Gully erosion susceptibility mapping using machine learning method in Pirai drainage basin, Rio de Janeiro State, Brazil

Jorge da Paixão Marques Filho[1], Antonio José Teixeira Guerra[1], Carla Bernadete Madureira Cruz[1] , Maria do Carmo Oliveira Jorge[1]

[1]Department of Geography
Federal University of Rio de Janeiro
CEP. 21044-020
Rio de Janeiro, RJ
Brazil
corresponding author: marquesfilho.j.p@gmail.com

Abstract
Soil erosion is influenced by several controlling factors and causes several environmental problems. Due to the complexity inherent to the occurrence of gullies, it was decided to use the machine learning method, as it deals well with multiple data sources and allows for the reduction of subjectivity. Therefore, this research work aims at mapping the susceptibility to erosion by gullies in Piraí drainage basin. The machine learning method obtained robust results, referring to the metrics used, values > 0.84 and in understanding the behavior of the controlling factors used. The high and very high susceptibility classes occupy ⅕ of the study area (18.94%), and the adoption of conservation practices is recommended.

Keywords
Gully erosion, susceptibility, machine learning, drainage basin

1. Introduction

Soil erosion processes are influenced by precipitation (intensity and amount), slope angle, land use and management and soil properties. The clearance of vegetation, including deforestation, and agricultural activities are determining factors for the occurrence of soil erosion and surface runoff, induced by the rainfall regime (Guerra et al., 2017). Depending on the specificity of the controlling factors, erosion can be categorized as sheet, rill or gully erosion. However, in this research work we will only address gully erosion.

Gully can be defined as an erosive incision, in unconsolidated materials, resulting from the concentration of water flows, following intense rains or ice melting. Furthermore, gullies are erosive features with depth > 0.5 m and cannot be obliterated by agricultural machinery (SSSA, 2018). Gully erosion has been recognized worldwide as one of the most important forms of land degradation. It can be found in rural or urban areas.

In this sense, several national and international authors draw attention to this type of erosive process, which in addition to causing effects in the place where they occur, also, through surface and subsurface runoff, causes silting of water bodies and areas located further downstream (Fullen and Catt, 2004; Poesen et al., 2011).

Phillips (2016) considers that the process in which a given landscape is established is dynamic, unstable, unrepeatable and occurs in a portion of the geographic space. In this sense, understanding natural laws, the portion of geographic space and the history of formation of environmental conditions are ways of understanding the landscape.

Agriculture and pasture are activities that can result in different forms of land degradation, especially when these activities do not take into account the limits imposed by the environments, with regard to gully erosion. As a consequence, in addition to the loss of land for rural activities, there is an increase in river and reservoir silting (Ciccolini et al., 2024; Guerra et al., 2023).

Due to the complexity inherent in the occurrence of gullies, methods such as multi-criteria decision analysis by Geographic Information Systems (GIS-MCDA) (Arabameri et al., 2019) and machine learning (Pourghasemi et al., 2020) have become recurrent. GIS-MCDA can be understood as an approach that contributes to decision-making, combining data and geospatial considerations, according to their importance in understanding this issue (Malczewski, 2006).

However, the subjectivity implicit in GIS-MCDA occurs mainly due to the attribution of weights to the analysis factors and results that are often less accurate, in relation to the machine learning method (Vojtek et al., 2021). Furthermore, empirical models, usually due to the difficulty of dealing with multiple controlling factors, cannot correctly estimate areas susceptible to erosion (Pourghasemi et al., 2020).

Machine learning is an empirical method that uses regression, or classification, and is recommended for problem solving, where theoretical knowledge is not yet consolidated (Lary et al., 2016) and for data analysis. In mapping gully erosion susceptibility, the machine learning method provides satisfactory results, such as prediction and metrics for model evaluation/validation (Chuma et al., 2023; Pourghasemi et al., 2020).

Despite the importance of gully erosion, there has been little effort to develop reliable models for its formation and evolution. Therefore, using measurements of width, depth and length of gullies in the study area, together with the monitoring of several geomorphological features, this research work aims at mapping the susceptibility to erosion by gullies in the Piraí drainage basin.

2. Materials and Methods

2. 1 Study area characterization

Piraí drainage basin, located in Rio de Janeiro State, intersects the municipalities of Barra do Piraí, Engenheiro Paulo Frontin, Mendes, Piraí and Vassouras, totaling 1,019.87 km².

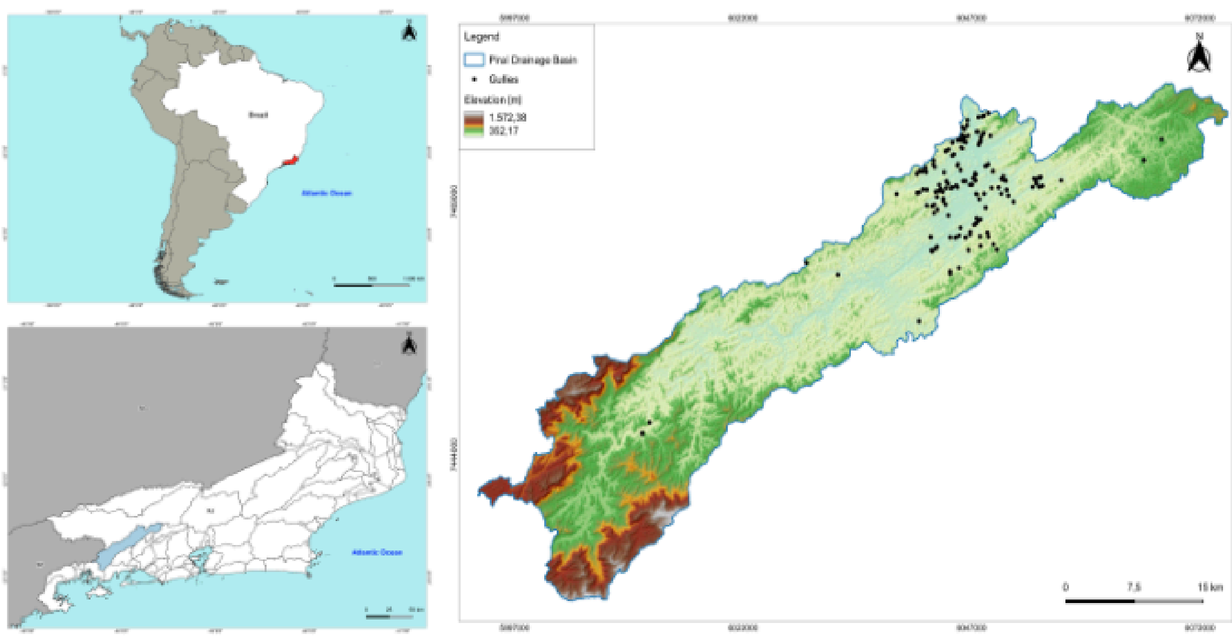

Fig. 1. Pirai drainage basin localization.

Furthermore, it is part of the context of the Depression of the Middle Valley of Paraíba do Sul, which plays an important role as a regional base level. In the Paraíba do Sul-Embu tectonic terrain, the Quirino complex stands out in this compartment, the largest outcropping area, especially in the Shear Zone of the Paraíba do Sul River, with a NE-SW direction. This condition has caused a depressed surface delimited by the Serras da Mantiqueira and do Mar, due to the Cenozoic tectonics that originated the Continental Rift of Southeast Brazil (Riccomini et al., 2004; Dantas et al., 2023).

According to Dantas (2001), the Depression of the Middle Vale Paraíba do Sul is a hemigrában, close to the mountain alignments of Serra da Mantiqueira (Front) and the mountain reverse of Serra do Mar. It is characterized by reliefs of hilly domains, which may include low hills and reliefs of hills and low hills and gentle to medium slopes. In the hilly domain, in the middle Paraíba do Sul valley, active geomorphological processes occur, which are observed by intense rills and gullies, relief inversion, drainage captures and structural concavities (Avellar and Coelho Netto., 1992; Coelho Netto., 2003) .

In relation to the Depression of the Middle Valley of the Paraíba do Sul River, where the Piraí River Hydrographic Basin is contained, Red-Yellow Oxisols, Red-Yellow Argisols, Yellow Oxisols and Yellow Argisols dominate in gentle slopes (Carvalho Filho et al., 2003). The occurrence of a sub-humid tropical climate predominates, with annual rainfall averages between 1,200 and 1,800 mm (Avellar and Coelho Netto, 1992) enabling the development of humid soils (Carvalho Filho et al., 2003; Dantas et al., 2023). Furthermore, the Atlantic Forest biome is predominant and in the study area, isolated forest fragments refer to secondary formations, arising from the abandonment of agricultural areas (Loureiro, 2019).

The settlement of the middle valley of the Paraíba do Sul river was characterized by several economic cycles, such as coffee growing and its subsequent replacement by dairy farming in the XIXth century, promoting soil depletion and accelerated erosion on the slopes, due to changes in the regional hydroclimatic dynamics (Dantas and Coelho Netto, 2018). Dantas (2001) considers that the structural control in the concavities, associated with the subsurface hydrological dynamics and soil use and management, favor the occurrence of erosion processes.

2. 2 Methodological Procedures

2. 2. 1 Gully Inventory Data

The samples from gullies used in the present study were acquired by the Geological Survey of Brazil (Bitar, 2016). However, in the technical report it was not clear which criteria made this differentiation. Although there are several classifications for gully measurements, it was decided to adopt the following criteria: > 0.5m for depth and > 0.5m for width. The main reason is that this classification can be used for both tropical and temperate environments, as several authors highlight in their articles (Barbosa et al., 2024; Ciccolini et al., 2024; Guerra et al., 2017, 2023; Poesen et al., 2011).

Therefore, in Google Earth, the width and depth were measured, using the criteria mentioned above. Among the 234 samples evaluated, 75 had georeferencing problems and were excluded, with only 159 samples being used. Subsequently, samples were generated without the occurrence of gullies, resulting in 159 samples, randomly (Baidah et al., 2023) to balance the previous sample set, totaling 318 samples. In the present article, the samples referring to the non-occurrence of gullies were called non-eroded (0) and with occurrence, eroded (1) and divided between training (70%) and test (30%) samples, according to Bammou et al.. 2024; Baidah et al., 2023; Chuma et al., 2023; Pourghasemi et al., 2020, for the predictive model of gully susceptibility.

2. 2. 2 Multicollinearity in Controlling factors for gully erosion

In studies for susceptibility mapping, it is necessary to consider multicollinearity in the controlling factors, as it may possibly influence the predictive model (Pourghasemi et al., 2020). In the present research study, the following thresholds were adopted, VIF < 10 and Tol > 0.1, adopted by Bammou et al,. (2024) and Chuma et al., (2023).

To prepare the controlling factors, referring to geomorphological and hydrological dynamics, the Forest and Building Removed Digital Elevation Model (Hawker et al., 2022) was used, with a spatial resolution of 30 meters. For hydrological conditioning in the digital elevation model, in the calculation of geomorphometric variables, it was decided to apply the filling method to remove spurious or erroneous depressions (Lindsay and Creed, 2005). In the analysis carried out by Reuter et al. (2009), digital elevation models, which undergo pre-processing, present more satisfactory results.

Geomorphometry is the science that aims to quantify and analyze the Earth's surface (Pike et al., 2009). In the calculation of geomorphometric variables, such as slope angle, profile curvature, plan curvature, topographic humidity index (Pourghasemi et al., 2020), stream power index (Chuma et al., 2023) and specific contribution area, the Whitebox package in R was used.

In the literature, regarding the mapping of susceptibility to erosion by gullies, controlling factors such as lithology, rainfall, distance to highways, distance to rivers (Pourghasemi et al., 2020), land use and land cover (Baidah et al., 2023) and soils (Al-Bawi et al., 2021) have commonly been adopted (Baidah et al., 2023; Pourghasemi et al., 2020). In this sense, for the present research work, the geological map was adopted at scale 1:400,000 (Heilbron et al., 2016), the land use and cover map at scale 1:100,000 (INEA, 2018) and soils map at a scale of 1:250,000 (Carvalho Filho et al., 2003). To calculate the distance to highways and distance to rivers, vector data were used, respectively from road sections and hydrography, at a scale of 1:25,000 (IBGE, 2018).

The selected controlling factors make it possible to understand respectively the geological framework, vegetation cover and soil management practices, soil properties, influence of river and anthropic dynamics, considering the adopted scale influence susceptibility to gully erosion.

In terms of rainfall, WorldClim climate data (Fick and Hijmans, 2017) version 2.1, was used, which has 1km2 spatial resolution, with a historical series between 1970-2000, and the average rainfall was subsequently calculated using map algebra. Subsequently, the data were resampled to a spatial resolution of 30 meters, and the data in vector format were also converted to matrix format, at the same resolution. These procedures were carried out to make data resolutions compatible in the predictive model.

The historical rainfall series is necessary to understand the constancy of Pirai Drainage Basin rainfall regime and its relationship and impact on gully erosion susceptibility.

2. 2. 3 Machine Learning Model

Witten et al., (2018) point out that the decision tree is a driving force for several applications, quickly, and that presents significantly accurate results. Maxwell et al., (2021) consider that the metrics for evaluating the predictive machine learning model using binary confusion matrix, take into account the relationship between true and false positives and negatives.

Random Forest machine learning algorithm is a non-parametric and random classifier, which consists of classifiers structured in decision trees. Attributes are randomly distributed and each decision tree sends its unit's "vote" to differentiate a class (Breiman, 2001). Random Forest model was chosen as it presents the most robust performance in gully erosion susceptibility models (Were et al., 2023, Huang et al., 2023, Pourghasemi et al., 2020, Gayen et al., 2019).

According to Maxwell et al. (2021), define that true positives or negatives correspond to the labeling, identically between the real and predicted classes. False positives correspond to the erroneous or incorrect classification of the predicted classes, while in false negatives, a class is wrongly assigned for subsequent classification.

Among the most commonly used metrics in gully erosion susceptibility mapping, accuracy, precision, recall, F1-Score and ROC Curve stand out. Accuracy corresponds to the degree of measurement between modeling and reality, precision refers to the proportion between true positives and sums that were predicted as positive in the model. Recall is similar to precision, however it only considers the sum of truly positive data and F1-Score is the harmonic mean between precision and recall, that is, if F1-Score has higher values, the more robust is the relationship between the two metrics (Sammut and Webb, 2017). The ROC Curve is defined as the relationship between sensitivity and specificity, where the first is the proportion of positives classified correctly and the second, the proportion of negatives, classified correctly (Sammut and Webb, 2017).

Furthermore, it is possible to identify the most significant contributions of the most significant factors or variables in the model prediction. The Gini Index calculates the impurity of the data, the probability of occurrence in the classification, the labeling of classes. Thus, the Gini Index is fundamental for identifying the importance of variables and the higher the value obtained by the factor, the greater its contribution to the predictive model and vice versa (Breiman, 2001).

When using this model, the following hyperparameters were considered: number of trees (82), mTry (4) and Target Node Size (1) for the classification between the Non-Eroded (0) and Eroded (1) labels in the model. The choice of these values was based and adapted to the methodological proposal developed by MapBiomas (2023) and after the classification, a probability matrix was generated for the recognition of geospatial data patterns of susceptibility to gully in Random Forest, with the Natural Jenks method . Natural breaking is an approach that groups similar data and accentuates the differentiation between them (Smith et al., 2024). The application of the Random Forest machine learning algorithm and the metrics to evaluate its performance were done in R with the randomForest and pROC packages. The Random Forest algorithm was chosen as it presents the most robust performance in gully erosion susceptibility models (Bammou et al., 2024; Pourghasemi et al., 2020).

2.2.5 Minimum Mapping Unit

Due to the multiplicity of data scales of the controlling factors and ensuring the reliability of the proposed mapping scale, the procedure was carried out to identify the most appropriate minimum mappable area. According to Buckley (2008), the minimum mapping unit is the smallest feature that can be captured when imaging by remote sensing.

For this, we used the matrix cartographic generalization proposed by Spínola et al., (2011), which suggests the calculation between the ratio of the minimum mappable area (40 ha) referring to the mapping scale, 1.100,000, and the size of the pixel area, 900 m². Enabling the appropriate choice of the movable window, in this case 6x6, as it is close to the result of the number of pixels relative to the previously mentioned ratio, 44.4 pixels.

3. Results

3.1 Multicollinearity in Controlling factors for gully erosion

In Table 1, the results on the VIF and Tol statistical tests to identify correlations between the controlling factors were performed. Among the 13 controlling factors, according to the data below, no significant correlation or connections were identified in the variables used to map susceptibility to erosion by gullies.

Table 1. Multicollinearity in Controlling factors for gully erosion.

Controlling Factors	VIF	TOL
Elevation	4.261653	0.2346507
Slope Angle	4.241980	0.235739
Profile Curvature	1.170134	0.854603
Plan Curvature	1.058737	0.9445216
Topographic Wetness Index	3.002529	0.3330526
Specific Contributing Area	3.684414	0.2714136
Stream Power Index	4.350771	0.2298443
Lithology	1.037845	0.963535
Land Use and Land Cover	1.169974	0.8547198
Rainfall	2.708888	0.3692642
Distance to Rivers	1.169855	0.8548068
Distance to Roads	1.273714	0.7851056
Soils	1.913443	0.5226181

3.2 Performance of Machine Learning Model

Table 2 and Figure 3 show the results regarding the metrics used to evaluate the predictive model for gully erosion susceptibility mapping. To evaluate the model's performance, we chose to use five performance metrics, Accuracy, Precision, Recall, F1-Score and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

In the Random Forest machine learning algorithm, the relative robustness over the classes referring to non-eroded (0) and eroded (1) was noticeable. By evaluating these metrics, it is possible to understand the performance of the model, for subsequent improvement. In this sense, it is clear that Accuracy in both classes is identical, with a noticeable difference between Precision, Recall and F1-Score. The Random Forest model, in relation to AUC-ROC, obtained a value of 91.5% in AUC-ROC, representing an excellent predictive capacity.

Table 2. Model performance based on performance metrics.

Metrics	Non-Eroded	Eroded
Accuracy	0.9148936	0.9148936
Precision	0.893617	0.9361702
Recall	0.9333333	0.8571429
F1-Score	0.9130435	0.8949153

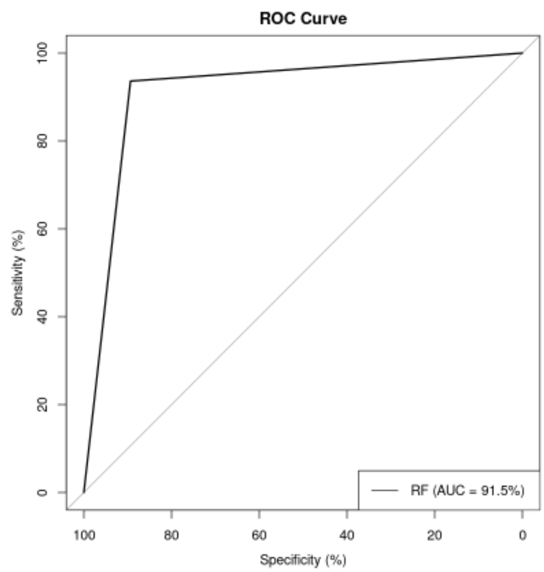

Fig. 2. ROC Curve analysis.

3.3 Variables Importance

The result refers to the contribution of the controlling factors to the prediction of the Random Forest machine learning algorithm (Fig. 4.) shows that the 13 selected variables were considered in the predictive model.

In this case, four factors stood out from the aforementioned set, Elevation (20.6991985), Land Use and Cover (16.6784218), Rainfall (15.7114718) and Slope (15.4537312). The other controlling factors like Stream Power Index (9.5436686), Profile Curvature (7.0812813), Topographic Wetness Index (5.9981457), Plan Curvature (5.7461969), Soils (3.9533505), Distance to Roads (3.5455912), Lithology (3.3660106), Distance to Rivers (2.9281166) and Specific Contributing Area (0.8787656) did not reach values greater than 10.

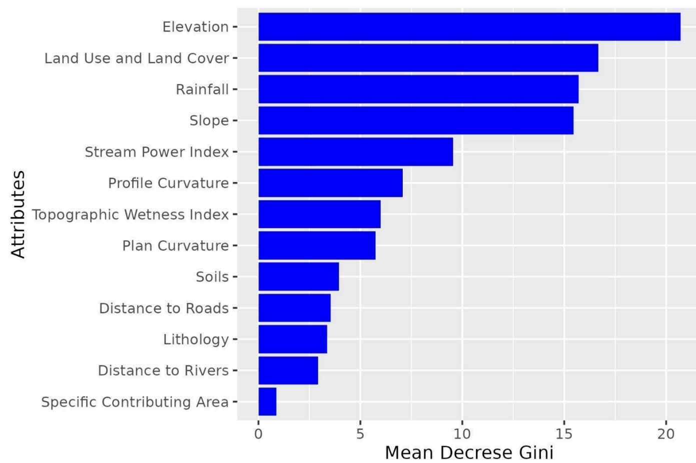

Fig. 3. Contribution of controlling factors to mapping susceptibility to gully erosion.

3.3 Gully Erosion Susceptibility Mapping

In recognizing the geospatial patterns of susceptibility classes (Fig. 4), the Natural Jenks method (Longley et al., 2024) was used to identify the areas most prone to gullying (Table 3). The ranges defined by the natural break for the susceptibility classes were: 0 - 0.10 (very low), 0.10 - 0.25 (low), 0.25 - 0.45 (intermediate), 0.45 - 0.69 (high) and 0.69 - 1 (very high).

The very low (0 - 0.10) (44.59%) and low (31.33) susceptibility classes are predominant in the study area, totaling 75.92%. The intermediate susceptibility class corresponds to 5.14% and is the least representative in Piraí River drainage basin. Among the classes corresponding to high (11.75%) and very high (7.19%) susceptibility, it is clear that in total, 18.94%, or approximately ⅕ of the study area is susceptible to gullies.

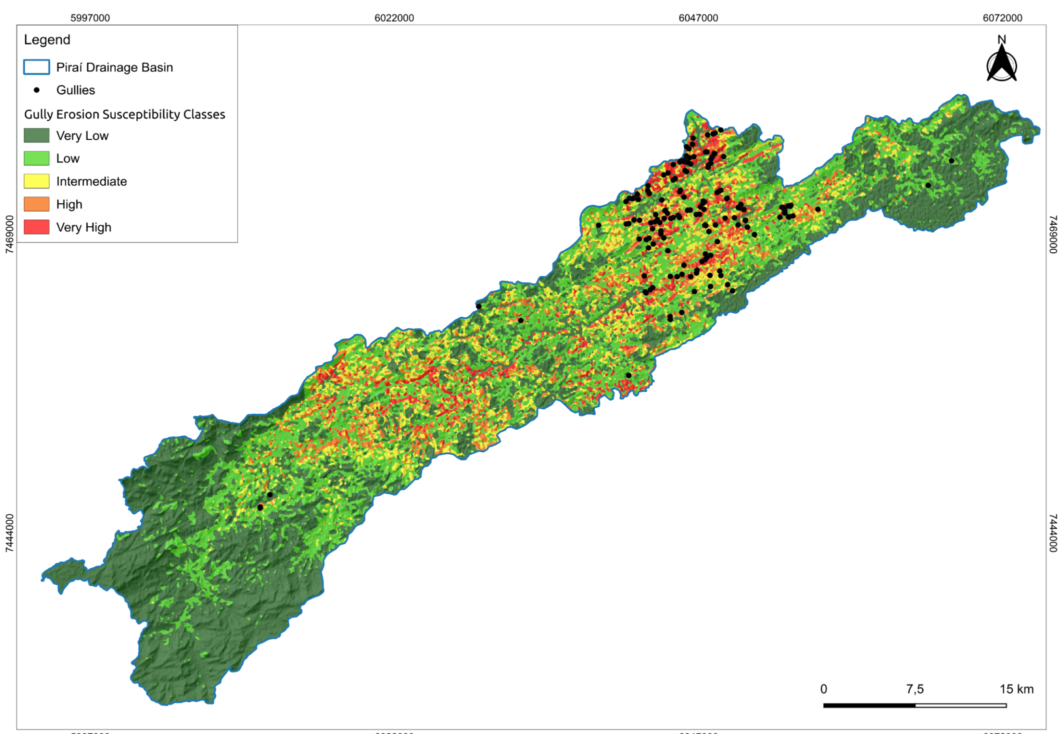

Fig. 4. Gully erosion susceptibility mapping in Piraí drainage basin.

Table 3. Percentage of susceptibility classes to erosion by gullies.

Susceptibility Classes	Area	Percent (%)
Very Low	454.73	44.59
Low	319.57	31.33
Intermediate	52.38	5.14
High	119.83	11.75
Very High	73.36	7.19

4. Discussion

4.1. Performance of Machine Learning Model

The non-eroded and eroded classes (Table 2) presented identical and satisfactory accuracy results, as it is an assessment that measures the relationship between model and reality, highlighting the degree of representativeness of the modeling (Sammut and Webb, 2017). Comparing the two classes, based on the Precision metric, it is understood that the proportion of truly positive data and the sum of positive predictions in the model in the present research work were satisfactory in the eroded class (Sammut and Webb, 2017).

However, in relation to the Recall metric, the lower performance of the eroded class compared to the non-eroded class is noted, as this metric only considers the sum of truly positive and unpredicted data (Sammut and Webb, 2017). Therefore, highlighting a slight discrepancy between the predicted and truly real data, as indicated by the two metrics. Furthermore, the F1-Score metric, which evaluates the harmonic mean between Precision and Recall, shows that the prediction of positive and truly positive data is robust and considered satisfactory. However, as it is a more sensitive metric, to lower values, it highlights the discrepancy between predicted and truly real data (Sammut and Webb, 2017). The ROC curve in the present model achieved adequate performance, as it considers the proportion of truly positive and negative data in the classification, highlighting the reliability of the predictive model (Sammut and Webb, 2017) for mapping susceptibility to gully erosion.

4.2. Variables Importance

It is possible to highlight that the controlling factors for gully erosion such as Elevation, Land Use and Land Cover, Rainfall and Slope Angle were the variables that contributed most to the susceptibility model. However, it is clear that the Stream Power Index, Profile Curvature and Topographic Wetness Index moderately corroborate the model of susceptibility to erosion by gullies. However, Plan Curvature, Soils, Distance to Roads, Lithology, Distance to Rivers and Specific Contributing Area did not contribute significantly to the modeling.

Elevation, as well as in studies carried out by Chuma et al., (2023) and Pourghasemi et al., (2020) was significant in the gully erosion susceptibility model. Two other factors that corroborate the present study are Rainfall and the Normalized Difference Vegetation Index, which can be interpreted as an analogy to land use and cover. However, factors such as distance to highways (Chuma et al., 2023; Pourghasemi et al., 2020) and distances to rivers (Pourghasemi et al., 2020) played more fundamental roles in these researches, than in the present study, a reverse situation, that occurs with slope, highlighting the importance of geomorphometric variables in this research work.

In this sense, understanding the dynamics of the landscape, the perspective of natural laws, and understanding the history of formation of environmental conditions over the portion of geographic space (Phillips, 2016), are essential for choosing the specificity of controlling factors for the gully erosion susceptibility model.

It is possible to relate that the altimetric variability in the Piraí drainage basin, conditioned by the historical degradation in the occupation of land use and cover, the spatial distribution of rainfall and the inclination of the relief forms, definitely contribute to the process of origin of the gullies (Guerra et al., 2017; Dantas and Coelho Netto, 2018). As well as, secondarily, the river drainage incision capacity, the geometry or shapes of the slopes and the surface runoff corroborate the hypothesis formulated by Dantas (2001) about the occurrence of gullies in Piraí drainage basin.

4.3. Gully Erosion Susceptibility Mapping

Considering the very low to low susceptibilities classes, the main occurrences identified are located in the upper course of the Piraí River watershed and in areas of isolated forest fragments, due to secondary vegetation of the resilience of abandoned pastures (Loureiro, 2019).

The intermediate susceptibility classes are normally close to the high and very high ones, located in the middle and mainly in the lower reaches of the study area. Although high and very high susceptibility correspond to approximately ⅕ of the territorial extension of Piraí drainage basin, they predominantly refer to pastures and degraded areas, and in features with low altimetric amplitude, hills and gentle to medium slopes (Dantas, 2001).

5. Conclusions

Mapping susceptibility to erosion by gullies using machine learning allows for multiple data, from different sources, referring to controlling factors, as opposed to only empirical models and the reduction of subjectivity of these classes. In this sense, the appropriate choice of controlling factors, based on the literature review and/or understanding the environmental dynamics, is essential to establish the development of the appropriate predictive model. It is therefore recommended to choose the controlling factors based on these two criteria.

The gully erosion susceptibility model presented evaluation metrics greater than 0.84, highlighting the robustness of this modeling as well as the Random Forest machine learning algorithm itself. Furthermore, the interpretation of the importance of the variables corroborates the current literature regarding the occurrence of gullies, as well as the history of environmental conditions in the study area.

Through the spatialization of susceptibility class patterns, it was possible to identify that ⅕ or 18.94% of the study area is located in susceptible areas between high and very high, recommending the adoption of soil conservation practices. Since soil degradation, as well as gully occurrence, originates from the inadequate use of the history of land use and settlement, as indicated in the literature and in this modeling.

Acknowledgements: The authors thank FAPERJ (Rio de Janeiro Funding Agency) and CNPq (Brazilian Funding Agency) for the financial support. The authors also thank Ms. Ana Karolina dos Santos Tome and Ms. Thainara Cordeiro Pereira for measuring the gullies.

References

Al-Bawi, A. J., Al-Abadi, A. M., Pradhan, B., & Alamri, A. M. (2021). Assessing gully erosion susceptibility using topographic derived attributes, multi-criteria decision-making, and machine learning classifiers. *Geomatics, Natural Hazards and Risk*, 12(1), 3035–3062. https://doi.org/10.1080/19475705.2021.1994024

Avelar, A. S., & Coelho Netto, A. L. (1992). Fraturas e desenvolvimento de unidades geomorfológicas côncavas no Médio Vale do Rio Paraíba do Sul. *Revista Brasileira de Geociências*, 22(2), 222- 227.

Arabameri, A., Cerda, A., & Tiefenbacher, J. P (2019). Spatial Pattern Analysis and Prediction of Gully Erosion Using Novel Hybrid Model of Entropy-Weight of Evidence. *Water*, 11(6), 1129. https://doi.org/10.3390/w11061129

Baiddah, A., Krimissa, S., Hajji, S., Ismaili, M., Abdelrahman, K., El Bouzekraoui, M., Eloudi, H., Elaloui, A., Khouz, A., Badreldin, N., & Namous, 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 Sciences*, 11, 1184038. https://doi.org/10.3389/feart.2023.1184038

Bammou, Y., Benzougagh, B., Abdessalam, O., Brahim, I., Kader, S., Spalevic, V., Sestras, P., Ercisli, S. Machine learning models for gully erosion susceptibility assessment in the Tensift catchment, Haouz Plain, Morocco for sustainable development. *Journal of African Earth Sciences,* 213, 105229. https://doi.org/10.1016/j.jafrearsci.2024.105229

Barbosa, W. C. S., Guerra, A. J. T., & Valladares, G. S. (2024). Soil erosion modeling using the revised universal soil loss equation and a geographic information system in a watershed in the northeastern Brazilian Cerrado. *Geosciences*, 14(3), 78. https://doi.org/10.3390/geosciences14030078

Bitar, O. Y. (2016). *Cartas de suscetibilidade a movimentos gravitacionais de massa e inundações – 1:25.000 nota técnica explicativa*. IPT – Instituto de Pesquisas Tecnológicas do estado de São Paulo and CPRM – Serviço Geológico do Brasil.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Buckley, A. R. (2008). Minimum Mapping Unit (MMU). Kemp, K. K. *Encyclopedia of Geographic Information Science* (pp. 287-288). SAGE Research Methods.

Ciccolini, U., Buffalini, M., Materazzi, M., & Dramis, F. (2024). Gully erosion development in drainage basins: A new morphometric approach. *Land*, 13 (6), 792. https://doi.org/10.3390/land13060792

Chuma, G. B., Mugumaarhahama, Y., Mond, J. M., Bagula, E. M., Ndeko, A. B., Lucungu, P. B., Katume, K., Mushagalusa, G. N., & Schmitz, S. (2023). Gully erosion susceptibility mapping using four machine learning methods in Luzinzi watershed, eastern Democratic Republic of Congo. *Physics and Chemistry of the Earth, Parts A/B/C* 123, 103295. https://doi.org/10.1016/j.pce.2022.103295

Carvalho Filho, A., Lumbreras, J. F., Wittern, K. P., Lemos, A. L., Santos, R. D., Calderano Filho, B., Oliveira, R. P., Aglio, M. L. D., Souza, J. S., Chaffin, C. E., Mothci, E. P., Larach, J. O. I., Conceição, M., Tavares, N. P., Santos, H. G., Gomes, J. B. V., Calderano, S. B., Goncalves, A. O., Martorano, L. G., Barreto, W. O., Claessen, M. E. C., Paula, J. L., Souza, J. L. R., Lima, T. C., Antonello, L. L., & Lima, P. C. (2003). *Levantamento de reconhecimento de baixa intensidade dos solos do Estado do Rio de Janeiro*. Embrapa Solos.

Coelho Netto, A. L. (2003). Evolução de cabeceiras de drenagem no médio Vale do Rio Paraíba do Sul (SP/RJ): bases para um modelo de formação e crescimento da rede de canais sob controle estrutural. *Revista Brasileira de Geomorfologia*, 4(2), 69-100.

Dantas, M. E., Ferreira, C. E. O., & Shinzato, E. (2023). Relevo. Pereira, M. G., Anjos, L. H. C., & Neto, E. C. S. *Solos do Rio de Janeiro - Gênese, classificação e limitações ao uso agrícola* (pp. 19-59). Editora Atena.

Dantas, M. E., Coelho Netto, A. L. (2018). A denudação antropogênica da paisagem: processos erosivo deposicionais no médio Vale do Rio Paraíba do Sul. Oliveira, R., Lazos, A. *Geografia histórica do café no vale do rio Paraíba do Sul* (pp. 107-126). Editora Puc-Rio.

Dantas, M. E. (2001). Geomorfologia do estado do Rio de Janeiro. CPRM – Serviço Geológico do Brasil. *Projeto Rio de Janeiro* (95-195). CPRM – Serviço Geológico do Brasil.

Fick, S. E. & Hijmans, R. J. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology,* 37 (12), 4302-4315. https://doi.org/10.1002/joc.5086

Fullen, M. A., & Catt, J. A. (2004). *Soil management – problems and solutions*. Oxford University Press.

Gayen, A., Pourghasemi, H. R., Saha, S., Keesstra, S., & Bai, S. (2019). Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Science of The Total Environment*, 668, 124-138. https://doi.org/10.1016/j.scitotenv.2019.02.436

Guerra, A. J. T., Bezerra, J. F. R., & Jorge, M. C. O. (2023). Recuperação de voçorocas e de áreas degradadas, no Brasil e no mundo - estudo de caso da voçoroca do Sacavém - São Luís – MA. *Revista Brasileira de Geomorfologia*, 24 (00). https://doi.org/10.20502/rbg.v24i00.2306

Guerra, A. J. T., Fullen, M. A., Jorge, M. C. O., Bezerra, J. F. R., & Shokr, M. S. (2017). Slope processes, mass movement and soil erosion: a review. *Pedosphere*, 27 (1), 27–41. https://doi.org/10.1016/S1002-0160(17)60294-7

Hawker, L., Uhe, P., Paulo, L., Sosa, J., & Savage, J.; Sampson, C.; Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17(2), 024016. https://doi.org/10.1088/1748-9326/ac4d4f

Heilbron, M., Almeida, J. C. H., & Eirado, L. G. *Geologia e recursos minerais do estado do Rio de Janeiro*. CPRM – Serviço Geológico do Brasil.

Huang, D., Su, L., Zhou, L., Tian, Y., & Fan, H. (2023). Assessment of gully erosion susceptibility using different DEM-derived topographic factors in the black soil region of Northeast China. *International Soil and Water Conservation Research*, 11. 97-111. https://doi.org/10.1016/j.iswcr.2022.04.001

Instituto Brasileiro de Geografia e Estatística. (2018). *Base comum vetorial do estado do Rio de Janeiro versão 2018 edgv 3.0.* https://www.ibge.gov.br/geociencias/downloads-geociencias.html?caminho=cartas_e_mapas/bases_cartograficas_continuas/bc25/rj/versao2018_edgv_3.0/

Instituto Estadual do Ambiente. (2018). Mapeamento de uso do solo e cobertura vegetal 2018. https://www.arcgis.com/home/item.html?id=634cd01b3473409c96d87bb41caeb87e

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geoscience and remote sensing. *Geoscience Frontiers*, 7(1), 3-10. https://doi.org/10.1016/j.gsf.2015.07.003

Lindsay, J. B., & Creed, I. F. (2005). Removal of artefact depressions from DEMs: towards a minimum impact approach. *Hydrological Processes* ,19 (16), 3113–3126. https://doi.org/10.1002/hyp.5835

Loureiro, H. A. S. (2019). *Diagnóstico de erosão por voçorocas: experimentos com geotecnologias e solos na bacia do Alto rio Piraí - Rio Claro-RJ*. Thesis (Doutorado em Geografia – Planejamento e Gestão Ambiental) – Programa de Pós Graduação em Geografia, Instituto de Geociências, Universidade Federal do Rio de Janeiro.

Malczewski, J. (2006). GIS‐based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, 20(7), 703–726. https://doi.org/10.1080/13658810600661508

Maxwell, A. E., Warner, T. A., & Guillén, L. A. (2021). Accuracy assessment in convolutional neural network-based Deep Learning remote sensing Studies — part 1: literature review. Remote Sensing, 13(13), 2591. https://doi.org/10.3390/rs13132591

MapBiomas. (2023). Mapeamento anual do estoque de carbono orgânico do solo no Brasil 1985-2021 (coleção beta). https://doi.org/10.58053/MapBiomas/DHAYLZ

Nelson, A., Reuter, H. I., Gessler, P., & Soille, P. (2009). Preparation of DEMs for geomorphometric analysis. Hengl, T., & Reuter, H. I. *Geomorphometry: concepts, software, applications* (pp. 65-86). Elsevier. https://doi.org/10.1016/S0166-2481(08)00004-4

Pike, R. J., Evans, I., & Hengl, T. (2009). Geomorphometry: a brief guide. Hengl, T., &Reuter, H. I. *Geomorphometry: concepts, software, applications* (pp. 3-30). Elsevier. https://doi.org/10.1016/S0166-2481(08)00001-9

Phillips, J. D. (2016). Laws, place, history and the interpretation of landforms. *Earth Surface Processes and Landforms*,, 42(2), 347–354,. https://doi.org/10.1002/esp.4083

Poesen, J., Torri D., Vanwalleghem, T. (2011). Gully erosion: procedures to adopt when modelling soil erosion in landscapes affected by gullying. Morgan, R., Nearing, M. *Handbook of Erosion Modelling* (pp. 360-386). Blackwell Publishing Ltd.

Pourghasemi, H. R., Sadhavisam, N., Kariminejad, N., & Collins, A. L. Gully erosion spatial modelling: Role of machine learning algorithms in selection of the best controlling factors and modelling process. *Geosciences Frontiers*, 11(6) , 2207-2019. https://doi.org/10.1016/j.gsf.2020.03.005

R Core Team (2023). *R: a language and environment for statistical computing, Version 4.3.1*. R Foundation for Statistical Computing, Vienna.

Riccomini, C., Sant'anna, L. G., & Ferrari, A. L. (2004). Evolução geológica do Rift continental do sudeste do Brasil. Mantesso-Neto, V., Bartorelli, A., Carneiro, C. D. R., & Brito-neves, B. B. *Geologia do continente Sul-Americano: evolução da obra de Fernando Flávio Marques de Almeida* (pp. 383-405). Editora Beca.

Sammut, C., & Webb, G. I. (2017). *Encyclopedia of machine learning and data mining* (2nd ed.). Springer.

Smith, M. J., Goodchild, M. F., & Longley, P. A. (2024). *Geospatial Analysis: A Comprehensive Guide to Principles Techniques and Software Tools* (7[t]th end). https://www.spatialanalysisonline.com/extractv7.pdf

Soil Science Society of America. (2008). *Glossary of soil science terms*. Soil Science Society of America.

Spínola, D. N., Filho, E. I. F., Portes, R. C., Resck, B. C. (2011). Comparação entre dois métodos de generalização cartográfica semi-automática em ambiente matricial. *Anais XV Simpósio Brasileiro de Sensoriamento Remoto*, 2294-2301.

Were, K., Kebeney, S., Churu, H., Mutio, J. M., Njoroge, R., Mugaa, D., Alkamoi, B., Ng'etich, W., & Singh, B. R. (2023). Spatial prediction and mapping of gully erosion susceptibility using machine learning techniques in a degraded semi-arid region of Kenya. *Land*, 12, 890.

Witten. H., Frank, E., & Hall, M. A. (2018). *Data mining practical machine learning tools and techniques* (4th ed.) Morgan Kaufmann Publishers.

Vojtek, M., Vojteková, J., Costache, R., Pham, Q. B., Lee, S., Arshad, A., … Anh, D. T. (2021). Comparison of multi-criteria-analytical hierarchy process and machine learning-boosted tree models for regional flood susceptibility mapping: a case study from Slovakia. *Geomatics, Natural Hazards and Risk*, 12(1), 1153–1180. https://doi.org/10.1080/19475705.2021.1912835