VNU Journal of Science: Earth and Environmental Sciences, Vol. 40, No. 4 (2024) 97-104

VNU Journal of Science: Earth and Environmental Sciences

Journal homepage: https://js.vnu.edu.vn/EES

Original Article

Predict the Riverbank Erosion Susceptibility for the Ham Luong River Using the Logistic Regression Model

Le Van Quyen, Luc Anh Tuan, Doan Van Binh*

Vietnamese-German University, Ben Cat, Binh Duong, Vietnam

Received 27 August 2024
Revised 26 October 2024; Accepted 20 December 2024

Abstract: Riverbank erosion, once considered an inevitable process, has emerged as a severe and unpredictable issue, exacerbated by climate change and human activities. The Vietnamese Mekong Delta (VMD) has faced significant erosion, leading to considerable infrastructural damage and economic losses. This study aims to predict riverbank erosion susceptibility along the Ham Luong River using a logistic regression (LR) model for 100 riverbank locations, classified as eroded or stable, with twelve conditioning variables. The LR model achieved an overall accuracy (ACC) of 0.83 and an area under the receiver operating characteristic curve (AUC) of 0.86 through 15-fold cross-validation. Sensitivity analysis identified the bank slope, soil moisture, and bank height as key factors influencing erosion susceptibility. Probability analysis revealed that an increased bank slope may cause greater riverbank instability, whereas higher soil moisture levels may reduce erosion susceptibility. These findings highlight the importance of stabilizing bank slopes and maintaining soil moisture to mitigate the riverbank erosion susceptibility effectively, emphasizing the need for managing water levels in rivers and canals during the dry season as part of disaster risk management in the VMD.

Keywords: Riverbank erosion, Susceptibility, Probability, Logistic Regression, Machine learning.

1. Introduction

Considered an inevitable process along river courses, riverbank erosion has become more severe and unpredictable nowadays. Among various drivers, this natural hazard has been aggravated by climate change and anthropogenic activities, such as sand mining and deforestation [1, 2]. In the Vietnamese Mekong Delta (VMD), there have been many erosion locations recorded recently. Notably, bank collapse events in the Vam Nao Channel in 2017 and the Co Chien

* Corresponding author.
E-mail address: binh.dv@vgu.edu.vn

https://doi.org/10.25073/2588-1094/vnuees.5234

River in Vinh Long in 2022 caused extreme damage to infrastructures (i.e., roads, and houses). Unfortunately, people have passively reacted to this phenomenon, resulting in serious losses in economic and, possibly, human lives.

In recent years, numerous researchers have investigated riverbank erosion in the VMD. Hung et al., (2006) used satellite imagery over eight years combined with empirical models to predict erosion, demonstrating improvements in accuracy and contributing to risk reduction for both lives and property [3]. Similarly, Khoi et al., (2020) assessed changes in the VMD's riverbanks from 1989 to 2014, using the Modified Normalized Difference Water Index (MNDWI) for water body detection [4]. While effective, the study faced challenges with image resolution. Binh et al., (2020) advanced this work by applying multiple indices, including Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and MNDWI, showing that MNDWI and NDWI outperformed other indices for tracking riverbank erosion, while NDVI was less suitable for this purpose [5]. These studies underscore the potential of integrating remote sensing (RS) techniques to improve erosion monitoring in the VMD. However, there remains a significant gap in research addressing riverbank erosion susceptibility (RBES), particularly in integrating remote sensing techniques with robust models for comprehensive risk assessment.

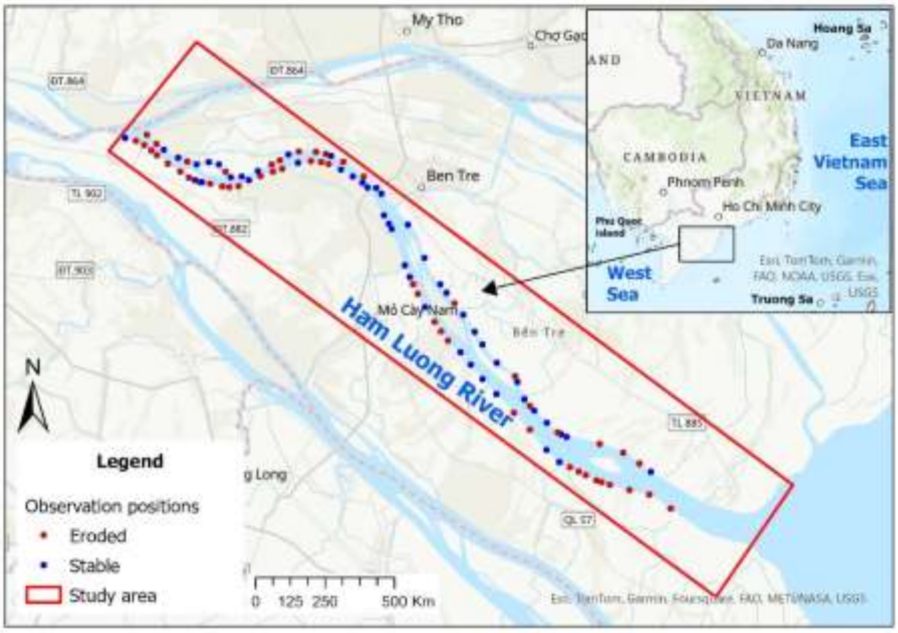

Figure 1. The Ham Luong River with observed locations of riverbank status and controlling factors.

Disaster susceptibility has been studied by researchers around the world. Nevertheless, they usually focus on gully erosion [6], soil erosion [7], and flooding [8]. These studies applied advanced techniques such as machine learning and artificial intelligence and achieved many valuable results. Azlinda et al. (2020) found that the NARX-QR Factorization model, integrating hydraulic and soil properties, enhanced predictions but required complex inputs [9]. Garosi et al., (2019) identified Random Forest (RF) as the best model for gully erosion mapping, outperforming the Support Vector Machine (SVM), Naïve Bayes, and Generalized Additive model (GAM), while Rahmati et al., (2017) confirmed RF, RBF-SVM, and Boosted

Regression Trees (BRT) as highly reliable [6]. However, BRT is computationally demanding, though it excels in soil erosion prediction compared to simpler models like Multiple Linear Regression (MLR) [7]. Meanwhile, logistic regression (LR) is a model that requires less complex input data and has been utilized in numerous studies, demonstrating very promising performance [10].

Although machine learning models have demonstrated effectiveness in predicting erosion, research on riverbank erosion susceptibility remains underdeveloped, particularly in the VMD. Developing a robust RBES predictive model is essential to mitigate erosion impacts. This requires a detailed understanding of the likelihood that a specific riverbank section will erode or collapse, a challenge that demands advanced methodologies and extensive datasets.

Therefore, this research selected the Ham Luong River, one of the primary branches of the VMD, as a study area (Figure 1). To understand the controlling factors of riverbank erosion behaviour in this region, this study aims to apply the LR model to i) Predict riverbank erosion susceptibility; ii) Determine key driving feature variables for prediction; and iii) Determine the thresholds of those variables for the riverbank to change from erosion to accretion and vice versa.

2. Data and Methods

The study was implemented through three primary stages, namely data collection and processing, model training and evaluation, and sensitivity-probability analysis. Each stage includes several tasks as illustrated in Figure 2.

2.1. Data Collection and Processing

The first working stage aimed to prepare the dataset through data collection and processing. Initially, we conducted a field survey (May 10th – 20th, 2023) along the Ham Luong River to examine the current status of its banks. As a result, information at 100 positions was recorded and separated into two types of bank status (50 eroded and 50 stable). These locations were chosen based on the historical riverbank evolution [11]. Erosion's driving factors were also observed, i.e., geomorphological and hydrological factors. Notably, the near bank flow velocity was derived from the numerical model simulation work of the "Nature-based Solutions for Food Security under Climate Change Effects for Sustainable Development in the Mekong Delta" (NbS-SUCCEED) project. Additionally, we collected soil samples in such locations to examine the soil physical properties. These samples were stored and analyzed (i.e., sample analysis, SA) in the laboratory in accordance with Vietnamese standards. Such factors were selected due to their recognized importance in estimating riverbank erosion rates through alternative methods, including empirical modeling approaches [10]. Consequently, a training dataset was generated with 12 features (Table 1).

2.2. Model Training and Evaluation

As one of the most influential analytic techniques in the social and natural sciences, LR is marvelously suited for discovering the link between features and particular outcomes [12]. This probabilistic classifier makes use of supervised machine learning, with its name originating from the logistic function:

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (1)$$

$$z = \left(\sum_{i=1}^{n} w_i x_i\right) + b \quad (2)$$

In Equation (1), the logistic function $\sigma(z)$ is used to create probabilities (i.e., 0 to 1) based on values of the logits z (i.e., -∞ to +∞). Meanwhile, w_i is the weight of the corresponding input feature x_i, and b is the bias term or the intercept. Essentially, a factor with a higher weight is considered more important. These weight and intercept values allow us to establish a linear line represented by Equation (2). Besides, locations with $\sigma(z)$ values larger than 0.5 will be classified as eroded points, and those smaller than 0.5 will be classified as stable positions.

Based on the prepared dataset, the LR model was trained and tested. The applied parameters for the LR were selected through a hyper-parameter tuning process, using the grid search method. To increase the accuracy of the training task, 15-fold cross-validation was used. We used overall accuracy (ACC) and the area under the receiver operating characteristic curve (AUC) to report the performance of the predictive model. The method to calculate these statistical metrics as well as the architecture of the LR model was comprehensively introduced by Merghadi et al., (2020) [13].

Table 1. Riverbank erosion conditioning factors

No.	Factor	Unit	Data range	Mean value	Std.	Data source	Group
1	River width	meter	177 ÷ 3,196	1,040	652	Field survey	Geomorphological features
2	Bank height	meter	0.36 ÷ 2.95	1.36	0.51	Field survey	
3	Bank slope (tanα)	[-]	0.12 ÷ 6.88	1.1	1.13	Field survey	
4	Vegetation (*NDVI*)	[-]	-0.29 ÷ 0.76	0.48	0.19	RS	
5	Position	[-]	Inner or outer bank			Field survey	
6	Near bank flow velocity	m/s	0.005 ÷ 2.1	0.567	0.302	Hydraulic model	Hydrological features
7	Sediment concentration	ppm	32 ÷ 9,978	665	1,166	Field survey	
8	Soil moisture	%	24.03 ÷ 48.61	31.55	4.31	SA	Soil physical properties
9	Cohesive coefficient	kPa	0.25 ÷ 52.36	13.97	9.61	SA	
10	Grain diameter (*D50*)	mm	0.045 ÷ 0.66	0.25	0.14	SA	
11	Density	g/cm^3	2.33 ÷ 2.78	2.60	0.085	SA	
12	Porosity	%	27.67 ÷ 46.84	34.83	3.54	SA	

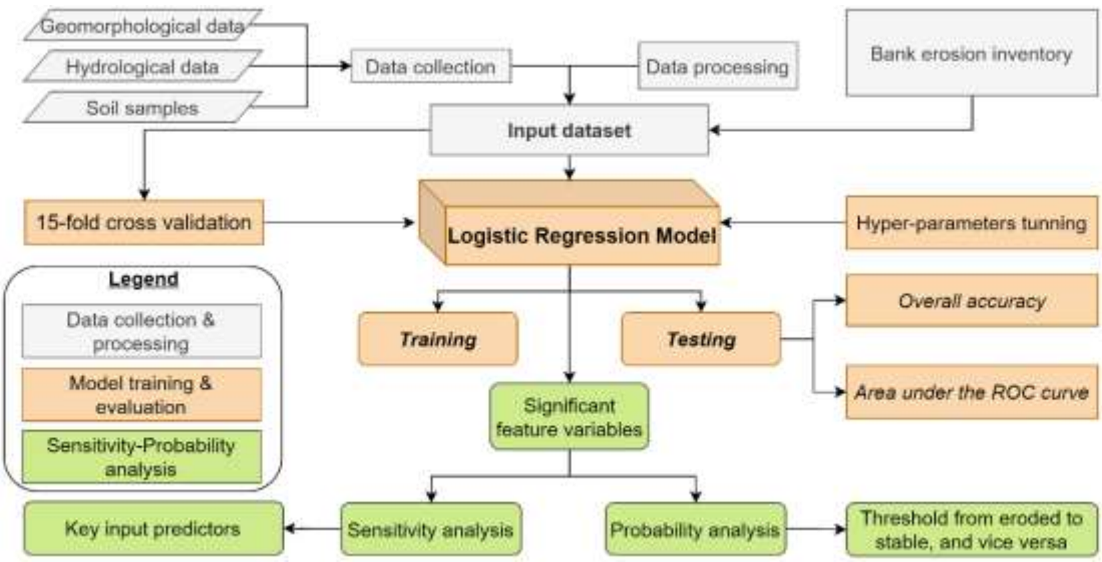

Figure 2. Comprehensive workflow for riverbank erosion susceptibility assessment, including three main phrases: Data collection and preprocessing, model training and evaluation, and sensitivity-probability analysis.

2.3. Sensitivity-probability Analysis

To better understand the relationship between riverbank erosion susceptibility and its controlling factors, we conducted sensitivity and probability analyses. These processes aim to determine the appropriate independent variables and their value thresholds where the riverbank changes from eroded to stable, and vice versa.

In the sensitivity analysis, we considered four combinations of feature variables based on their important level (i.e., the weight value w_i). They are C0 (the whole dataset), C1 (the six most important factors), C2 (the three most important factors), and C3 (the two most important factors). Such combinations were then used to re-train and re-evaluate the model, and the best combination with the highest statistical metrics was selected.

For the probability analysis, riverbank positions were initially classified into three groups: stable ($\sigma < 0.5$), low erosion ($\sigma > 0.6$), and high erosion ($\sigma > 0.8$). By testing different values of significant feature variables one by one, we examined the probability of change to determine the abovementioned thresholds.

The model training, evaluation, and sensitivity-probability analyses were implemented in the Python environment using packages in the Scikit-Learn library.

3. Results and Discussion

3.1. Model Performance

Using the tuned hyper-parameters (Table 2), the LR model was trained with the input data. As a result, the values of ACC and AUC were 0.83 and 0.86, respectively. These scores show that its predictive capability is excellent [14], and better than multiple linear and multiple nonlinear approaches [10]. This performance, however, is quite low in comparison to other research. For instance, Lana et al., (2022) reported that the LR model achieved an outstanding performance in gully erosion susceptibility prediction [15]. This limitation might stem from a limited training dataset, or the multicollinearity effect [16]. This will be confirmed by the sensitivity analysis result in the following section of this article.

3.2. Influential Factors for RBES

Theoretically, each feature variable has a different impact (w_i) on the model's capability to predict riverbank erosion susceptibility. The present study found that the bank slope has the highest importance score (2.764), indicating that such geomorphological characteristic strongly influences the predictability of the LR model. Bank slope is closely related to the historical evolution of a river [17]. It was not only the driver but also the result of the bank erosion processes during the development of watercourses.

Table 2. Applied hyper-parameters for the LR model

Parameter	Value
C	78.476
penalty	L1
solver	liblinear

The second most important factor is soil moisture (1.384), while the bank's height holds the third position (0.790). Besides, some more significant factors are grain diameter, cohesive coefficient, and near bank flow velocity, with weight values of 0.637, 0.591, and 0.518, respectively. Such results imply that geomorphological factors, together with hydrological features, are driving the riverbank erosion susceptibility. The impact values of the remaining features are quite low. Therefore, we choose the first six important variables to establish testing combinations for sensitivity analysis, as explained in Sub-section 2.2, including C0 (whole dataset), C1 (bank's slope, moisture, bank height, grain diameter, cohesive coefficient, near bank flow velocity), C2 (bank's slope, moisture, bank's height), and C3 (bank's slope, moisture).

Sensitivity analysis results are shown in Figure 3, illustrating that C2 is the most appropriate feature variable combination for predicting riverbank erosion susceptibility using the LR model. With this combination, the predictive model achieves the highest performance (ACC = 0.83 and AUC = 0.91). The combinations C1 and C3 also have high AUC scores (0.90 and 0.91, respectively), but ACC values are relatively low (0.82 and 0.80, respectively). Notably, C0 is ordered at the last position, demonstrating that multicollinearity effects exist in the original dataset [16]. By subtracting less important feature variables, such effects were removed and the model could predict more precisely. Given that achieving the training features through field surveys is costly and labour-intensive, the C3 combination is recommended for research with limited resources.

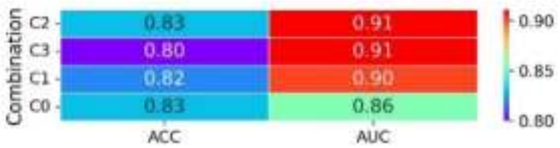

Figure 3. Model's performance in different combinations of feature variables.

3.3. The Thresholds for RBES Changing

Based on the outcomes of the model training phase, three samples' groups were selected for probability analysis: i) Stable bank positions; ii) Low erosion positions; and iii) High erosion positions. These groups were evaluated for the two most influential features, i.e., bank slope and moisture. The results of this analysis are presented in Figure 4. We found that the steeper slopes will increase the erosion probability of all sample sites (Figure 4a). If the slope is greater than 3.4, all samples will become completely eroded (probability = 1). Riverbank erosion probability is more sensitive in the range of the bank slope from 0.1 to 3.4. We also found that the threshold at which the stable position becomes eroded (erosion probability > 0.5) when the slope = 1.3 or the riverbank angle = 52°. On the other hand, the thresholds for the low and high erosion positions become stable when the slopes are 0.52 and 0.32 (i.e., angle = 27.47° and 17.74°), respectively. It demonstrates that these low and high erosion positions are more sensitive to the bank slope's variation. These results reveal the importance of minimizing the bank slope by implementing robust protective measures to mitigate erosion risks in practical applications.

In the experiment involving variations in soil moisture, the trend in riverbank erosion probability was opposite to that observed with bank slope. As soil moisture increased, all test samples exhibited a reduced erosion probability (Figure 4b). Notably, the soil moisture threshold for the stable position to become eroded is 27.25%. Alternately, the low and high erosion positions will become stable when soil moisture is 36.24% and 39.28%, respectively. This indicates that as soil moisture decreases, the stable location is less susceptible to erosion than locations currently eroded.

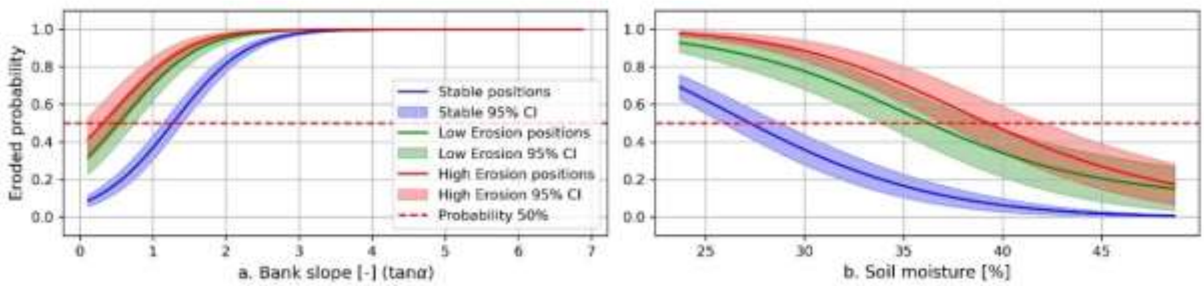

Figure 4. Riverbank erosion probability as a function of bank slope (a) and soil moisture (b).

The lines represent the mean values for each sample group. The probability at which the riverbank exchanges between stable and eroded is 0.5, from which the changing thresholds are determinedThis finding may relate to recent increases in riverbank erosion during the dry season in the VMD. Low water levels reduce soil moisture in riverbanks, accelerating erosion and potentially leading to collapse. Thus, managing water levels in rivers and canals during the dry season should be integrated into disaster risk management strategies, with a focus on preventing riverbank erosion.

4. Conclusion

This study demonstrates the effectiveness of the LR model in predicting riverbank erosion susceptibility along the Ham Luong River in the VMD. The LR model provided highly reliable prediction using twelve controlling variables. The bank slope, soil moisture, and bank height are the most significant features in riverbank erosion susceptibility prediction. Nevertheless, the performance of the model is not optimal compared to other studies, which may be attributed to limitations in the sample size utilized. Besides, several other important factors have not been considered in this study, such as anthropogenic influences and flow direction. Sensitivity analysis revealed that strategic combinations of key variables improved model performance, highlighting multicollinearity in the dataset. Practically, the developed model can be applied to predict RBES for a wider area (e.g., the whole VMD) using only three abovementioned significant features.

Additionally, probability analysis indicates that steeper bank slopes increase erosion probability, while higher soil moisture levels decrease it. These results highlight the importance of reducing bank slopes and maintaining soil moisture to manage erosion risks effectively. Given the recent rise in riverbank erosion during the dry season (due to lower soil moisture), the study recommends proactive water level management in disaster risk frameworks. This research offers crucial insights for assessing erosion susceptibility and provides practical guidance for preserving riverbank integrity in at-risk areas.

Acknowledgements

This research is funded by the Ministry of Education and Training (grant number B2023-VGU-01) and is part of the first author's Master's thesis. We also acknowledge partial funding from the JASTIP-Net and DPRI Collaborative Research (2023IG-01) projects of Kyoto University. We gratefully acknowledge the NbS-SUCCEED project for providing near bank flow velocity data.

References

[1] B. Abernethy, I. D. Rutherfurd, Where Along a River's Length Will Vegetation Most Effectively Stabilize Stream Banks?, Geomorphology, Vol. 23, No. 1, 1998, pp. 55-75, https://doi.org/10.1016/S0169-555X(97)00089-5.

[2] J. P. Bravard, M. Goichot, S. Gaillot, Geography of Sand and Gravel Mining in the Lower Mekong River, EchoGéo, No. 26, 2013, https://doi.org/10.4000/echogeo.13659.

[3] L. M. Hung, H. Tanaka, N. T. Tu, N. T. Viet, Prediction of Riverbank Erosion in the Lower Mekong River Delta, Vietnam-Japan Estuary Workshop, 2006, pp. 169-178.

[4] D. N. Khoi, T. D. Dang, L. T. H. Pham, P. T. Loi, N. T. D. Thuy, N. K. Phung, N. T. Bay, Morphological Change Assessment from Intertidal to River-Dominated Zones Using Multiple-Satellite Imagery: A Case Study of the Vietnamese Mekong Delta, Regional Studies in Marine Science, Vol. 34, 2020, https://doi.org/10.1016/j.rsma.2020.101087.

[5] D. V. Binh, B. Wietlisbach, S. Kantoush, H. H. Loc, E. Park, G. D. Cesare, D. H. Cuong, N. X. Tung, and T. Sumi, A Novel Method for River Bank Detection from Landsat Satellite Data: A Case Study in the Vietnamese Mekong Delta, Remote Sensing, Vol. 12, No. 20, 2020, https://doi.org/10.3390/rs12203298.

[6] Y. Garosi, M. Sheklabadi, C. Conoscenti, H. R. Pourghasemi, K. V. Oost, Assessing the Performance of GIS-Based Machine Learning Models with Different Accuracy Measures for Determining Susceptibility to Gully Erosion, Science of The Total Environment, Vol. 664, 2019, pp. 1117-1132, https://doi.org/10.1016/j.scitotenv.2019.02.093.

[7] H. Sahour, V. Gholami, M. Vazifedan, S. Saeedi, Machine Learning Applications for Water-Induced Soil Erosion Modeling and Mapping, Soil and Tillage Research, Vol. 211, 2021, https://doi.org/10.1016/j.still.2021.105032.

[8] W. Chen, Y. Li, W. Xue, H. Shahabi, S. Li, H. Hong, X. Wang, H. Bian, S. Zhang, B. Pradhan, B. B. Ahmad, Modeling Flood Susceptibility using Data-Driven Approaches of Naive Bayes Tree, Alternating Decision Tree, and Random Forest Methods, Science of The Total Environment, Vol. 701, 2020, pp. 134979, https://doi.org/10.1016/j.scitotenv.2019.134979.

[9] A. Saadon, J. Abdullah, N. S. Muhammad, J. Ariffin, Development of Riverbank Erosion Rate Predictor for Natural Channels using NARX-QR Factorization Model: a Case Study of Sg. Bernam, Selangor, Malaysia, Neural Computing and Applications, Vol. 32, No. 18, 2020, pp. 14839-14849, https://doi.org/10.1007/s00521-020-04835-5.

[10] A. Saadon, J. Abdullah, N. S. Muhammad, J. Ariffin, P. Y. Julien, Predictive models for the estimation of riverbank erosion rates, Catena, Vol. 196, 2021, https://doi.org/10.1016/j.catena.2020.104917.

[11] L. V. Quyen, D. V. Binh, Assessing Riverbank Erosion in the Ham Luong River by Integrating Remote Sensing with Machine Learning and Digital Shoreline Analysis System, Journal of Hydro-meteorology, Vol. 8, No. 764, 2024, pp. 38-52, https://doi.org/10.36335/vnjhm.2024%28764%29. 38-52, (in Vietnamese).

[12] D. Jurafsky, J. H. Martin, Logistic Regression, In: Speech and Language Processing, 3rd Edition, 2023, pp. 75-93, https://web.stanford.edu/~jurafsky/slp3/ed3book_dec302020.pdf (accessed on: June 1st, 2024).

[13] A. Merghadi, A. P. Yunus, J. Dou, J. Whiteley, B. ThaiPham, D. T. Bui, R. Avtar, B. Abderrahmane, Machine Learning Methods for Landslide Susceptibility Studies: A Comparative Overview of Algorithm Performance, Earth-Science Reviews, Vol. 207, 2020, pp. 103225, https://doi.org/10.1016/j.earscirev.2020.103225.

[14] D. W. Hosmer, J. S. Lemeshow, R. X. Sturdivant, Applied Logistic Regression, John Wiley & Sons, 2013.

[15] J. C. Lana, P. D. T. A. Castro, C. E. Lana, Assessing Gully Erosion Susceptibility and Its Conditioning Factors in Southeastern Brazil using Machine Learning Algorithms and Bivariate Statistical Methods: A Regional Approach, Geomorphology, Vol. 402, 2022, https://doi.org/10.1016/j.geomorph.2022.108159.

[16] A. Alin, Multicollinearity, Wiley Interdisciplinary Reviews Computational Stastistics, Vol. 2, 2010, pp. 370-374, https://doi.org/10.1002/wics.84.

[17] A. J. Henshaw, C. R. Thorne, and N. J. Clifford, Identifying Causes and Controls of River Bank Erosion in a British Upland Catchment, Catena, Vol. 100, 2013, pp. 107-119, https://doi.org/10.1016/j.catena.2012.07.015.