), while the NDVI is related to the terrain's natural defense against erosion, through elements that directly protect against the impact of rainfall.

Drainage Density relates to the distance that the water must travel to the riverbed. The lower the drainage density, the greater the distance that the water must travel. This is directly related to flooding capacity and increased sediment transport rate (FONSECA; SILVA FILHO, 2017).

Lithology influences erosive processes through the mineralogical and textural characteristics of the rocks present in the geological substrate, affecting the permeability and ease of transporting loose particles (CAMPOS, 2019). Land use and occupation refers to development activities and the possible aggravation of the erosive processes due to soil structure damage and sediment loss (PERUSI; CARVALHO, 2008). Soil type is the chemical and physical properties that affect erosive processes.

Erosivity expresses the potential of rainfall to cause erosion, based on the ratio between monthly and annual precipitation (FONSECA, 2017). Land Surface Temperature (LST) influences daily and annual fluctuations in soil temperature that affect biological and chemical processes in the soil, including decomposition, rates of soil organic matter uptake, and release of CO2 (CARNEIRO et al 2014).

Data layers	Method and/or equation	Variable type	Source
Elevation (ELV)	-	Numerical	TOPODATA (IBGE)
Slope (SLP)	$D = \left(\sqrt{\frac{d_z}{d_y}^2 + \frac{d_z}{d_x}^2}\right)$ $\frac{d_z}{d_y}$ = rate of change in direction **x** $\frac{d_z}{d_x}$ = rate of change in direction **y**	Numerical	TOPODATA (IBGE)
Aspect (ASP)	$AS = 57{,}29578 * \left(\frac{d_z}{d_y} - \frac{d_z}{d_x}\right)$ NE→ '1'; E→ '2'; SE→ '3'; S→ '4'; SW→ '5', W→ '6'; NW→ '7'; N→ '8'	Numerical	TOPODATA (IBGE)
Slope curvature (CV)	-	Numerical	TOPODATA (IBGE)
Compound Topographic Index (CTI)	$CTI = log\ log\ \frac{AS}{tan\ tan\ \beta}$ AS = specific capture area across a unit width of the contour; β = slope gradient in degrees	Numerical	TOPODATA (IBGE)
Stream Power Index (SPI)	$SPI = ln\left[(R_{Fluxo\ acu} + 0{,}001) * \frac{R_{Decliv}}{100} + 0{,}001\right]$ $R_{acu\ flow}$ = Raster accumulated flow; R_{slope} = Raster slope	Numerical	TOPODATA (IBGE)
Lineament Density (DL)	*Line Density - Spatial Analyst Tool* (ArcGIS)	Numerical	CPRM
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$ ρ_{NIR} = near infrared; ρ_{RED} = red	Numerical	LANDSAT 8/OLI
Drainage Density (DD)	*Line Density - Spatial Analyst Tool* (ArcGIS)	Numerical	TOPODATA (IBGE)
Lithology (LIT)	-	Categorical	CPRM
Land use and occupation (USE)	Bhattacharya Supervised Classification 99%	Categorical	LANDSAT 8/OLI
Soil type (SOIL)	-	Categorical	PLANAFLORO /RO
Erosivity (EROS)	$Elm = p2\ /P$ Elm = average erosivity index $p2$ = average monthly precipitation (mm) P = average annual precipitation	Numerical	EMBRAPA
Land Surface Temperature (LST)	$LST = \frac{K_2}{ln\left(\frac{\varepsilon_{NB}\ K_1}{L_{\lambda,6}} + 1\right)}$ LST = Land Surface Temperature $L_{\lambda,6}$ = spectral radiance of the thermal band; ε_{NB} = emissivity K_2 and K_1 = constant sensor	Numerical	LANDSAT 8/TIRS

Table 1 - Layers of data used in the study

In terms of categorical variables, the classes were ordered according to the degree of susceptibility to particle transport. Soil type classification in relation to susceptibility follows the indications described by Ross (1994) (Table 2).

The weighting of the lithology variable (Table 3) is related to the weathering that rocks undergo when exposed to the surface. It is the beginning of a long-term process that continues with the erosion and deposition of material, subsequent diagenesis, leading to the formation of sedimentary rocks (CPRM, 2014).

Soil classes	Predominant characteristics	Weight
Latosols	Slope of 0-2%, well drained and loamy to clayey texture.	0
Gleysols	Slope varies from 0-2%, it is poorly drained and loamy.	1
Cambisols	Declivity between 8 and 30%, well drained, clayey, rocky.	2
Argisols	Predominant slope of 2-8%, with sites exceeding 30%, well drained, clayey, and slightly rocky.	3
Neosols	2-8% slope, well drained and sandy.	4

Table 2 - Instability by soil class

Group	Unit	Predominant rocks	Weight
C	Mafic-Ultramafic Trench Complex	Granitoids	0
B	Utiariti Formation	Quartz arenites	1
A	Avila River Formation	Bimodal arenites	2
E	Corumbiara Formation	Immature polymictic conglomerates	3
D	Undifferentiated Sedimentary Coverage and Alluvial Deposits	Sands, silts, and clays	4

Table 3 - Instability by lithological class

The classification used to map land use and occupation was based on the application of supervised classification in Spring 5.6 to a segmented image. For the supervised classification, Bhattacharya was used with 99.9% acceptance. The Kappa agreement index was 0.71, a value considered very good. The representative classes of the image are described in Table 4.

Susceptibility was attributed according to the degree of protection that a certain type of land cover offers to the soil, following the protection capacity hierarchy described by Ross (1994). The class ‘water resources’ was added due to the frequent presence of ravines and gullies near waterways.

Class	Description	Weight
Native vegetation	Characterized by forests and natural or regenerated forest fragments after human intervention.	0
Pasture	Characterized by the presence of grazing vegetation intended for animal production.	1
Agriculture	Characterized by areas under agricultural production.	2
Water resources	Characterized by rivers, lakes, and other bodies of water found in the image.	3

Table 4 - Land use and occupation classes.

The data maps used in this study are based on a cell size of 30 × 30 m, resampled for the Geographic Coordinate System and projected on Datum WGS 1984. The work was carried out using the ArcGIS 10.5 GIS software environment. Summary statistics can be found in Table 5. For mapping classification purposes, erosion susceptibility was divided into five classes (Low, Very Low, Moderate, High and Very High) based on Jenks' natural breakdown algorithm.

Numeric Parameters	Max	Min	Average	SD	Interval
Elevation	565.06	204.73	361.01	91.25	360.56
Slope	0.20	74.57	10.98	8.25	74.36
Aspect	0	359.99	177.34	107.26	359.99
CV	-0.14	0.14	-0.001	0.02	0.28
CTI	-8.82	2.04	-4.95	1.27	10.86
SPI	6.30	-12.68	-2.96	3.84	-6.38
LD	0	92.50	42.04	25.61	92.50
NDVI	-0.13	0.11	-0.04	0.022	0.25
DD	0	150.51	77.95	21.75	150.51
ERO	8055.42	6999.56	7449.16	285.87	1155.86
LST	21.21	32.22	24.84	1.27	11.01
Categorical Parameters			**Distribution of Area in %**		
Lithology*	A	B	C	D	E
	78.77	5.19	5.80	6.67	3.57
Land use	Native vegetation	Pasture	Agriculture	Water resources	
	17.12	79.44	2.12	1.30	
Soil type	Latosol	Gleysol	Cambisol	Argisol	Neosol
	32.03	3.58	7.26	52.39	4.75

Table 5 - Descriptive statistics of the input layer parameters.

Nota. SD – standard deviation; CV - curvature of the slope; CTI – composite topographic index; SPI – stream power index; LD – lineament density; NDVI – normalized difference vegetation index; DD – drainage density; LST – land surface temperature; Lithology* - refers to the lithostratigraphic units corresponding to the predominant type of rock.

The predictor parameters used in the analysis can be found in Figure 2.

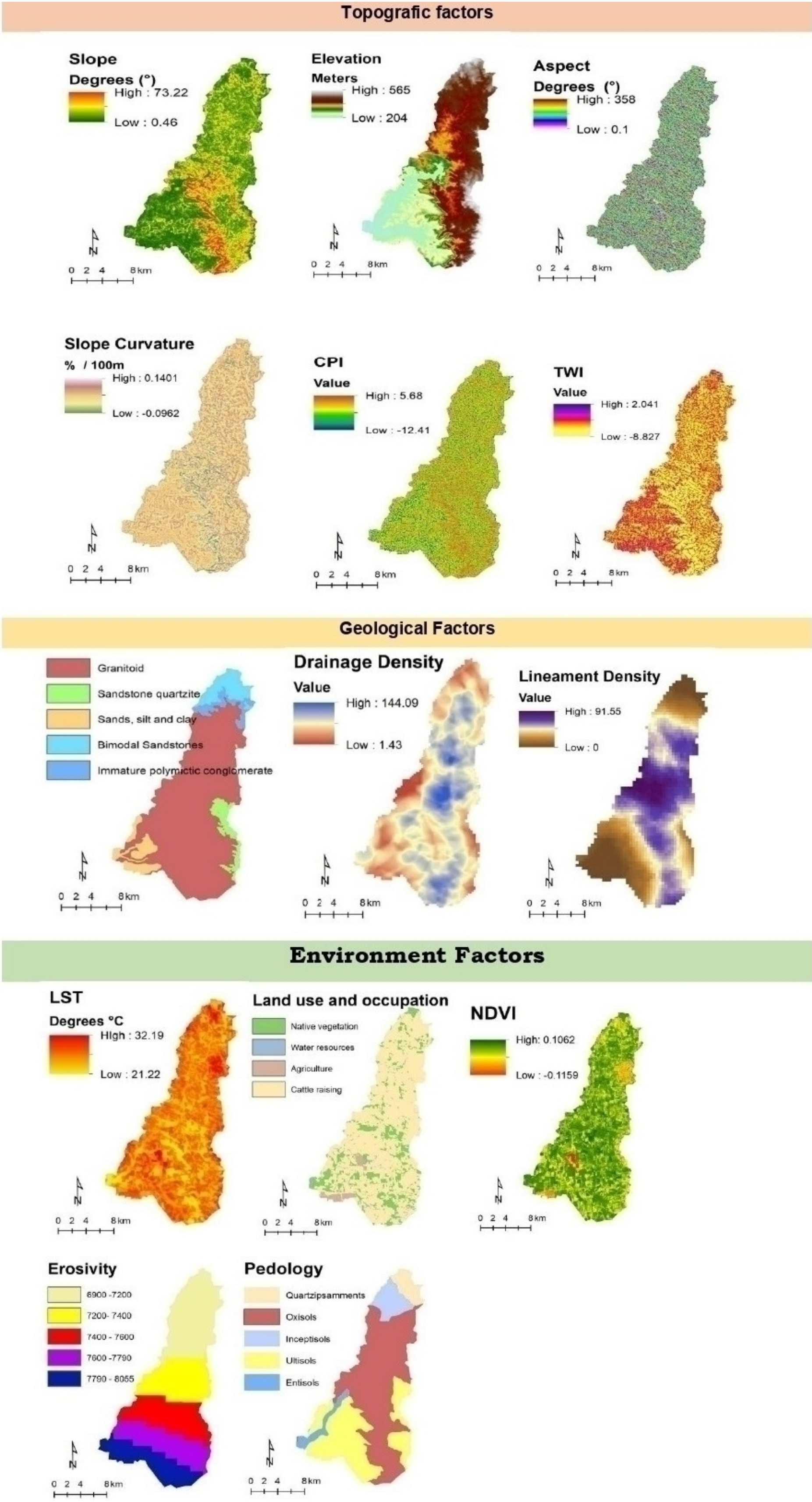

Figure 2 - Spatial distribution of raster layers.

RESULTS

The accuracy of the BLR model reached an overall precision of 77% (Table 6). The matrix indicates the percentage of success and error of the model for the two possible answers. Thus, the model correctly classified 80% of the samples from the 50 sampling units with the presence of erosion, and 74% of the 50 samples without erosion.

	Observed		Predicted		
			Reply		Correct percentage
			0	1	
Stage 1	Reply	0	37	13	74.0
		1	10	40	80.0
	Overall percentage				77.0

Table 6 - Binary Logistic Regression Confusion Matrix

The model sensitivity and specificity are expressed in the area under the curve (AUC). The AUC model was 0.888 (88.8%) for the RLB model and 0.808 (80.8%) for the ANN model.

The ANN model was classified with all samples, for a total of 77 sample units for training and 23 sample units for testing. The training phase employed the scaled conjugate gradient algorithm with the sigmoid activation function. The synaptic weights of the neural network that achieved better results were for two hidden layers, with seven neurons in the first layer and two neurons in the second layer.

The model with two hidden layers correctly predicted erosion susceptibility for 72.4% of the training data sample units and 79.2% of the test data units (Table 7). Considering that the model was developed specifically to identify areas susceptible to erosion, the accuracy of predicting the presence of erosion, or '1', was more appropriate to be considered in this study.

Sample	Observed	Predicted		
		0	1	Correct Percentage
Training	0	25	12	67.6
	1	9	30	76.9
	Overall Percentage	44.7	55.3	72.4
Test	0	10	3	76.9
	1	2	9	81.8
	Overall Percentage	50.0	50.0	79.2

Note: 0 - absence of erosion; 1 - presence of erosion

Table 7 - Artificial Neural Network Confusion Matrix

The synthesis maps from the BLR statistical model and ANN machine learning model are shown in Figure 3. The raster output produced by the BLR and ANN methods highlighted susceptibility to erosion between the values 0 and 1, where 0 represents a low probability of soil erosion and 1 the highest probability of soil erosion in the study area.

Areas that are susceptible to erosion correspond to those classified as moderate, high, and very high, and represent 57.71% and 54.80% of the area for BLR and ANN models, respectively (Table 8, Figure 4).

Erosion Susceptibility	Model BLR		Model ANN	
Class	No. Pixels	% of area	No. Pixels	% of area
Very low	67085	19.44	69913	20.27
Low	78779	22.83	85977	24.93
Moderate	121053	35.09	125638	36.42
High	73153	21.20	61819	17.92
Very high	4856	1.40	1579	0.46
Total	**344926**	**100**	**344926**	**100**

Table 8 - Erosion susceptibility class and area of coverage by BLR and ANN.

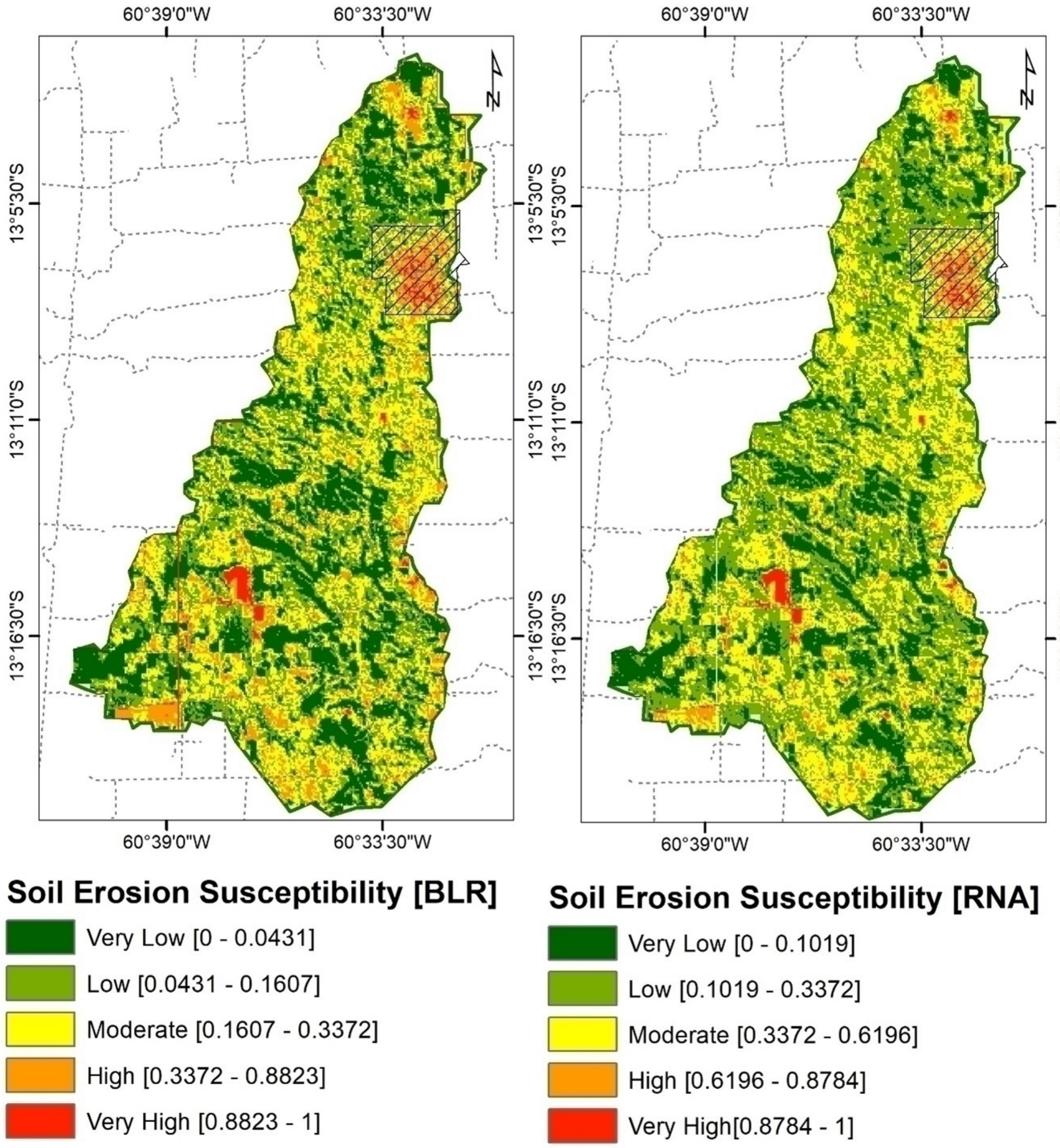

Figure 3 – Map of susceptibility to soil erosion of the Sete Voltas River Sub-basin - Binary Logistic Regression Model [BLR] and Artificial Neural Network [ANN].

DISCUSSION

Mapping revealed the existence of a belt of low erosion susceptibility in the southeast of the sub-basin (Figure 3). The lowest risk of erosion is associated with the maintenance of native forest, since vegetation cover can mitigate soil loss processes through protection against the direct impact of rainfall on the surface and water dispersion by interception and evaporation before it reaches the ground (BERTONI; LOMBARDI NETO, 2008).

Low levels of erosion risk can also be observed in the extreme southwest. (Figura 3). These are lowland areas that can flood at certain times of the year, making agricultural activities unfeasible. The preservation of natural vegetation, depending on its characteristics, can help to ensure protection of the soil.

In adjacent regions, agricultural practices increase the risk of soil loss. These are newly converted areas to agriculture, and soil exposure to the kinetic action of rainfall leads to the loss of the surface layer. Loss of surface layers results in a reduction in fertility due to fewer macro and micronutrients and less organic matter in the soil. For agriculture, the reduction of these elements has a direct impact on crop productivity and production costs, as low levels of nutrient availability for plants necessitates the use of fertilizers.

In general, the ravine formation process in arable areas is associated with different soil textures, microtopography, and vegetation cover. When there is erosion in cultivated land, the nutrients present in the upper layers are lost as they are incorporated into the eroded soil due to their high solubility and rapid absorption by fine soil particles (BERTONI; LOMBARDI NETO 2008). Consequently, there is a reduction in the productive capacity of plant biomass and soil protection (MAFRA, 2014), along with an increased risk of flooding due to the deposition of solid material in water resources.

In the extreme north of the sub-basin, susceptibility to erosion is related to the conversion of Cerrado vegetation, or transition zones between Ombrophilous Forest and Cerrado, into pasture (Figure 3). These are areas composed of Neosols located on the edge of the Chapada dos Parecis mountain range which is heavily influenced by weathering agents.

When evaluating erosional conditions in Quartzarenic Neosols in the municipality of Colorado do Oeste, Fonseca (2017) found that the predisposition to sediment loss stems from the predominance of very coarse and coarse sand, high soil density values, low total porosity, soil compaction related to cattle trampling, and an increase in the average resistance to penetration, mainly in the first 10 cm.

The central area of the sub-basin and the region close to the urban center show moderate susceptibility. The greatest potential risks are associated with degraded pasture and lowland areas close to water bodies (Figure 4).

Fonseca et al. (2018) classified the stages of pasture degradation in the same study area and found that areas with some degree of degradation have similar characteristics, such as: the variety of the forage vegetation (brachiaria brizantha); low forage height; low plant population per m2; presence of weeds; high grazing pressure; and poor pasture formation and management.

Erosion surrounding drainage networks is a major issue in the study area. The rivers in the sub-basin are mostly first-order channels, which are fragile, intermittent, with low water runoff volume, and high susceptibility to anthropogenic pressure (Figure 5).

When analyzing the morphometry of the hydrographic sub-basins of the municipality studied herein, Fonseca and Silva Filho (2017) identified the aforementioned characteristics in relation to water resources, pointing out that the first-order channels flow directly into the main river and have low flow. Without adequate management of agricultural activities, they may cease to exist.

Based on previous analyses, it is clear that susceptibility mapping techniques are important tools in understanding the phenomenon and in the management of areas where soil loss already occurs. According to Arabameri et al. (2018), mapping is a basic method to understand the mechanisms behind erosive events. In addition to enabling the identification and measuring of the relevant conditioning factors (ARABAMERI et al. 2020), such maps can be used to inform planning, identify suitable areas for infrastructure development (BRAGAGNOLO et al. 2020), optimize land use, and mitigate the effects of inappropriate land use.

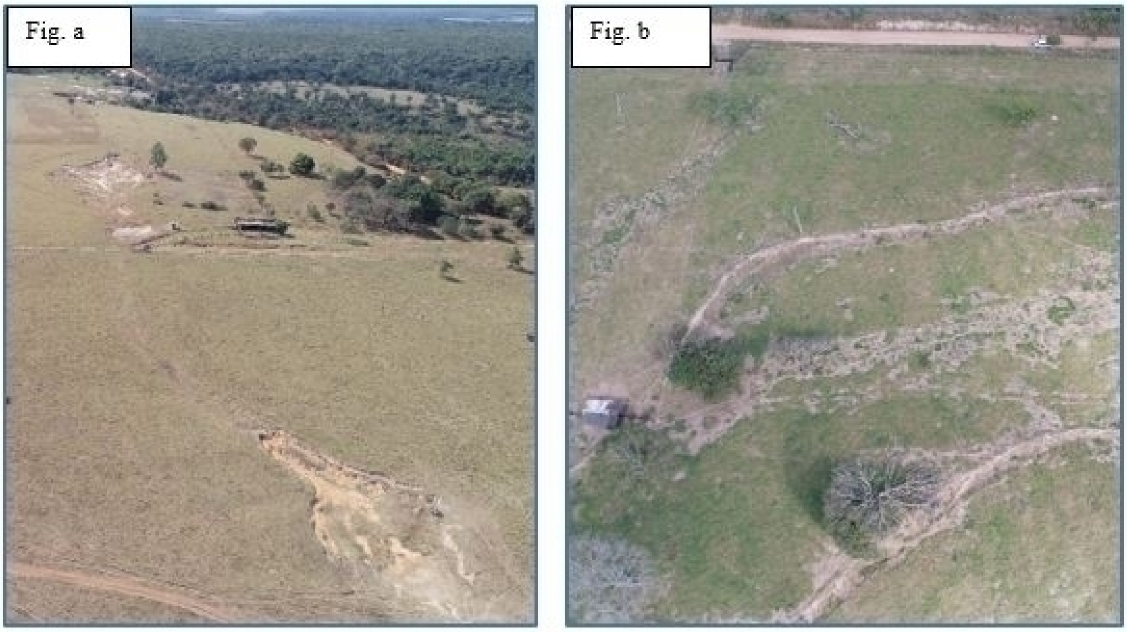

Figure 4 - Aerial photographs of cultivated pastures with a high degree of degradation and erosion. Fig. 4a shows the beginning of the ravine formation process; Fig. 4b shows areas of laminar erosion with consequent soil exposure to rainfall. Source: Author (July/2020)

Figure 5 - Aerial photographs of erosion close to water resources. Source: Author (July/2020).

AUC values of 0.888 (88.8%) for the BLR model and 0.808 (80.8%) for the ANN model are classified as very good and good according to the qualitative-quantitative relationship described by Polo and Miot (2020). Thus, they can be considered satisfactory for subsequent application in the mapping of soil erosion susceptibility. AUC estimates have been used as acceptance parameters for a model's performance by several authors (RAHMATI et al. 2016, SARKAR; MISHARA, 2018; BRAGAGNOLO et al. 2020).

As for the models studied herein, Bissacot (2015) states that logistic regression is a standard technique that is well established as a decision-making tool, in addition to being preferable when the dependent variable is categorical dichotomous (NARDI et al. 2019). According to Conforti et al. (2014), ANNs must be applied to deal with complex data patterns, identify subtle patterns present in the training input, and solve problems with unidentified patterns present in the input data.

According to Bragagnolo et al. (2020), the use of ANN presents several advantages for soil loss susceptibility studies, ranging from a methodology that applies learning algorithms that define their own architecture, to high versatility in that they can be trained with a variety of databases with different input parameters and still provide satisfactory results.

The distortions and errors that may occur are due to the impossibility of a comprehensive understanding of the physical behavior of the phenomena addressed by the modeling, among other factors. The limitations of the study are also related to the sensitivity of the results, the quality of the thematic data layers, and the specific characteristics of certain regions susceptible to erosion that may not have been considered (THIERY et al. 2013).

Despite the limitations inherent to the modeling processes, Shahri et al. (2019) list several benefits of these techniques, including the use of publicly available data from satellite images and geological maps. As such, these techniques do not depend on expensive and time-consuming geotechnical investigations. Despite the uncertainty embedded in the maps produced, they can be used as screening tools to identify areas where more detailed investigations should be carried out, in addition to offering a more efficient use of economic and social resources.

CONCLUSIONS

The results show that the two models were able to effectively identify the relationships between the conditioning factors and generate susceptibility maps consistent with the local reality. This performance is based on the area under the curve (AUC), in which the BLR model presented an AUC of 0.888 and the ANN an index of 0.808. The areas susceptible to soil loss represent 57.71% of the study area based on BLR and 54.8% based on ANN.

The greatest potential risks are verified in places with no vegetation cover associated with agricultural practices, and close to the drainage network. Mapping revealed low erosion susceptibility in the southeast of the sub-basin associated with the maintenance of the native forest. Low levels of erosion risk can also be observed in the extreme southwest. These are lowland areas that can flood at certain times of the year. The central area of the sub-basin and the region close to the urban center show moderate susceptibility. In the extreme north of the sub-basin, susceptibility to erosion is related to the conversion of Cerrado vegetation into pasture.

The main advantages of using such models include quicker identification of susceptible areas; the possibility of using freely available data; and different types and combinations of input variables. Applied algorithms can also learn complex patterns and consider nonlinear relationships between dependent and independent variables.

Modeling techniques can provide valuable information for managers to prevent soil erosion, especially in regions with similar landscape characteristics. In addition, they can help to gain a better understanding of erosive processes, their evolution over space and time, and inform strategies for the sustainable use and management of soil and water resources.

REFERENCES

Albuquerque, A.R.C.; Vieira, A.F.S.G. Erosão dos solos na Amazônia. In: Degradação dos solos no Brasil. Guerra, A.J.T.; Jorge, M.C.O.(Org). Rio de Janeiro. Ed. Bertrand Brasil, 230p. 2014.

Al-Najjar, H.H.; Pradhan, B. Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. Geoscience Frontiers 12, 625–637, 2021. https://doi.org/10.1016/j.gsf.2020.09.002.

Al-Shawwa, M.; Abu-Naser, S.S. Predicting Birth Weight Using Artificial Neural Network. International Journal of Academic Health and Medical Research (IJAHMR) Vol. 3 Issue 1, 2019.

Arabameri A. et al. Spatial modelling of gully erosion using evidential belief function, logistic regression, and a new ensemble of evidential belief function–logistic regression algorithm. Land Degrad Dev.; 29:4035–4049. 2018. https://doi.org/10.1002/ldr.3151

Arabameri, A.; Blaschke, T.; Pradhan, B.; Pourghasemi, H.R.; Tiefenbacher, J.P.; Bui, D.T. Evaluation of Recent Advanced Soft Computing Techniques for Gully Erosion Susceptibility Mapping: A Comparative Study. Sensors 2020, 20, 335. https://doi.org/10.3390/s20020335

Arabameri, A. et al. Prediction of gully erosion susceptibility mapping using novel ensemble machine learning algorithms. Geomatics, Natural Hazards and Risk, v. 12, n. 1, p. 469–498, 2021. https://doi.org/10.1080/19475705.2021.1880977

Arruda, W. C.; Lima, H. N.; Forsberg, B. R.; Teixeira W. G. Estimativa de erosão em clareiras através da mudança do relevo do solo por meio de pinos. In: 10 Workshop Técnico-Científico da Rede CT-Petro Amazônia, 2004, Manaus. 02-04 Sept.

Bertoni, J.; Lombardi Neto, F. Conservação do solo. 6.ed. São Paulo: Ícone, 2008. 355p.

Bissacot, A.C.G. Estudo Comparativo entre Regressão Logística Binária e Redes Neurais Artificiais na Avaliação dos Resultados Clássicos de Hosmer, Lemeshow e Sturdivant. (Dissertação) Programa de Pós-Graduação em Engenharia de Produção. 2015. Universidade Federal de Itajubá, Itabujá.

Bragagnolo, L.; Silva, R.V.A.; Grzybowski, J.M.V. Artificial neural network ensembles applied to the mapping of landslide susceptibility. Catena 184, 104240. 2020. https://doi.org/10.1016/j.catena.2019.104240

BRASIL. Departamento Nacional da Produção Mineral. Projeto RADAMBRASIL. Folha SD 20 Guaporé: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra. Rio de Janeiro, 368p, 1979

Campos, A.A.C. Condicionantes dos processos erosivos na área urbana de Buriticupu – MA: o caso da voçoroca do bairro Santos Dumont. Dissertação (Mestrado). 2019. Curso de Pós-Graduação em Geografia, Natureza e Dinâmica do Espaço. Universidade Estadual do Maranhão.

Capoane, V. Determinação do índice de potência de escoamento para o município de Palmitinho/RS utilizando modelos digitais de elevação. Estudos Geográficos, Rio Claro, 13(2): 106-117, 2015.

Carmo, A.M.; et al. Avaliação de suscetibilidade à movimentos de massa, utilizando as variáveis morfométricas, para as serras da porção sul do maciço central do Ceará. R. B. de Cartografia Nº 68/9, 2016.

Carneiro, R.G. et al. Variabilidade da temperatura do solo em função da liteira em fragmento remanescente de mata atlântica. Revista Brasileira de Engenharia Agrícola e Ambiental v.18, n.1, 2014. p.99-108.

Christofoletti, A. Modelagem de Sistemas Ambientais. São Paulo: Edgard Blücher, 256p. 1999.

Conforti, M., Pascale, S., Robustelli, G., Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment. Catena 113, 236–250. 2014. https://doi.org/10.1016/j.catena.2013.08.006.

CPRM – SERVIÇO GEOLÓGICO DO BRASIL. Geologia e Recursos Minerais do Estado de Rondônia: texto explicativo e mapa geológico do Estado de Rondônia, escala 1.1.000.000, Brasília: CPRM. 1999.

CPRM - Serviço Geológico Do Brasil. Residência de porto velho geologia e recursos minerais da folha Pimenteiras (SD.20.X.D). Org. Gilmar José Rizzotto. Porto Velho-Rondônia. 2010.

Ermini, L.; Catani, F.; Casagli, N. Artificial Neural Networks applied to landslide susceptibility assessment. Geomorphology (66), 2005, 327–343. https://doi.org/10.1016/j.geomorph.2004.09.025.

Fonseca, E.L. Processos erosivos em superfícies tabulares com evolução de voçorocamento em áreas de pastagens cultivadas (Braquiária brizantha cv. marandu) no município de Colorado do Oeste – Rondônia. Dissertação. Fundação Universidade Federal de Rondônia – UNIR. Porto Velho. 2017.

Fonseca, E.L.; Locatelli, M.; Silva Filho, E.P, NDVI aplicado na detecção de degradação de pastagens cultivadas, Confins [Online], 35 | 2018, DOI: https://doi.org/10.4000/confins.13180

Fonseca, E.L.; Silva Filho, E.P. Análise fisiográfica como subsídio ao estudo da suscetibilidade erosiva em bacias hidrográficas. ACTA Geográfica, Boa Vista, v.11, n.25. 2017. pp. 137-158. https://doi.org/10.5654/acta.v11i25.4029.

Geron, A. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Published by O'Reilly Media, 2017.

Goyes-Peñafiel, P., & Hernandez-Rojas, A. Doble evaluación de la susceptibilidad por movimientos en masa basada en redes neuronales artificiales y pesos de evidencia. Boletín De Geología, 43(1), 173–191,2021. https://doi.org/10.18273/revbol.v43n1-2021009

Guadagnin, P.M.A.; Trentin, R. Compartimentação geomorfométrica da bacia hidrográfica do arroio Caverá – RS. Geo UERJ. Rio de Janeiro - Ano 16, nº. 25, v. 1, 2014, pp.183-199.

Guerra, A.J.T. O início do processo erosivo. In: Erosão e conservação do solo: conceitos, temas e aplicações. Org. Guerra, A.J.T.; Silva, A.S.; Botelho, R.G.M. 9ª ed., Rio de Janeiro: Berthrand Brasil, 340p. 2014.

Jaafari, A., Najafi, A., Pourghasemi, H.R. et al. GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. Int. J. Environ. Sci. Technol. 11, 909–926 (2014). https://doi.org/10.1007/s13762-013-0464-0

Lee, S.; Hong, S.-M.; Jung, H.-S. A Support Vector Machine for Landslide Susceptibility Mapping in Gangwon Province, Korea. Sustainability 2017, 9, 48. https://doi.org/10.3390/su9010048

Mafra, N.M.C. Erosão e planificação de uso do solo. In: Erosão e conservação do solo: conceitos, temas e aplicações. Org. Guerra, A.J.T.; Silva, A.S.; Botelho, R.G.M. 9ª ed., Rio de Janeiro: Berthrand Brasil, 340p. 2014.

Malhotra, N.K. Pesquisa de marketing: uma orientação aplicada. 7ª ed. Editora: Bookman. 2019.

Meena, S.R., Ghorbanzadeh, O., van Westen, C.J. et al. Rapid mapping of landslides in the Western Ghats (India) triggered by 2018 extreme monsoon rainfall using a deep learning approach. Landslides 18, 1937–1950 (2021). https://doi.org/10.1007/s10346-020-01602-4

Meliho, M., Khattabi, A. & Mhammdi, N. A GIS-based approach for gully erosion susceptibility modelling using bivariate statistics methods in the Ourika watershed, Morocco. Environ Earth Sci 77, 655 (2018). https://doi.org/10.1007/s12665-018-7844-1

Mosavi, A.; Sajedi-Hosseini, F.; Choubin, B.; Taromideh, F.; Rahi, G.; Dineva, A.A. Susceptibility Mapping of Soil Water Erosion Using Machine Learning Models. Water 2020, 12, 1995. https://doi.org/10.3390/w12071995

Nardi, I.R. O desenvolvimento de um modelo matemático para a previsão da aprovação da disciplina de cálculo 1 utilizando regressão logística. Braz. J. of Develop., Curitiba, v. 5, n. 10, 2019.

Perreault, L.M.; Yager, E.M.; Aalto, R. Effects of gradient, distance, curvature, and aspect on steep burned and unburned hillslope soil erosion and deposition. Ear. Surf. Proc. and Landf., 42(7), 1033–1048. 2016. https://doi.org/10.1002/esp.4067

Perusi, M. C.; Carvalho, W. A. Comparação de Métodos para Determinação da Estabilidade de Agregados por Vias Seca e Úmida em Diferentes Sistemas de Uso e Manejo do Solo. Geociências, São Paulo, v. 27, n. 2, 2008, p. 197-206.

Pinheiro, H.S.K. Métodos de mapeamento digital aplicados na predição de classes e atributos dos solos da Bacia Hidrográfica do Rio Guapi. (Tese). 2015. Universidade Federal Rural do Rio de Janeiro. Rio de

Janeiro.

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

Raja, N.B., Çiçek, I., Türkoğlu, N. et al. Landslide susceptibility mapping of the Sera River Basin using logistic regression model. Nat Hazards 85, 1323–1346 (2017). https://doi.org/10.1007/s11069-016-2591-7

RONDONIA, Secretaria de Estado do Planejamento. Plano agroflorestal e Pecuária de Rondônia – PLANAFLORO (bando de dados geográfico). Porto Velho. 2002.

RONDÔNIA. Secretaria de Estado do Desenvolvimento Ambiental (SEDAM). Boletim Climatológico de Rondônia, ano 2008, Porto Velho, 36p. 2010.

Ross, J.L.S. Análise Empírica da Fragilidade dos Ambientes Naturais e Antropizados. In: Revista do Departamento de Geografia n° 8, DG-FFLCH-USP, São Paulo, p. 63-74, 1994.

Roy, J.; Saha, S. Integration of artificial intelligence with meta classifiers for the gully erosion susceptibility assessment in Hinglo river basin, Eastern India. Advances in Space Research, v. 67, n. 1, p. 316–333, 2021.https://doi.org/10.1016/j.asr.2020.10.013

Sarkar, S., Roy, A.K. & Martha, T.R. Landslide susceptibility assessment using Information Value Method in parts of the Darjeeling Himalayas. J Geol Soc India 82, 351–362 (2013). https://doi.org/10.1007/s12594-013-0162-z

Sarkar, T., Mishra, M. Soil Erosion Susceptibility Mapping with the Application of Logistic Regression and Artificial Neural Network. J geovis spat anal 2, 8 (2018). https://doi.org/10.1007/s41651-018-0015-9

Shahri, A.A.; Spross, J.; Johansson, F.; Larsson, S. Landslide susceptibility hazard map in southwest Sweden using artificial neural network. CATENA, Volume 183, 2019. https://doi.org/10.1016/j.catena.2019.104225

Shit, P.K.; Pourghasemi, H.R. Gully Erosion Susceptibility Mapping Based on Bayesian Weight of Evidence. n. January, 2020.

Soma, A. S., & Kubota, T. Landslide susceptibility map using certainty factor for hazard mitigation in mountainous areas of Ujung-loe watershed in South Sulawesi. Forest and Society, 2(1), 79-91, 2018. https://doi.org/10.24259/fs.v2i1.3594

Souza, C.K. et al. Influência do relevo e erosão na variabilidade espacial de um latossolo em Jaboticabal (SP). R. Bras. Ci. Solo, 27. 2003.

Takaki, A.J.H. Caraterização de processos erosivos como instrumento de apoio ao planejamento urbano de Manaus – AM. Dissertação (Mestrado). Manaus: UFAM, 128p. 2002.

Thiery, Y., Maquaire, O.; Fressard, M. Application of expert rules in indirect approaches for landslide susceptibility assessment. Landslides 11, 411–424 (2014). https://doi.org/10.1007/s10346-013-0390-8

Vaezia, S.S. et al. Application of artificial neural networks to describe the combined effect of pH, time, NaCl and ethanol concentrations on the biofilm formation of Staphylococcus aureus. Microbial Pathogenesis 141, 2020. https://doi.org/10.1016/j.micpath.2020.103986.

Vieira, I.C.G.; Jardim, M.A.G.; Rocha, E.J.P. Amazônia em tempo: estudos climáticos e socioambientais. Belém: Universidade Federal do Pará. Embrapa Amazônia Oriental, 462 p. 2014.

Wang, LJ., Guo, M., Sawada, K. et al. A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. Geosci J 20, 117–136 (2016). https://doi.org/10.1007/s12303-015-0026-1

Xing, X., Wu, C., Li, J. et al. Susceptibility assessment for rainfall-induced landslides using a revised logistic regression method. Nat Hazards 106, 97–117 (2021). https://doi.org/10.1007/s11069-020-04452-4

Yavari S.; Maroufpoor, S.; Shiri, J. Modeling soil erosion by data-driven methods using limited input variables. Hydrology Research | 49.5. 2018. https://doi.org/10.2166/nh.2017.041

Author's Affiliation

Fonseca, E.L. - Professor at Federal Institute of Education, Science and Technology of Rondônia, Colorado do Oeste (RO), Brazil.
Filho, E.P.S. - Professor at Federal University of Rondônia, Porto Velho (RO), Brazil.

Authors' Contribution

Fonseca, E.L. - The author proposed the research, collected data, and analyzed the data.
Filho, E.P.S. - The author reviewed the analysis and reviewing the results.

Editors in Charge

Jader de Oliveira Santos
Lidriana de Souza Pinheiro