remote sensing

MDPI

Article

Gully Erosion Susceptibility Prediction Using High-Resolution Data: Evaluation, Comparison, and Improvement of Multiple Machine Learning Models

Heyang Li [1,2,†], Jizhong Jin [1,2,†], Feiyang Dong [1,2], Jingyao Zhang [1,2], Lei Li [1,2] and Yucheng Zhang [1,*]

[1] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China; liheyang23p@ict.ac.cn (H.L.); jinjizhong22s@ict.ac.cn (J.J.); dongfeiyang22s@ict.ac.cn (F.D.); zhangjingyao23p@ict.ac.cn (J.Z.); lilei22b1@ict.ac.cn (L.L.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: zhangyucheng@ict.ac.cn
† These authors contributed equally to this work.

Abstract: Gully erosion is one of the significant environmental issues facing the black soil regions in Northeast China, and its formation is closely related to various environmental factors. This study employs multiple machine learning models to assess gully erosion susceptibility in this region. The primary objective is to evaluate and optimize the top-performing model under high-resolution UAV data conditions, utilize the optimized best model to identify key factors influencing the occurrence of gully erosion from 11 variables, and generate a local gully erosion susceptibility map. Using 0.2 m resolution DEM and DOM data obtained from high-resolution UAVs, 2,554,138 pixels from 64 gully and 64 non-gully plots were analyzed and compiled into the research dataset. Twelve models, including Logistic Regression, K-Nearest Neighbors, Classification and Regression Trees, Random Forest, Boosted Regression Trees, Adaptive Boosting, Extreme Gradient Boosting, an Artificial Neural Network, a Convolutional Neural Network, as well as optimized XGBOOST, a CNN with a Multi-Head Attention mechanism, and an ANN with a Multi-Head Attention Mechanism, were utilized to evaluate gully erosion susceptibility in the Dahewan area. The performance of each model was evaluated using ROC curves, and the model fitting performance and robustness were validated through Accuracy and Cohen's Kappa statistics, as well as RMSE and MAE indicators. The optimized XGBOOST model achieved the highest performance with an AUC-ROC of 0.9909, and through SHAP analysis, we identified roughness as the most significant factor affecting local gully erosion, with a relative importance of 0.277195. Additionally, the Gully Erosion Susceptibility Map generated by the optimized XGBOOST model illustrated the distribution of local gully erosion risks.

Keywords: gully erosion; machine learning; unmanned aerial vehicle; susceptibility mapping; geo-environmental factors

Citation: Li, H.; Jin, J.; Dong, F.; Zhang, J.; Li, L.; Zhang, Y. Gully Erosion Susceptibility Prediction Using High-Resolution Data: Evaluation, Comparison, and Improvement of Multiple Machine Learning Models. *Remote Sens.* **2024**, *16*, 4742. https://doi.org/10.3390/rs16244742

Academic Editor: Sandro Moretti

Received: 5 November 2024
Revised: 14 December 2024
Accepted: 17 December 2024
Published: 19 December 2024

1. Introduction

Gully erosion is a severe soil degradation process, primarily characterized by the formation of linear or channel-like erosional landforms due to concentrated surface runoff scouring [1]. This globally widespread phenomenon represents one of the most destructive forms of soil degradation, threatening agricultural sustainability and environmental stability worldwide. This erosion type not only results in soil loss but also destroys farmland structure, reduces land productivity, and poses a significant threat to the sustainable development of agriculture [2]. It also severely impacts water resources and ecosystem health, leading to substantial environmental and economic losses in affected regions. The initiation and progression of gully erosion are influenced by a multitude of factors, whose interactions determine the spatial distribution characteristics and temporal evolution patterns of gully erosion [3].

Research methods for gully erosion have evolved from qualitative descriptions to quantitative analyses, and from single-factor studies to comprehensive multi-factor assessments [4]. Over the past decade, with the rapid advancements in GIS and RS technologies, quantitative assessment methods combined with machine learning algorithms have gradually emerged as research hotspots [5]. In the field of gully distribution modeling, the primary approaches include topographic threshold models, traditional statistical methods, and machine learning algorithms. Topographic threshold models, such as the physically based model proposed by Montgomery and Dietrich, are simple and intuitive but often neglect the influence of other key factors [6]. Traditional statistical methods, such as discriminant analysis used by Arabamer et al., can account for multiple influencing factors but exhibit limitations in capturing complex nonlinear relationships [7]. These conventional methods also struggle with processing large-scale spatial data and handling the intricate interactions among multiple environmental variables. In contrast, machine learning algorithms demonstrate significant advantages in handling high-dimensional data and complex nonlinear relationships [8]. Their superior capability in processing massive datasets, capturing subtle patterns, and automatically extracting features makes them particularly suitable for gully erosion susceptibility assessment.

Since 2015, machine learning algorithms have been widely applied and thoroughly investigated in gully erosion research, with many studies employing multiple methods within a single study. For instance, Arabameri et al. [9] and Conoscenti et al. [10] utilized LR, Avand et al. [11] employed KNN and RF methods, and Pourghasemi et al. [12] applied the CART method in comparison with other machine learning algorithms. As a prevalent machine learning method in this field, Amir [13], Arabameri [14], Garosi [15], and Rahmati [16] all employed the RF method. Chen [17] utilized both BRT and XGBOOST methods. Hembram [18] utilized both RF and BRT. In 2021, Arabameri [19] applied the XGBOOST model for the spatial mapping of gully erosion susceptibility, and Eloudi et al. [20] used ADABOOST, XGBOOST, and RF in their study. Arabameri [21] utilized the ANN model in their study, and similarly, Garosi [22] also employed the ANN model. Sara [23] applied the CNN model in their study. In addition, different from the above-mentioned studies focusing on gully erosion susceptibility assessment, various methodological approaches have evolved for gully detection and monitoring applications. Conventional methodologies employ GPS technology to achieve precise measurements and the continuous monitoring of gully morphological characteristics [24]. Subsequently, advanced semi-automated approaches have emerged, integrating Geographic Object-Based Image Analysis with machine learning algorithms for systematic gully network detection [25]. Furthermore, researchers have implemented deep learning-based methodologies for automated gully segmentation utilizing remote sensing imagery [26,27].

Based on a systematic review of the existing literature and the availability of high-resolution RS data, this study selects 11 key controlling factors and employs nine native machine learning algorithms and three optimized machine learning algorithms to systematically assess gully erosion susceptibility in the Dahewan region. In this study, we established a comprehensive gully erosion inventory through high-resolution (0.2 m) UAV surveys and DOM and DEM data. After field validation, 64 gully and 64 non-gully sites were carefully selected as sample plots, providing a robust foundation for subsequent machine learning model training and validation. For traditional machine learning models (excluding ANN and CNN), a 5-fold cross-validation strategy was implemented to ensure model reliability and stability. The native machine learning models include LR, KNN, CART, RF, BRT, ADABOOST, XGBOOST, ANNs, and CNNs. The optimized machine learning algorithms consist of optimized XGBOOST (using Optuna, Gradient Early Stopping, and the Synthetic Minority Over-Sampling Technique) and CNN and ANN models enhanced with Multi-Head Attention mechanisms. By comparing the performance of these models on high-resolution data using evaluation metrics such as AUC-ROC, Accuracy, Cohen's Kappa statistic, RMSE, and MAE, this study aims to provide a more precise and reliable methodological foundation for gully erosion susceptibility assessment and offer

scientific evidence for the formulation of regional soil conservation strategies and ecological environmental protection.

The primary objectives of this study include the following: (1) improving the performance of the top-performing machine learning model through optimization techniques and enhancing the performance of CNN and ANN models by incorporating Multi-Head Attention mechanisms; (2) comprehensively comparing the performance of 12 models on a high-resolution dataset, with traditional machine learning models (excluding ANNs and CNNs) evaluated through 5-fold cross-validation, using metrics such as AUC-ROC and Accuracy; and (3) generating the GESM for the study area based on the optimal model and assessing the relative importance of gully erosion control factors. The combination of model optimization techniques and high-resolution data analysis is expected to enhance the Accuracy and reliability of gully erosion susceptibility assessment. Through these efforts, this study will thoroughly evaluate the application of machine learning models in high-resolution gully erosion susceptibility assessments, provide more accurate and reliable prediction tools for soil conservation planning, and contribute to the development of targeted erosion control strategies. The improved prediction accuracy and factor importance analysis will directly support decision-makers in implementing cost-effective soil and water conservation measures, particularly in regions highly susceptible to gully erosion.

2. Materials and Methods

2.1. Study Area

The study area is located in Dahewan Town, Zalantun City, Inner Mongolia Autonomous Region, China, as shown in Figure 1. The topography of the region is predominantly characterized by gentle slopes and rolling hills, and the climate is classified as a temperate continental semi-humid monsoon climate. The region experiences an annual mean temperature of 2.4 °C, spring mean temperatures ranging from 5.8 °C to 9.1 °C, and an annual effective accumulated temperature of 2442.5 °C. Spring is generally dry, while summer experiences abundant rainfall, with the annual precipitation ranging between 400 and 500 mm, primarily concentrated between June and August. As a key area for gully erosion prevention, the region has a gully coverage area of 0.197 km^2, which affects 34.75 km^2 of the arable land [28]. Figure 1 shows the geographical context and gully characteristics of the study area: (a) the geographical location of the study area; (b) the spatial distribution of gully and non-gully areas derived from UAV imagery; (c) field photographs showing typical gully features; and (d) a UAV-captured gully image with the highlighted gully area.

2.2. Experimental Steps

Based on a systematic review of the existing literature and the availability of high-resolution remote sensing data, this study selected 11 key influencing factors and employed nine native machine learning algorithms along with three optimized algorithms to systematically assess gully erosion susceptibility in the Dahewan region. The research methodology comprised five main steps. In Step 1, DOM and DEM were used to generate a gully inventory map and obtain a gully dataset comprising 11 influencing factors. The second step involved standardizing the dataset constructed with the 11 influencing factors and performing multicollinearity analysis using VIF to remove factors with high multicollinearity. Following this, the performance of nine machine learning methods was evaluated and compared using AUC-ROC through 5-fold cross-validation to ensure robust model assessment and to select the best-performing model. The fourth step focused on enhancing the performance of the best-performing model using appropriate optimization methods and improving the CNN and ANN models by incorporating Multi-Head Attention mechanisms. Finally, the GESM was generated using the best-performing model, and the relative importance of influencing factors on gully erosion was assessed using the SHAP importance analysis method. Through these steps, the study provided a comprehensive evaluation of gully erosion susceptibility, ensuring methodological rigor and the reliability of the results. The methodological sequence for this study is depicted in Figure 2.

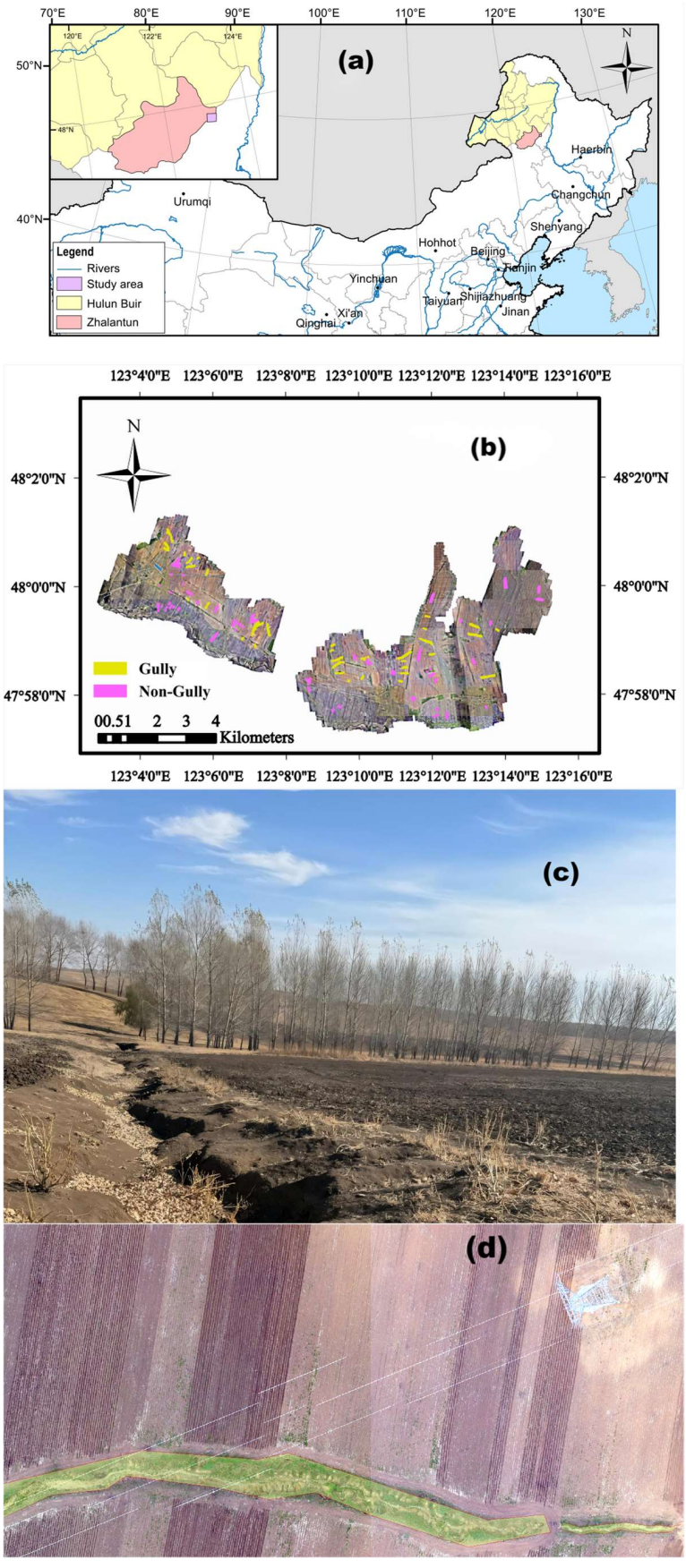

Figure 1. (**a**) The schematic location of the study area; (**b**) a display of the study area; (**c**) a field photograph of the gully; and (**d**) a UAV-captured gully image with the highlighted gully area.

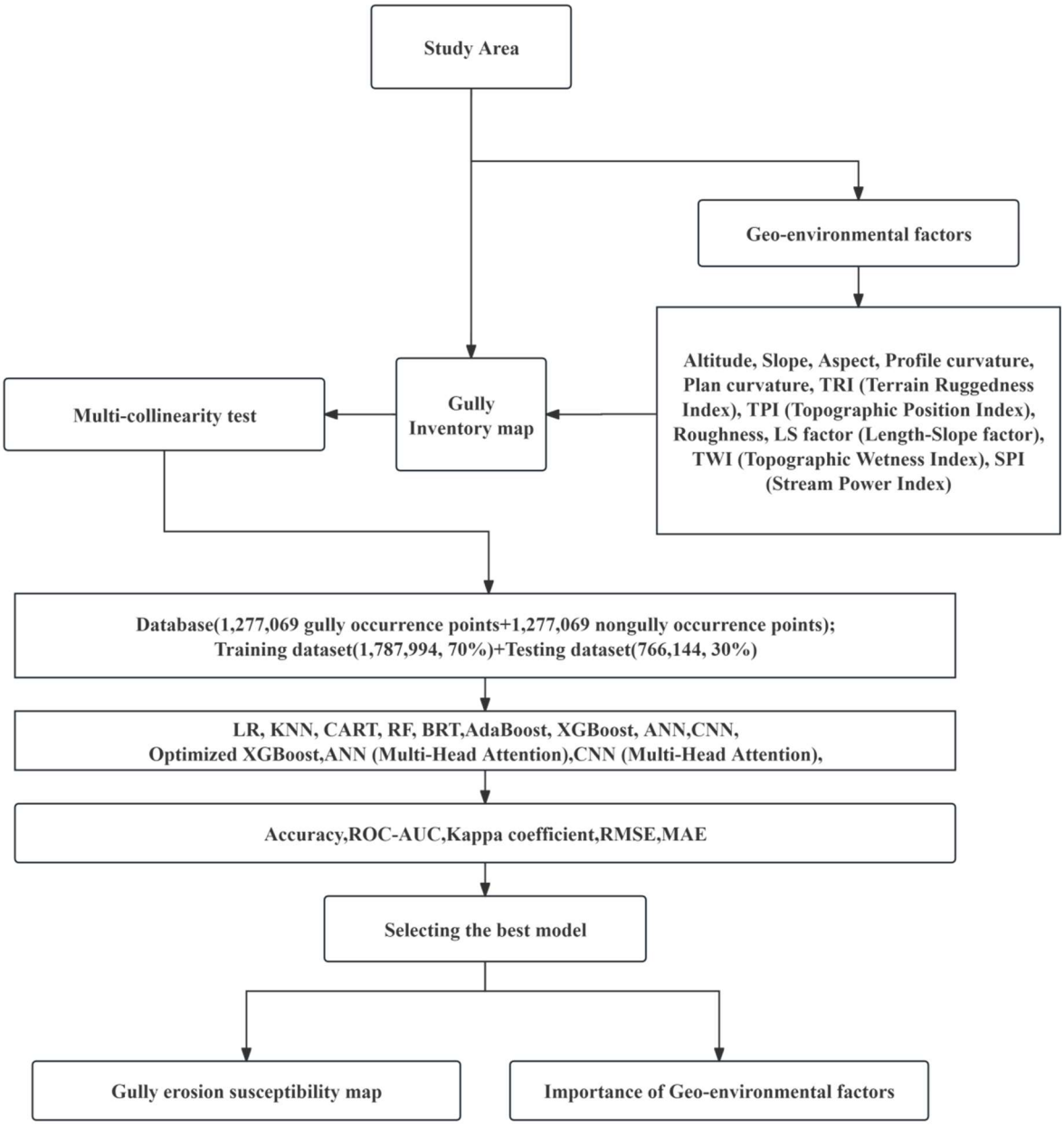

Figure 2. Flowchart of the methodology used in this study.

2.3. Data and Data Preprocessing

The fundamental data comprised DOM and DEM with a resolution of 0.2 m, obtained using a DJI Phantom 4 UAV. The identification and validation of the gully samples followed a two-step process. First, clearly visible gullies were delineated directly from the high-resolution UAV imagery, where their morphological characteristics were distinct. Second, for less obvious gully features, field investigations were conducted to verify their presence and characteristics. During the field validation process, gullies that could be crossed by agricultural machinery or those that typically disappear after rainy seasons were classified as shallow gullies and were excluded from the study. The dataset comprised 64 gully landforms and 64 non-gully landforms as experimental samples. To create a dataset suitable for modeling, we processed the DEM data using ArcMap10.5, QGIS3.36.1, and the rasterio1.4.3 and GDAL3.6.2 modules in Python. The Profile curvature, Plan curvature, and the Aspect and slope data were obtained using ArcMap; the TRI, TPI, and roughness data were acquired from QGIS. The LS-factor, TWI, and SPI were first derived from flow data obtained using ArcMap and subsequently calculated using Python's rasterio package.

After obtaining 11 raster layers, the GDAL module in Python was used to convert each of the 11 different TIF files into their corresponding CSV files. After eliminating invalid values and aligning the coordinates, the dataset was split into a training set (comprising 1,787,994 samples) and a testing set (comprising 766,144 samples) in a 70% to 30% ratio.

2.3.1. Geo-Environmental Factors (GEFs)

Currently, there is no clearly defined global standard for determining the GEFs influencing gully erosion across different study areas [29,30]. Therefore, based on previous studies [31–33] and field investigations, eleven key factors were selected. These factors included Altitude, slope, Aspect, Profile curvature, Plan curvature, TRI, TPI, roughness, LS Factor, TWI, and SPI (Figure 3). The selection of these factors was determined by their proven significance in gully erosion processes, the specific environmental characteristics of the study area, and the availability of high-resolution data. Altitude represents the vertical distance from the mean sea level, affecting local climate, vegetation, and surface processes. Slope (in degrees) indicates surface inclination, controlling hydrological processes and soil erosion [34]. Aspect (0–360°) refers to slope orientation, affecting local sunlight, moisture conditions, and vegetation growth [34]. Profile curvature indicates slope changes in the steepest direction, with positive values showing convex surfaces and negative values showing concave surfaces [35]. Plan curvature describes contour line curvature, determining water flow convergence and divergence patterns [35]. TRI measures terrain heterogeneity and surface complexity [36]. TPI identifies landform features by comparing point elevation with the surrounding average elevation [37]. Roughness quantifies surface irregularity, influencing water and wind dynamics [38]. LS Factor combines slope length and steepness to assess erosion potential [39]. TWI indicates topographic moisture accumulation patterns [36,40]. SPI measures potential water flow erosive power [41].

2.3.2. Multicollinearity Analysis

In multivariate statistical analysis, multicollinearity is a problem that requires special attention. Multicollinearity refers to the phenomenon of high intercorrelations among predictor variables, which can lead to instability in model estimates, reduce the reliability of statistical inferences, and affect the predictive performance of the model [42]. To identify and quantify multicollinearity, researchers typically employ VIF, a widely used statistical metric [43]. VIF provides an intuitive method to measure the degree of multicollinearity for each predictor variable. It achieves this by quantifying the extent to which the variance of regression coefficients increases due to inter-variable correlations [44]. The calculation of VIF is based on the coefficient of determination (R^2) obtained by regressing each predictor variable against all other predictor variables. Specifically, for the predictor variable X_i, its VIF is calculated using Equation (1):

$$\mathrm{VIF} = \frac{1}{(1 - R_i^2)} \quad (1)$$

where R_i^2 is the coefficient of determination obtained by regressing X_i as the dependent variable against all other predictor variables as independent variables. Generally, a VIF value of 1 indicates no multicollinearity, while a VIF value exceeding 10 is considered indicative of severe multicollinearity.

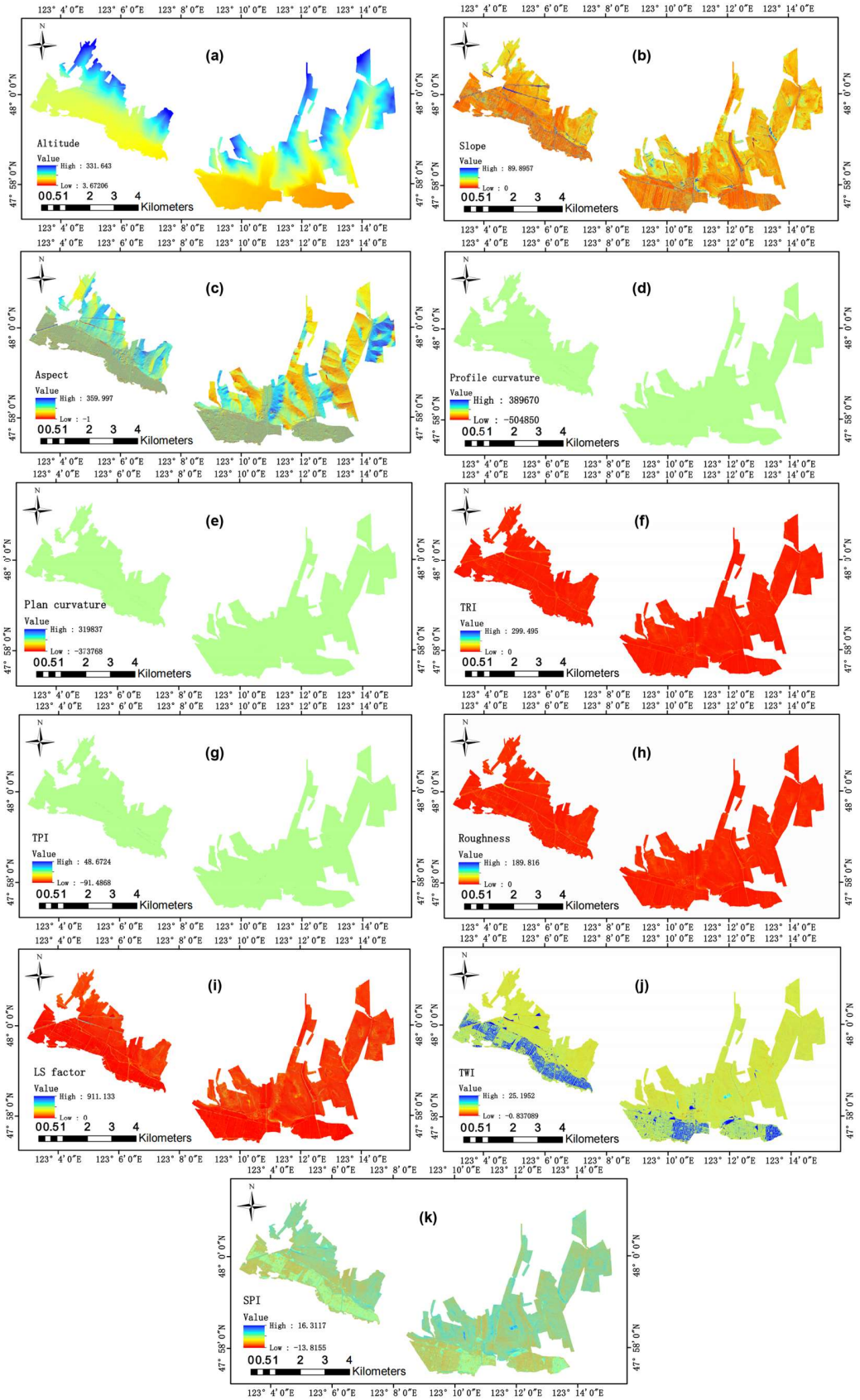

Figure 3. Maps of geo-environmental factors (GEFs): (**a**) Altitude, (**b**) slope, (**c**) Aspect, (**d**) Profile curvature, (**e**) Plan curvature, (**f**) Topographic Ruggedness Index, (**g**) Topographic Position Index, (**h**) roughness, (**i**) LS Factor, (**j**) Topographic Wetness Index, and (**k**) Stream Power Index.

2.4. Methods

2.4.1. Logistic Regression (LR)

Logistic Regression, a widely used statistical method, is employed to model the probability of binary outcomes based on one or more predictor variables [45]. Despite the inclusion of the term "regression" in its name, LR is actually a classification algorithm primarily used to predict categorical dependent variables [46]. The logistic function (also known as the Sigmoid function) is the core component of LR, translating the linear combination of input features into probability values ranging from 0 to 1. LR utilizes maximum likelihood estimation to estimate the coefficients of input features, aiming to find the best fit that most effectively explains the observed data [47]. In this study, the LR model parameters were set as follows: penalty = 'l2', C = 0.1, solver = 'LBFGS', and maximum iterations = 1000.

2.4.2. K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a simple yet effective non-parametric supervised learning algorithm, widely applied in classification and regression tasks [48]. The core idea of KNN is based on the assumption that similar instances possess similar attributes by calculating the distance between the instance to be classified and all instances in the training set and selecting the nearest K neighbors for prediction. In classification tasks, KNN determines the class of the instance to be classified through a majority voting mechanism; whereas in regression tasks, it predicts the outcome by computing the arithmetic mean or weighted average of the K-Nearest Neighbors. Common distance metrics include the Euclidean distance, the Manhattan distance, and the Minkowski distance [49]. In this study, the KNN model parameter was set as follows: number of neighbors = 10.

2.4.3. Classification and Regression Trees (CART)

Classification and Regression Trees is a widely used decision tree algorithm applicable to both classification and regression problems. The algorithm constructs the tree structure through recursive binary splitting, selecting at each split the feature and threshold that maximally reduce impurity (for classification problems) or minimize the squared error (for regression problems). CART uses the Gini impurity or information gain as the splitting criterion for classification problems, and employs the mean squared error for regression problems [50]. A notable feature of CART is its ability to handle both continuous and categorical variables while automatically performing feature selection. The tree-building process continues until the stopping criteria are met, such as the minimum number of samples per node or the maximum tree depth. To prevent overfitting, CART typically employs pruning techniques, evaluating subtrees of varying complexity using a validation set to select the model with the best generalization performance [51]. In this study, the CART model parameters were set as follows: criterion = 'Gini impurity', maximum depth = 10, minimum samples for split = 5, and minimum samples per leaf = 2.

2.4.4. Random Forest (RF)

Random Forest is a robust ensemble learning method proposed by Breiman in 2001 [52]. It enhances the overall performance and generalization capability of a model by constructing multiple decision trees and combining their predictions. The core idea of RF is to reduce the risk of overfitting by introducing randomness while maintaining high predictive accuracy. The main steps of the algorithm include the following: creating multiple subsets from the original training set using bootstrap sampling; for each subset, randomly selecting a subset of features at each node split instead of considering all features; and allowing the decision trees to grow fully without pruning [52]. During the prediction phase, RF derives the final result by aggregating the predictions of all trees through majority voting for classification problems or averaging for regression problems. An important feature of RF is its built-in validation mechanism, known as the Out-of-Bag error, which provides an

unbiased estimate of the model's performance [53]. In this study, the RF model parameters were set as follows: number of estimators = 500 and maximum depth = 50.

2.4.5. Boosted Regression Trees (BRT)

The Boosted Regression Trees method is a powerful ensemble learning method that combines decision trees with gradient-boosting algorithms [54]. BRT iteratively trains a series of simple decision trees, where each new tree attempts to correct the residuals of the preceding trees, thereby progressively enhancing the model's performance. The core principle of this approach is to combine multiple "weak learners" into a robust predictive model [55]. The main steps of BRT include the following: initializing the model's predictions to a constant value; calculating the current model's residuals; training a new decision tree to fit these residuals; multiplying the new tree's predictions by a small learning rate and adding them to the current model; and repeating this process until a stopping criterion is met. The advantages of BRT include its ability to automatically handle interactions between features, accommodate various types of predictor variables, exhibit robustness to outliers, and produce highly accurate predictions [56]. In this study, the BRT model parameters were set as follows: number of estimators = 500, learning rate = 0.05, and maximum depth = 5.

2.4.6. Adaptive Boosting (ADABOOST)

Adaptive Boosting is a significant ensemble learning algorithm developed by Freund and Schapire in 1995. The core concept of ADABOOST is to construct a strong classifier by adjusting sample weights and combining multiple weak learners. The main steps of the algorithm include the following: initializing each training sample with equal weights; iteratively training weak learners, adjusting the sample weights based on the classification results of the previous iteration, increasing the weights for misclassified samples and decreasing the weights for correctly classified ones; calculating the weight of each weak learner, with better classification performance leading to higher weights; and finally, combining all weak learners with their respective weights to form the final strong classifier. A notable characteristic of ADABOOST is its sensitivity to outliers, which can lead to overfitting in specific scenarios. However, in practice, ADABOOST generally demonstrates good generalization capabilities, a phenomenon known as the "boundary effect" [57,58]. In this study, the ADABOOST model parameters were set as follows: number of estimators = 500, learning rate = 0.1, and base estimator maximum depth = 3.

2.4.7. Extreme Gradient Boosting (XGBOOST)

Extreme Gradient Boosting is an efficient, flexible, and scalable implementation of gradient boosting trees, proposed by Chen et al. in 2016. As an improved version of GBDT, XGBOOST introduces innovations at both the algorithmic and system optimization levels. Its core principle is to integrate multiple weak learners (decision trees) into a strong learner by adding new trees in each iteration to fit the residuals of the previous model. The main features of XGBOOST include the following: utilizing second-order Taylor expansion to approximate the objective function, thereby enhancing convergence speed; introducing regularization terms to control model complexity, effectively preventing overfitting; supporting column sampling to further reduce overfitting risk; and implementing efficient distributed computing to handle large-scale data [59]. In this study, the XGBOOST model parameters were set as follows: number of estimators = 500, maximum depth = 50, learning rate = 0.1, and minimum child weight = 1.

2.4.8. Artificial Neural Network (ANN)

An Artificial Neural Network is a machine learning model inspired by biological neural networks, capable of learning and simulating complex nonlinear relationships. ANN consists of interconnected artificial neurons organized into multiple layers, including an input layer, one or more hidden layers, and an output layer [60]. Each neuron receives

weighted inputs from the preceding layer, processes these inputs through an activation function, and then transmits the results to the next layer. The learning process of an ANN is primarily realized through the backpropagation algorithm, which calculates the error between predicted and actual values and propagates these errors backward through the network to adjust the network weights [61]. In this study, the ANN model comprised an input layer and four hidden layers (with 128, 256, 128, and 64 neurons, respectively), each followed by ReLU activation and dropout (dropout rate = 0.5). The output layer consisted of a single neuron utilizing Sigmoid activation. The model employed binary cross-entropy loss and the Adam optimizer (learning rate α = 0.001), with a training batch size = 32, trained for 50 epochs.

2.4.9. Convolutional Neural Network (CNN)

A Convolutional Neural Network is a type of deep learning model specifically designed to handle grid-like topological data structures, achieving remarkable success in the field of computer vision [62]. The core concept of a CNN is to automatically learn and extract local features from data through the use of convolutional layers, a structure that effectively captures spatial hierarchical relationships within images. A typical CNN architecture comprises convolutional layers, pooling layers, and fully connected layers. Convolutional layers employ a series of learnable filters to detect local patterns in the input data; pooling layers reduce the spatial dimensions of the data through downsampling, enhancing the model's translation invariance; and fully connected layers are utilized for the final classification or regression tasks [63]. In this study, the CNN model comprised three convolutional layers (with 16, 32, and 64 filters, respectively), each followed by ReLU activation and maximum pooling. Subsequently, there were three fully connected layers (with 128, 64, and 1 neurons, respectively), culminating in a Sigmoid activation function. The model utilized binary cross-entropy loss and the Adam optimizer (learning rate α = 0.001), with training batch size = 32 and trained for 50 epochs.

2.4.10. Multi-Head Attention Mechanism

The Multi-Head Attention mechanism was introduced by Vaswani et al. in 2017 [64]. This mechanism enhances a model's ability to capture information from different subspaces by parallelly computing multiple attention "heads." Each attention head independently learns a set of linear projections for queries, keys, and values, and then performs scaled dot-product attention calculations. Specifically, the computation process of Multi-Head Attention includes the following: linearly projecting the input sequence into multiple subspaces; independently computing attention weights in each subspace; and concatenating the outputs of each head and integrating them through a linear layer. This design allows the model to simultaneously focus on different positions within the sequence and extract features from multiple perspectives, thereby enhancing the model's expressive capacity and performance. Multi-Head Attention has demonstrated outstanding performance not only in Natural Language Processing tasks but also in various other fields, such as computer vision and recommendation systems [65].

2.4.11. Optuna

Optuna is an automatic Hyperparameter Optimization framework developed by Preferred Networks and released in 2019 [66]. Its design objective is to provide a flexible, efficient, and user-friendly tool for optimizing the hyperparameters of machine learning models. The core principle of Optuna is to treat hyperparameter optimization as a Black-box Optimization Problem, finding the optimal solution by sampling and evaluating various hyperparameter configurations.

2.4.12. Gradient Early Stopping

Gradient Early Stopping is a regularization technique used to prevent overfitting in deep learning models. This method determines when to halt the training process by

monitoring the model's performance on the validation set. The core principle is to cease training when the model's performance on the validation set begins to decline or no longer shows significant improvement, as continued training may result in overfitting the training data. Implementing Gradient Early Stopping typically involves the following steps: evaluating the model's validation metrics on the validation set after each training epoch; updating the best model if performance improves; and stopping training and reverting to the best model if performance does not improve for several consecutive epochs [67].

2.4.13. Synthetic Minority Over-Sampling Technique (SMOTE)

The Synthetic Minority Over-Sampling Technique is an over-sampling technique used to address imbalanced classification problems, introduced by Chawla et al. in 2002 [68]. This method increases the number of minority class samples by generating synthetic samples in the feature space, thereby balancing the dataset. The core principle of SMOTE is to perform interpolation on minority class samples instead of simply duplicating existing samples. Its main steps include the following: identifying the K-Nearest Neighbors for each minority class sample in the feature space; randomly selecting one of these neighbors; and generating new synthetic instances by randomly choosing a point along the line connecting the original sample and the selected neighbor [69]. The advantage of SMOTE lies in its ability to generate new and diverse minority class samples, thereby avoiding model overfitting caused by simple duplication.

2.4.14. SHapley Additive exPlanations (SHAP)

SHAP, a game theory-based approach, is employed to quantify feature importance in machine learning model predictions. The methodology calculates the contribution of each feature by evaluating its marginal effect across all possible feature combinations, thereby providing both global importance rankings and local feature effects. Unlike traditional feature importance methods, SHAP allocates the prediction value fairly among features while accounting for feature interactions and non-linearity. The absolute SHAP values are averaged across all samples to determine feature importance rankings, where higher values indicate a stronger influence on model predictions. This approach ensures consistent and locally accurate interpretations while maintaining a strong theoretical foundation based on cooperative game theory [70].

2.5. Model Evaluation and Validation

Accuracy is a fundamental metric for evaluating the overall performance of classification models, defined as the proportion of correctly predicted samples to the total number of samples. Its calculation formula is Equation (2):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Here, True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) represent the four possible outcomes of classification results. TP and TN denote the number of pixels correctly classified as gully occurrence and non-gully occurrence pixels, respectively. FP indicates the number of non-gully occurrence pixels incorrectly classified as gully occurrence, while FN represents the number of gully occurrence pixels incorrectly classified as non-gully occurrence. The kappa coefficient (κ) is used to evaluate the degree of agreement between the performance of the classification model and that of random classification [71]. Its calculation formula is Equation (3):

$$\kappa = \frac{P_{obs} - P_{exp}}{1 - P_{exp}} \tag{3}$$

$$P_{obs} = \frac{TP + TN}{n} \tag{4}$$

$$P_{exp} = \frac{[(TP + FN)(TP + FP) + (FP + TN)(FN + TN)]}{n^2} \quad (5)$$

Here, n is the total number of samples. The Kappa value typically ranges between $[-1, 1]$, where 0 indicates agreement comparable to random classification, and 1 indicates perfect agreement.

The Area Under the ROC Curve (AUC) provides a single numerical measure of model performance. The AUC represents the probability that a randomly chosen positive instance is ranked higher than a randomly chosen negative instance. The AUC value ranges between [0, 1], where 0.5 indicates random guessing and 1 indicates perfect classification. The advantage of the AUC is that it is not affected by class imbalance and evaluates the model's ranking ability independently of a specific classification threshold [72,73].

The Root Mean Squared Error (RMSE) is a metric that measures the average deviation between predicted and actual values and is more sensitive to large errors. Its calculation formula is Equation (6):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_1 - x_2)^2}{n}} \quad (6)$$

where x_1 represents the actual value, x_2 represents the predicted value, and n is the number of samples. The unit of the RMSE is the same as that of the dependent variable, facilitating interpretation.

The Mean Absolute Error (MAE) measures the average absolute deviation between the predicted and actual values [74]. Its calculation formula is Equation (7):

$$\text{MAE} = \frac{\sum_{i=1}^{n}|x_1 - x_2|}{n} \quad (7)$$

The MAE assigns equal weight to all errors and is less sensitive to outliers compared to the RMSE.

3. Results and Discussion

3.1. Results of Multicollinearity Analysis

To assess the multicollinearity among the selected geological environmental factors, VIF analysis was conducted (Figure 4). The initial analysis included 11 features, among which TRI had the highest VIF value of 31.349, significantly exceeding the critical threshold of 10.0 and indicating a severe multicollinearity issue. After the first iteration of analysis, the TRI feature with the highest VIF value was removed. In the updated VIF values, TPI emerged as the feature with the highest VIF value of 11.431, still surpassing the threshold. In the second iteration, the TPI feature was further removed. The observed exceedance of VIF thresholds by TRI and TPI can be attributed to their computational dependency on elevation differentials, which suggests inherent correlations with slope-derived factors, collectively contributing to multicollinearity. The final results, as shown in Figure 4, indicate that the remaining nine features have VIF values below the acceptable threshold of 5.0.

3.2. Model Performance

After conducting multicollinearity analysis and dataset construction, this study trained and evaluated nine machine learning models, including LR, KNN, CART, RF, BRT, ADABOOST, XGBOOST, an ANN, and a CNN. Model performance was quantified using multiple evaluation metrics, including Accuracy, the Kappa coefficient (κ), the AUC, the RMSE, and the MAE. The performance metrics of the nine models are presented in Table 1.

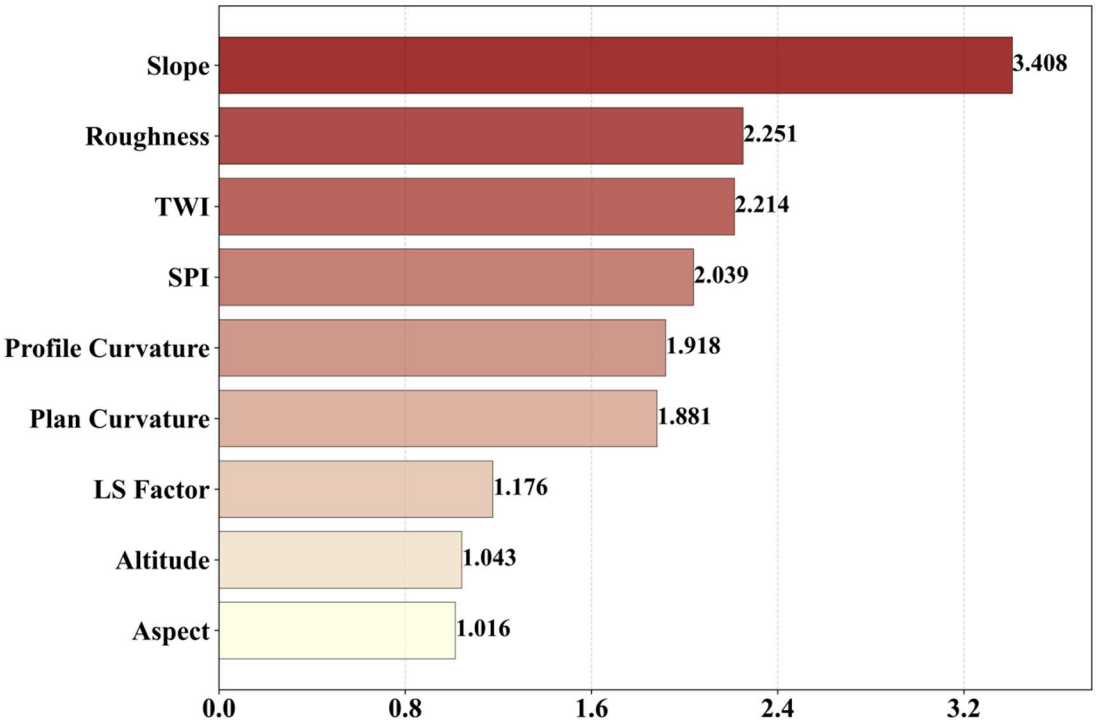

Figure 4. Multicollinearity analysis of the geo-environmental factors.

Table 1. Performance metrics of standard models (* represents optimal values).

Metrics	LR	KNN	CART	RF	BRT	ADABOOST	XGBOOST	ANN	CNN
Accuracy	0.8759 (−0.0761)	0.9369 (−0.0150)	0.9329 (−0.0190)	0.9492 (−0.0027)	0.9460 (−0.0059)	0.9413 (−0.0107)	0.9520 *	0.9039 (−0.0480)	0.9351 (−0.0168)
K	0.7518 (−0.1521)	0.8738 (−0.0301)	0.8659 (−0.0380)	0.8985 (−0.0054)	0.8921 (−0.0118)	0.8826 (−0.0213)	0.9039 *	0.8079 (−0.0960)	0.8702 (−0.0337)
AUC	0.9403 (−0.0500)	0.9794 (−0.0109)	0.9814 (−0.0089)	0.9892 (−0.0011)	0.9879 (−0.0024)	0.9867 (−0.0036)	0.9903 *	0.9683 (−0.0220)	0.9837 (−0.0066)
RMSE	0.3523 (+0.1331)	0.2512 (+0.0320)	0.2589 (+0.0398)	0.2253 (+0.0061)	0.2323 (+0.0131)	0.2423 (+0.0231)	0.2192 *	0.2633 (+0.0441)	0.2193 (+0.0001)
MAE	0.1241 (+0.0761)	0.0631 (+0.0150)	0.0671 (+0.0190)	0.0508 (+0.0027)	0.0540 (+0.0059)	0.0587 (+0.0107)	0.0480 *	0.1468 (+0.0988)	0.1030 (+0.0550)

From Table 1, it can be observed that ensemble learning methods, particularly tree-based models, exhibit significant advantages in overall performance. The XGBOOST model achieved optimal performance across all evaluation metrics, with an Accuracy of 95.20%, a κ of 0.9039, an AUC of 99.03%, an RMSE of 0.2192, and an MAE of 0.0480. These results strongly indicate that XGBOOST possesses exceptional performance and stability when handling the dataset used in this study. The RF model closely followed this, with performance metrics marginally inferior to those of XGBOOST. RF attained an Accuracy of 94.92%, which is only 0.28 percentage points lower than that of XGBOOST, and an AUC of 98.92%, differing by a mere 0.11 percentage points from XGBOOST. This demonstrates that the RF model performs excellently in both classification tasks and ranking capabilities. Other ensemble learning methods, such as BRT and ADABOOST, also exhibited strong performance. BRT achieved an Accuracy of 94.60%, ADABOOST achieved 94.13%, and both demonstrated AUC values exceeding 98.6%. These findings further validate the effectiveness of ensemble learning methods in handling complex datasets.

Among the neural network models, the CNN outperformed the traditional ANN. The CNN demonstrated a significantly higher Accuracy (93.51%) and AUC (98.37%) compared

to the ANN. Notably, the CNN exhibited an outstanding performance in the RMSE metric, being only 0.0001 higher than the best-performing XGBOOST model, indicating a potential advantage of the CNN in reducing prediction errors.

Among traditional machine learning methods, the KNN algorithm displayed remarkable performance, with an Accuracy of 93.69% and an AUC of 97.94%, surpassing that of certain deep learning models. This finding emphasizes the significant practical value of traditional machine learning methods in certain specific tasks. The CART model performed similarly to the KNN, whereas the LR model exhibited relatively weaker performance across all metrics, particularly in Accuracy and κ, lagging behind the best model by 7.61 and 15.21 percentage points, respectively.

The relatively poor performance of LR can be attributed to its inherent linear assumptions, which may not adequately capture the complex nonlinear relationships between gully erosion and environmental factors. The superior performance of the CNN over the ANN indicates that CNN's convolutional operations are more effective in extracting spatial patterns and local feature dependencies from the high-resolution terrain data, while traditional ANN architecture may limit its ability to capture these spatial correlations, resulting in suboptimal performance.

From Table 2, it can be observed that ensemble learning methods, particularly tree-based models, exhibit significant advantages in overall performance. The optimized XGBOOST model achieved the highest level of performance across all evaluation metrics, significantly outperforming the other two models. Specifically, the optimized XGBOOST attained an Accuracy of 95.38%, a κ of 0.9077, an AUC of 99.09%, and RMSE and MAE values of 0.1850 and 0.0747, respectively. These metrics were all denoted as optimal (marked with *), highlighting the superior performance and stability of the optimized XGBOOST when handling the given dataset.

Table 2. Performance metrics of optimized models (* represents optimal values).

Metrics	Optimized XGBoost	ANN (Multi-Head Attention)	CNN (Multi-Head Attention)
Accuracy	0.9538 *	0.9227 (−0.0311)	0.9402 (−0.0136)
K	0.9077 *	0.8453 (−0.0624)	0.8804 (−0.0273)
AUC	0.9909 *	0.9770 (−0.0139)	0.9875 (−0.0034)
RMSE	0.1850 *	0.2380 (+0.0530)	0.2093 (+0.0243)
MAE	0.0747 *	0.1086 (+0.0339)	0.0861 (+0.0114)

The CNN with the Multi-Head Attention model ranked second in overall performance. Its Accuracy was 94.02%, only 1.36 percentage points lower than the optimized XGBOOST. The model also exhibited excellent performance with a κ of 0.8804 and an AUC of 98.75%, showing differences of 0.0273 and 0.0034 from the optimal model, respectively. Notably, while the attention-enhanced CNN slightly lagged behind the optimized XGBOOST in RMSE (0.2093) and MAE (0.0861) metrics, it still maintained a low error level. This suggests that the attention-enhanced CNN has significant potential to capture complex features and reduce prediction errors.

Although the ANN with Multi-Head Attention model lagged behind the other two models across all metrics, its performance remains noteworthy. The model achieved an Accuracy of 92.27%, a κ of 0.8453, and an AUC of 97.70%. Despite these metrics being somewhat lower than those of the optimal model (with the Accuracy differing by 3.11 percentage points and the AUC by 1.39 percentage points), they still demonstrate the positive impact of the Multi-Head Attention mechanism on enhancing conventional ANN performance.

Table 3 presents a performance comparison of the various machine learning and deep learning models used in this experiment, including their optimized versions, across five key evaluation metrics: Accuracy, κ, AUC, RMSE, and MAE.

Table 3. Comprehensive performance comparison between standard and optimized models (* represents optimal values).

Metrics	LR	KNN	CART	RF	BRT	ADABOOST	XGBOOST	ANN	CNN	Optimized XGBoost	ANN (Multi-Head Attention)	CNN (Multi-Head Attention)
Accuracy	0.8759 (−0.0779)	0.9369 (−0.0169)	0.9329 (−0.0209)	0.9492 (−0.0046)	0.9460 (−0.0078)	0.9413 (−0.0125)	0.9520 (−0.0018)	0.9039 (−0.0499)	0.9351 (−0.0187)	0.9538 *	0.9227 (−0.0311)	0.9402 (−0.0136)
K	0.7518 (−0.1559)	0.8738 (−0.0339)	0.8659 (−0.0418)	0.8985 (−0.0092)	0.8921 (−0.0156)	0.8826 (−0.0251)	0.9039 (−0.0038)	0.8079 (−0.0998)	0.8702 (−0.0375)	0.9077 *	0.8453 (−0.0624)	0.8804 (−0.0273)
AUC	0.9403 (−0.0506)	0.9794 (−0.0115)	0.9814 (−0.0095)	0.9892 (−0.0017)	0.9879 (−0.0030)	0.9867 (−0.0042)	0.9903 (−0.0006)	0.9683 (−0.0226)	0.9837 (−0.0072)	0.9909 *	0.9770 (−0.0139)	0.9875 (−0.0034)
RMSE	0.3523 (+0.1673)	0.2512 (+0.0662)	0.2589 (+0.0739)	0.2253 (+0.0403)	0.2323 (+0.0473)	0.2423 (+0.0573)	0.2192 (+0.0342)	0.2633 (+0.0783)	0.2193 (+0.0343)	0.1850 *	0.2380 (+0.0530)	0.2093 (+0.0243)
MAE	0.1241 (+0.0761)	0.0631 (+0.0150)	0.0671 (+0.0190)	0.0508 (+0.0027)	0.0540 (+0.0059)	0.0587 (+0.0107)	0.0480 *	0.1468 (+0.0988)	0.1030 (+0.0550)	0.0747 (+0.0267)	0.1086 (+0.0606)	0.0861 (+0.0381)

The optimized XGBOOST model demonstrated superior performance across most evaluation metrics, achieving the highest Accuracy (95.38%), a κ of 0.9077, and an AUC of 99.09%, while also recording the lowest RMSE (0.1850). Compared to the original XGBOOST, the optimized version improved the Accuracy by 0.18 percentage points and the AUC by 0.06 percentage points, thereby substantiating the effectiveness of the optimization process.

Deep learning models also exhibited significant performance improvements after incorporating the Multi-Head Attention mechanism, particularly the CNN with the Multi-Head Attention model, which achieved an Accuracy of 94.02% and an AUC of 98.75%, noticeably higher than the original CNN model. This underscores the pivotal role of the attention mechanism in enhancing the expressive capacity of deep learning models. The CNN model optimized with the Multi-Head Attention mechanism saw an increase in Accuracy to 94.02 and AUC to 98.75, yet it still did not surpass the XGBOOST model. This indicates that, within the context of high-resolution UAV data applications, tree-based ensemble learning models may be more suitable for predicting gully erosion susceptibility.

Notably, despite the overall superior performance of the optimized models, certain traditional machine learning methods still demonstrated competitiveness on specific metrics. For instance, the RF model slightly lagged behind the optimized XGBOOST with an Accuracy of 94.92% and an AUC of 98.92%, yet it outperformed most other models in the MAE metric (0.0508). This phenomenon highlights that traditional methods may still be effective choices for specific tasks or datasets.

The study results also indicate that no single model leads absolutely across all metrics. For example, although the optimized XGBOOST performed best on most metrics, the original XGBOOST model slightly outperformed it in the MAE metric (0.0480 versus 0.0747). This phenomenon underscores the necessity of adopting a multi-dimensional perspective when evaluating model performance and weighing the importance of different metrics based on specific application requirements.

Overall, the optimized models, particularly the optimized XGBOOST, exhibited the most outstanding performance across the majority of metrics. Its built-in regularization mechanisms and parallel processing capabilities enhanced the model's generalization ability and computational efficiency [59]. In contrast, models such as KNN and CART were more susceptible to performance degradation when dealing with high-dimensional data. Although deep learning models like ANNs and CNNs possess powerful nonlinear fitting capabilities, their performance was potentially hindered by the high demand for large-scale, high-quality training data within the scope of this study [75]. Furthermore, the XGBOOST model's low sensitivity to parameter settings makes it easier to tune and optimize, contributing to its superior performance in this study.

UAV data offers advantages such as high spatial resolution (up to 0.2 m), flexible acquisition, and low cost, enabling the precise capture of subtle surface changes, including micro-gullies and minor slope variations, which are critical factors influencing gully erosion occurrence [76]. High-resolution data allow models to leverage richer feature information, thereby enhancing prediction accuracy. Additionally, advancements in UAV technology have significantly improved the timeliness and controllability of data acquisition, making it suitable for small-scale, detailed environmental studies [77].

3.3. Variable Importance Analysis

This study employed the SHAP method to quantitatively assess the geomorphological factors influencing gully formation, aiming to deepen the understanding of gully development mechanisms and provide a scientific basis for formulating precise soil and water conservation strategies. As shown in Figure 5, surface roughness and altitude are the most influential factors, with importance values of 0.277195 and 0.161360, respectively, accounting for 80.7 percent of the cumulative importance. This finding underscores the dominant role of microgeomorphic features and macro topography in the process of gully erosion. The Aspect, LS Factor, and slope follow, with importance values of 0.031666, 0.020345, and 0.016166, respectively, reflecting the continued relevance of these traditional topographic parameters in gully development. Although SPI, Profile curvature, TWI, and Plan curvature have relatively lower importance (all < 0.012), they still provide supplementary information to the model, contributing to more comprehensive and accurate predictions.

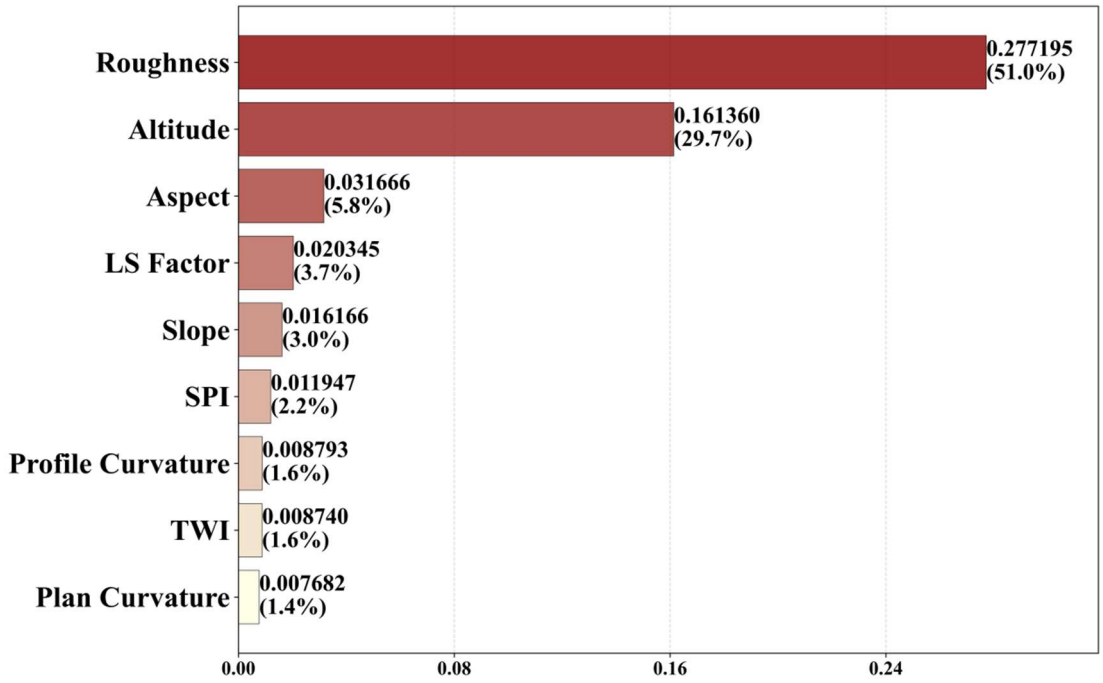

Figure 5. Relative importance of different geo-environmental factors.

Surface roughness reflects the degree of topographic undulation; areas with high roughness typically exhibit more complex landforms, such as gullies, cliffs, and steep slopes. These landforms tend to concentrate water flow, exacerbating erosion on the surface [38]. During rainfall events, water flow in areas with higher roughness is more likely to form turbulence, increasing erosive force and thereby promoting the formation and development of gullies. The significant impact of surface roughness indicates that microgeomorphic features play a crucial role in the formation of gullies, emphasizing the importance of microscale geomorphological management in erosion prevention and control. In areas prone to gully erosion, measures should be taken to reduce surface roughness, such

as land leveling or filling small gullies, to decrease water flow concentration and erosion opportunities.

The impact of altitude on gully erosion primarily lies in its regulation of environmental factors such as climate, vegetation, and soil. In higher altitude regions, there may be increased precipitation and lower temperatures, reduced vegetation cover, and poorer soil stability, making erosion more likely [78]. Conversely, changes in altitude also affect the topographic slope and water flow velocity, thereby influencing the erosion process. The importance of altitude in gully erosion highlights the central role of topographic gradients in gully development, suggesting the necessity of implementing gradient-based soil and water conservation measures in areas with varied topography.

Other topographic factors, such as the Aspect, LS Factor, slope, and Profile curvature, also have certain impacts on gully erosion. The slope is a crucial factor influencing the velocity of surface runoff; steep slopes increase the kinetic energy of water flow, intensifying erosive action [39]. Aspect affects the amount of solar radiation received, thereby influencing vegetation growth and soil moisture and indirectly impacting the erosion process [79]. While SPI and TWI showed relatively low importance in this study, these hydrological indices remain theoretically significant in erosion processes. Their lower importance values might be attributed to two factors: first, under high-resolution data conditions, these compound indices may be partially redundant with primary terrain factors like slope and roughness; second, the strong correlation between surface roughness and water flow patterns might have already captured the hydrological information that SPI and TWI typically represent. Moreover, the interaction between slope and surface roughness likely enhances their combined effect on erosion processes, as increased roughness on steeper slopes can significantly amplify water flow concentration and erosive force. In this study, the feature importance of Plan curvature was relatively low, possibly because the variations of this factor had a minor impact on gully erosion within the study area, or due to the inconspicuous spatial variability of this factor under high-resolution data. This data-driven importance analysis provides a quantitative foundation for accurately identifying key factors influencing gully development, thereby aiding in the formulation of more targeted and cost-effective soil and water conservation measures.

3.4. Gully Erosion Susceptibility Map

This study utilized the optimized XGBOOST model to generate the GESM of the study area, as depicted in Figure 6. The map intuitively displays the potential risk of gully erosion within the study area in a high-resolution spatial distribution format. The susceptibility map employed a binary classification method to divide the study area into two categories: high susceptibility (38.22 percent) and low susceptibility (61.78 percent). Although our model could generate multi-class susceptibility levels, this study adopted a binary classification approach considering the study area's practical management needs and local conservation resource constraints. This simplified classification provides clear, actionable guidance for local environmental protection departments to prioritize areas requiring gully erosion prevention measures. High susceptibility areas are depicted in deep red on the map, predominantly concentrated in regions with significant topographic variability and steep slopes, particularly in the central and northern parts of the study area. These areas may be more susceptible to gully erosion and thus require priority in the implementation of soil and water conservation measures. In contrast, low susceptibility areas are represented in light yellow, primarily located in relatively flat regions, such as the southern and eastern edges of the study area. The spatial distribution pattern of the GESM is highly consistent with topographical features, land use types, and previously identified key influencing factors (such as surface roughness and altitude). This consistency further validates the reliability and Accuracy of the model employed in this study. By identifying high susceptibility areas, scientific foundations can be provided for the formulation of land management and soil and water conservation measures.

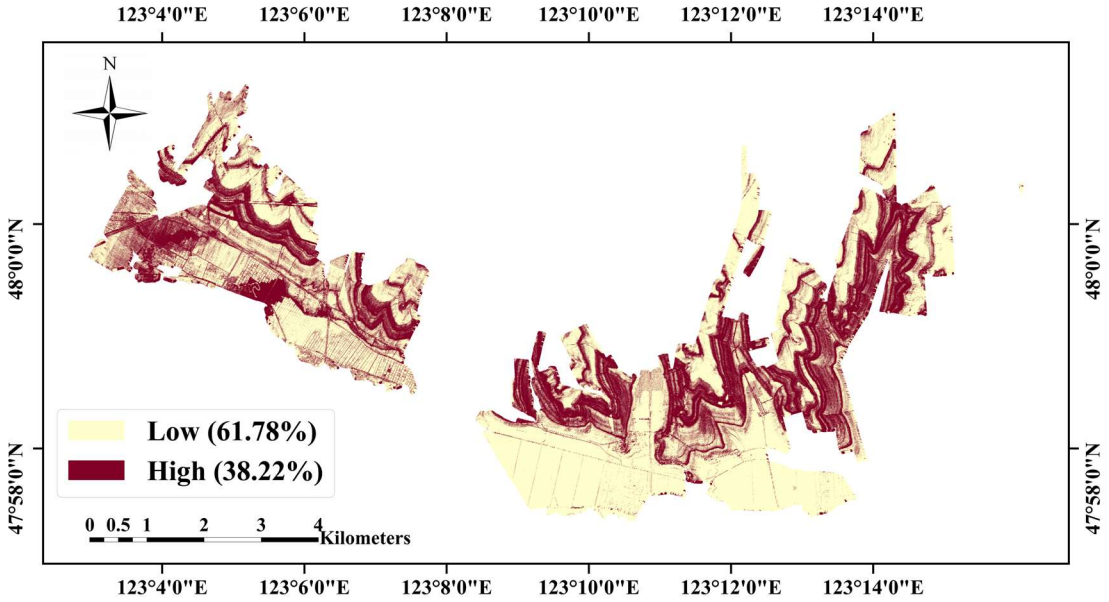

Figure 6. Gully erosion susceptibility mapping using the optimized XGBOOST mode.

3.5. Future Work

Future research could address several critical aspects of gully erosion susceptibility modeling. First, considering the data dependency of deep learning models, future studies could explore the integration of diverse data sources, including multi-temporal satellite imagery, hyperspectral data, soil organic matter content, and radioisotope measurements. This multi-source data integration could enhance model robustness and provide comprehensive insights into gully erosion processes. Additionally, to address the sensitivity of traditional models to high-dimensional data, research efforts could focus on developing advanced ensemble frameworks that combine the stability of tree-based models, the feature extraction capabilities of deep learning models, and the interpretability of traditional statistical methods. The development of adaptive modeling frameworks that automatically adjust to varying data quality conditions and dimensional characteristics would be valuable for broader applications.

Furthermore, several methodological improvements warrant investigation. The temporal dynamics of gully development require deeper examination through the incorporation of time series data to better understand and model the evolution process. The implementation of recursive feature elimination (RFE) techniques could provide a more systematic approach to feature selection, potentially optimizing input variable selection and improving model performance [80]. Moreover, future research should focus on validating the GESM results using independent regions with similar geomorphological characteristics or temporal datasets from different periods. This external validation would not only strengthen the model's reliability but also assess its transferability across different geographical contexts. Finally, model transferability and applicability need validation across different geographic contexts to ensure broader practical utility. These enhancements could collectively advance the field of gully erosion susceptibility assessment and contribute to more effective soil conservation strategies.

4. Conclusions

This study systematically evaluated the susceptibility of gully erosion in the Northeast China black soil region by comprehensively applying various advanced machine learning algorithms. Utilizing high-resolution UAV data and incorporating eleven key environmental factors such as topography and hydrology, a comprehensive predictive modeling framework was developed. Through rigorous multicollinearity analysis, we optimized the model inputs, ensuring the reliability and efficiency of the predictions.

In the model performance evaluation, the optimized XGBOOST algorithm demonstrated exceptional predictive capabilities, with an AUC value of 99.09 percent, significantly outperforming other traditional and deep learning models. This result not only validated the advantages of ensemble learning methods in handling complex geospatial data but also provided robust methodological support for future gully erosion research. Additionally, the Multi-Head Attention mechanism introduced in this study achieved significant improvements in enhancing the performance of CNN and ANN models, particularly in enhancing the models' spatial awareness and feature extraction efficiency. This innovation paves new research directions for the application of deep learning in geomorphological process modeling.

Through SHAP importance analysis, we identified roughness and altitude as the most critical factors influencing gully formation, with importance scores of 0.277195 and 0.161360, respectively. These two factors significantly outweighed others, jointly governing the occurrence of gully erosion. This finding enhanced our understanding of the mechanisms governing gully development in the Northeast China black soil region, emphasizing the pivotal roles of micro-topographical features and macro-scale topography in the erosion process. The GESM generated based on the optimal model revealed that 38.22 percent of the study area exhibits high erosion susceptibility, providing a scientific basis for the precise formulation of soil and water conservation strategies.

The innovative aspects of this study are primarily manifested in the following aspects: First, it represents the first systematic comparison of various machine learning models' performance in gully erosion prediction at a 0.2 m high-resolution scale. Second, by incorporating various optimization techniques and introducing a Multi-Head Attention mechanism, we significantly enhanced the application potential of machine learning and deep learning models in geomorphological process modeling. Lastly, the research findings provide precise spatial decision-making support for soil and water conservation practices in the black soil regions.

Overall, this study not only advanced the methodological development of gully erosion susceptibility assessment but also provided valuable empirical experience for the application of machine learning in geomorphological processes. The research findings hold significant practical implications for enhancing the precision and effectiveness of soil and water conservation measures, while simultaneously delineating new directions for future geomorphodynamic research.

Author Contributions: Conceptualization, H.L., J.J., F.D. and L.L.; Methodology, H.L., J.J. and F.D.; Software, H.L., J.J. and F.D.; Validation, H.L., J.J. and F.D.; Formal Analysis, H.L., J.J. and F.D.; Investigation, H.L., J.J., F.D., J.Z. and L.L.; Resources, F.D., J.Z. and L.L.; Data Curation, L.L., J.J. and J.Z.; Writing—Original Draft Preparation, H.L. and J.J.; Writing—Review and Editing, F.D., J.Z. and L.L.; Visualization, H.L., J.J. and F.D.; Supervision, L.L. and Y.Z.; Project Administration, L.L. and Y.Z.; Funding Acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the lnnovation Funding of the Institute of Computing Technology, Chinese Academy of Sciences, under Grant No. E261030.

Data Availability Statement: Data are available on request due to restrictions (project data privacy). The data presented in this study are available on request from the corresponding author.

Acknowledgments: We are grateful to all the staff who contributed to this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADABOOST	Adaptive Boosting
Adam	Adaptive Moment Estimation
ANN	Artificial Neural Network
AUC	Area Under the ROC Curve
AUC-ROC	Area Under the Curve-Receiver Operating Characteristic
BRT	Boosted Regression Trees
CART	Classification and Regression Trees
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
DEM	Digital Elevation Model
DOM	Digital Orthophoto Map
GDAL	Geographic Data Abstraction Library
GEFs	Geo-environmental Factors
GESM	Gully Erosion Susceptibility Map
GIS	Geographic Information System
KNN	K-Nearest Neighbors
LBFGS	Broyden–Fletcher–Goldfarb–Shanno
LR	Logistic Regression
MAE	Mean Absolute Error
QGIS	Quantum Geographic Information System
ReLU	Rectified Linear Unit
RF	Random Forest
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
RS	Remote Sensing
SMOTE	Synthetic Minority Over-Sampling Technique
SPI	Stream Power Index
SHAP	SHapley Additive exPlanations
TIF	Tagged Image File Format
TPI	Topographic Position Index
TRI	Topographic Ruggedness Index
TWI	Topographic Wetness Index
UAV	Unmanned Aerial Vehicle
VIF	Variance Inflation Factor
XGBOOST	Extreme Gradient Boosting

References

1. Poesen, J.; Nachtergaele, J.; Verstraeten, G.; Valentin, C. Gully Erosion and Environmental Change: Importance and Research Needs. *CATENA* **2003**, *50*, 91–133. [CrossRef]
2. Valentin, C.; Poesen, J.; Li, Y. Gully Erosion: Impacts, Factors and Control. *CATENA* **2005**, *63*, 132–153. [CrossRef]
3. Vanmaercke, M.; Poesen, J.; Van Mele, B.; Demuzere, M.; Bruynseels, A.; Golosov, V.; Bezerra, J.F.R.; Bolysov, S.; Dvinskih, A.; Frankl, A.; et al. How Fast Do Gully Headcuts Retreat? *Earth-Sci. Rev.* **2016**, *154*, 336–355. [CrossRef]
4. Castillo, C.; Gómez, J.A. A Century of Gully Erosion Research: Urgency, Complexity and Study Approaches. *Earth-Sci. Rev.* **2016**, *160*, 300–319. [CrossRef]
5. Gayen, A.; Pourghasemi, H.R.; Saha, S.; Keesstra, S.; Bai, S. Gully Erosion Susceptibility Assessment and Management of Hazard-Prone Areas in India Using Different Machine Learning Algorithms. *Sci. Total Environ.* **2019**, *668*, 124–138. [CrossRef]
6. Montgomery, D.R.; Dietrich, W.E. A Physically Based Model for the Topographic Control on Shallow Landsliding. *Water Resour. Res.* **1994**, *30*, 1153–1171. [CrossRef]
7. Arabameri, A.; Pourghasemi, H.R. Spatial Modeling of Gully Erosion Using Linear and Quadratic Discriminant Analyses in GIS and R. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 299–321, ISBN 978-0-12-815226-3.
8. Pourghasemi, H.R.; Yousefi, S.; Kornejady, A.; Cerdà, A. Performance Assessment of Individual and Ensemble Data-Mining Techniques for Gully Erosion Modeling. *Sci. Total Environ.* **2017**, *609*, 764–775. [CrossRef]

9. Arabameri, A.; Pradhan, B.; Rezaei, K.; Yamani, M.; Pourghasemi, H.R.; Lombardo, L. 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.* **2018**, *29*, 4035–4049. [CrossRef]
10. Conoscenti, C.; Angileri, S.; Cappadonia, C.; Rotigliano, E.; Agnesi, V.; Märker, M. Gully Erosion Susceptibility Assessment by Means of GIS-Based Logistic Regression: A Case of Sicily (Italy). *Geomorphology* **2014**, *204*, 399–411. [CrossRef]
11. Avand, M.; Janizadeh, S.; Naghibi, S.A.; Pourghasemi, H.R.; Khosrobeigi Bozchaloei, S.; Blaschke, T. A Comparative Assessment of Random Forest and K-Nearest Neighbor Classifiers for Gully Erosion Susceptibility Mapping. *Water* **2019**, *11*, 2076. [CrossRef]
12. Pourghasemi, H.R.; Sadhasivam, 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. *Geosci. Front.* **2020**, *11*, 2207–2219. [CrossRef]
13. Amiri, M.; Pourghasemi, H.R.; Ghanbarian, G.A.; Afzali, S.F. Assessment of the Importance of Gully Erosion Effective Factors Using Boruta Algorithm and Its Spatial Modeling and Mapping Using Three Machine Learning Algorithms. *Geoderma* **2019**, *340*, 55–69. [CrossRef]
14. 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. [CrossRef] [PubMed]
15. Garosi, Y.; Sheklabadi, M.; Conoscenti, C.; Pourghasemi, H.R.; Van Oost, K. Assessing the Performance of GIS- Based Machine Learning Models with Different Accuracy Measures for Determining Susceptibility to Gully Erosion. *Sci. Total Environ.* **2019**, *664*, 1117–1132. [CrossRef]
16. Rahmati, O.; Tahmasebipour, N.; Haghizadeh, A.; Pourghasemi, H.R.; Feizizadeh, B. Evaluation of Different Machine Learning Models for Predicting and Mapping the Susceptibility of Gully Erosion. *Geomorphology* **2017**, *298*, 118–137. [CrossRef]
17. Chen, W.; Lei, X.; Chakrabortty, R.; Chandra Pal, S.; Sahana, M.; Janizadeh, S. Evaluation of Different Boosting Ensemble Machine Learning Models and Novel Deep Learning and Boosting Framework for Head-Cut Gully Erosion Susceptibility. *J. Environ. Manag.* **2021**, *284*, 112015. [CrossRef]
18. Hembram, T.K.; Saha, S.; Pradhan, B.; Abdul Maulud, K.N.; Alamri, A.M. Robustness Analysis of Machine Learning Classifiers in Predicting Spatial Gully Erosion Susceptibility with Altered Training Samples. *Geomat. Nat. Hazards Risk* **2021**, *12*, 794–828. [CrossRef]
19. Arabameri, A.; Chandra Pal, S.; Costache, R.; Saha, A.; Rezaie, F.; Seyed Danesh, A.; Pradhan, B.; Lee, S.; Hoang, N.-D. Prediction of Gully Erosion Susceptibility Mapping Using Novel Ensemble Machine Learning Algorithms. *Geomat. Nat. Hazards Risk* **2021**, *12*, 469–498. [CrossRef]
20. Eloudi, H.; Hssaisoune, M.; Reddad, H.; Namous, M.; Ismaili, M.; Krimissa, S.; Ouayah, M.; Bouchaou, L. Robustness of Optimized Decision Tree-Based Machine Learning Models to Map Gully Erosion Vulnerability. *Soil Syst.* **2023**, *7*, 50. [CrossRef]
21. Arabameri, A.; Asadi Nalivan, O.; Chandra Pal, S.; Chakrabortty, R.; Saha, A.; Lee, S.; Pradhan, B.; Tien Bui, D. Novel Machine Learning Approaches for Modelling the Gully Erosion Susceptibility. *Remote Sens.* **2020**, *12*, 2833. [CrossRef]
22. Garosi, Y.; Sheklabadi, M.; Pourghasemi, H.R.; Besalatpour, A.A.; Conoscenti, C.; Van Oost, K. Comparison of Differences in Resolution and Sources of Controlling Factors for Gully Erosion Susceptibility Mapping. *Geoderma* **2018**, *330*, 65–78. [CrossRef]
23. Saha, S.; Sarkar, R.; Thapa, G.; Roy, J. Modeling Gully Erosion Susceptibility in Phuentsholing, Bhutan Using Deep Learning and Basic Machine Learning Algorithms. *Environ. Earth Sci.* **2021**, *80*, 295. [CrossRef]
24. Wu, Y.; Cheng, H. Monitoring of Gully Erosion on the Loess Plateau of China Using a Global Positioning System. *CATENA* **2005**, *63*, 154–166. [CrossRef]
25. Shahabi, H.; Jarihani, B.; Tavakkoli Piralilou, S.; Chittleborough, D.; Avand, M.; Ghorbanzadeh, O. A Semi-Automated Object-Based Gully Networks Detection Using Different Machine Learning Models: A Case Study of Bowen Catchment, Queensland, Australia. *Sensors* **2019**, *19*, 4893. [CrossRef]
26. Zhang, X.; Zhang, S.; Meng, X.; Zhang, G.; Zang, D.; Han, Y.; Ai, H.; Liu, H. Remote Sensing Image Segmentation of Gully Erosion in a Typical Black Soil Area in Northeast China Based on Improved DeepLabV3+ Model. *Ecol. Inform.* **2024**, *84*, 102929. [CrossRef]
27. Gafurov, A.M.; Yermolayev, O.P. Automatic Gully Detection: Neural Networks and Computer Vision. *Remote Sens.* **2020**, *12*, 1743. [CrossRef]
28. Dong, F.; Jin, J.; Li, L.; Li, H.; Zhang, Y. A Multi-Scale Content-Structure Feature Extraction Network Applied to Gully Extraction. *Remote Sens.* **2024**, *16*, 3562. [CrossRef]
29. Yang, A.; Wang, C.; Pang, G.; Long, Y.; Wang, L.; Cruse, R.M.; Yang, Q. Gully Erosion Susceptibility Mapping in Highly Complex Terrain Using Machine Learning Models. *IJGI* **2021**, *10*, 680. [CrossRef]
30. Mohebzadeh, H.; Biswas, A.; Rudra, R.; Daggupati, P. Machine Learning Techniques for Gully Erosion Susceptibility Mapping: A Review. *Geosciences* **2022**, *12*, 429. [CrossRef]
31. Setargie, T.A.; Tsunekawa, A.; Haregeweyn, N.; Tsubo, M.; Fenta, A.A.; Berihun, M.L.; Sultan, D.; Yibeltal, M.; Ebabu, K.; Nzioki, B.; et al. Random Forest–Based Gully Erosion Susceptibility Assessment across Different Agro-Ecologies of the Upper Blue Nile Basin, Ethiopia. *Geomorphology* **2023**, *431*, 108671. [CrossRef]
32. Bammou, Y.; Benzougagh, B.; Abdessalam, O.; Brahim, I.; Kader, S.; Spalevic, V.; Sestras, P.; Ercişli, S. Machine Learning Models for Gully Erosion Susceptibility Assessment in the Tensift Catchment, Haouz Plain, Morocco for Sustainable Development. *J. Afr. Earth Sci.* **2024**, *213*, 105229. [CrossRef]

33. Lei, X.; Chen, W.; Avand, M.; Janizadeh, S.; Kariminejad, N.; Shahabi, H.; Costache, R.; Shahabi, H.; Shirzadi, A.; Mosavi, A. GIS-Based Machine Learning Algorithms for Gully Erosion Susceptibility Mapping in a Semi-Arid Region of Iran. *Remote Sens.* **2020**, *12*, 2478. [CrossRef]
34. Zevenbergen, L.W.; Thorne, C.R. Quantitative Analysis of Land Surface Topography. *Earth Surf. Process. Landf.* **1987**, *12*, 47–56. [CrossRef]
35. Shary, P.A.; Sharaya, L.S.; Mitusov, A.V. Fundamental Quantitative Methods of Land Surface Analysis. *Geoderma* **2002**, *107*, 1–32. [CrossRef]
36. Różycka, M.; Migoń, P.; Michniewicz, A. Topographic Wetness Index and Terrain Ruggedness Index in Geomorphic Characterisation of Landslide Terrains, on Examples from the Sudetes, SW Poland. *Z. Geomorphol. Suppl. Issues* **2017**, *61*, 61–80. [CrossRef]
37. Reu, J.D.; Bourgeois, J.; Bats, M.; Zwertvaegher, A.; Gelorini, V.; Smedt, P.D.; Chu, W.; Antrop, M.; Maeyer, P.D.; Finke, P.; et al. Application of the Topographic Position Index to Heterogeneous Landscapes. *Geomorphology* **2013**, *186*, 39–49. [CrossRef]
38. Smith, M.W. Roughness in the Earth Sciences. *Earth-Sci. Rev.* **2014**, *136*, 202–225. [CrossRef]
39. Bircher, P.; Liniger, H.P.; Prasuhn, V. Comparing Different Multiple Flow Algorithms to Calculate RUSLE Factors of Slope Length (L) and Slope Steepness (S) in Switzerland. *Geomorphology* **2019**, *346*, 106850. [CrossRef]
40. Mattivi, P.; Franci, F.; Lambertini, A.; Bitelli, G. TWI Computation: A Comparison of Different Open Source GISs. *Open Geospat. Data Softw. Stand.* **2019**, *4*, 6. [CrossRef]
41. Moore, I.D.; Nieber, J.L. Landscape Assessment of Soil Erosion and Nonpoint Source Pollution. *J. Minn. Acad. Sci.* **1989**, *55*, 18–25.
42. Alin, A. Multicollinearity. *WIREs Comput. Stats.* **2010**, *2*, 370–374. [CrossRef]
43. O'brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **2007**, *41*, 673–690. [CrossRef]
44. Adnan, N.; Ahmad, M.H.; Adnan, R. A Comparative Study On Some Methods For Handling Multicollinearity Problems. *Matematika* **2006**, *22*, 109–119.
45. Maalouf, M. Logistic Regression in Data Analysis: An Overview. *Int. J. Data Anal. Tech. Strateg.* **2011**, *3*, 281. [CrossRef]
46. Peng, C.-Y.J.; Lee, K.L.; Ingersoll, G.M. An Introduction to Logistic Regression Analysis and Reporting. *J. Educ. Res.* **2002**, *96*, 3–14. [CrossRef]
47. LaValley, M.P. Logistic Regression. *Circulation* **2008**, *117*, 2395–2399. [CrossRef]
48. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inform. Theory* **1967**, *13*, 21–27. [CrossRef]
49. Garcia, S.; Derrac, J.; Cano, J.R.; Herrera, F. Prototype Selection for Nearest Neighbor Classification: Taxonomy and Empirical Study. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 417–435. [CrossRef]
50. Loh, W. Fifty Years of Classification and Regression Trees. *Int. Statistical Rev.* **2014**, *82*, 329–348. [CrossRef]
51. Loh, W. Classification and Regression Trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [CrossRef]
52. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
53. Biau, G. Analysis of a Random Forests Model. *J. Mach. Learn. Res.* **2012**, *13*, 1063–1095.
54. De'ath, G. Boosted trees for ecological modeling and prediction. *Ecology* **2007**, *88*, 243–251. [CrossRef] [PubMed]
55. Elith, J.; Leathwick, J.R.; Hastie, T. A Working Guide to Boosted Regression Trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [CrossRef]
56. Hutchinson, R.; Liu, L.-P.; Dietterich, T. Incorporating Boosted Regression Trees into Ecological Latent Variable Models. *AAAI* **2011**, *25*, 1343–1348. [CrossRef]
57. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
58. Schapire, R.E.; Singer, Y. Improved Boosting Algorithms Using Confidence-Rated Predictions. In Proceedings of the eleventh annual conference on Computational Learning Theory, Madison, WI, USA, 24 July 1998; ACM: New York, NY, USA, 1998; pp. 80–91.
59. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
60. Zou, J.; Han, Y.; So, S.-S. Overview of Artificial Neural Networks. In *Artificial Neural Networks: Methods and Applications*; Livingstone, D.J., Ed.; Humana Press: Totowa, NJ, USA, 2009; pp. 14–22, ISBN 978-1-60327-101-1.
61. Wu, Y.; Feng, J. Development and Application of Artificial Neural Network. *Wirel. Pers. Commun.* **2018**, *102*, 1645–1656. [CrossRef]
62. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
63. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent Advances in Convolutional Neural Networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]
64. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 6000–6010.
65. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2021**, arXiv:2010.11929.

66. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 25 July 2019; ACM: New York, NY, USA, 2019; pp. 2623–2631.
67. Yao, Y.; Rosasco, L.; Caponnetto, A. On Early Stopping in Gradient Descent Learning. *Constr. Approx.* **2007**, *26*, 289–315. [CrossRef]
68. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
69. Fernandez, A.; Garcia, S.; Herrera, F.; Chawla, N.V. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-Year Anniversary. *J. Artif. Intell. Res.* **2018**, *61*, 863–905. [CrossRef]
70. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 4768–4777.
71. Vieira, S.M.; Kaymak, U.; Sousa, J.M.C. Cohen's Kappa Coefficient as a Performance Measure for Feature Selection. In Proceedings of the International Conference on Fuzzy Systems, Barcelona, Spain, 18–23 July 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1–8.
72. Bradley, A.P. The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [CrossRef]
73. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
74. Willmott, C.; Matsuura, K. Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
75. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *J. Big Data* **2021**, *8*, 53. [CrossRef]
76. Tarolli, P. High-Resolution Topography for Understanding Earth Surface Processes: Opportunities and Challenges. *Geomorphology* **2014**, *216*, 295–312. [CrossRef]
77. Colomina, I.; Molina, P. Unmanned Aerial Systems for Photogrammetry and Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97. [CrossRef]
78. Montgomery, D.R. Slope Distributions, Threshold Hillslopes, and Steady-State Topography. *Am. J. Sci.* **2001**, *301*, 432–454. [CrossRef]
79. Zapata-Rios, X.; Brooks, P.D.; Troch, P.A.; McIntosh, J.; Guo, Q. Influence of Terrain Aspect on Water Partitioning, Vegetation Structure and Vegetation Greening in High-elevation Catchments in Northern New Mexico. *Ecohydrology* **2016**, *9*, 782–795. [CrossRef]
80. Barzani, A.R.; Pahlavani, P.; Ghorbanzadeh, O.; Gholamnia, K.; Ghamisi, P. Evaluating the Impact of Recursive Feature Elimination on Machine Learning Models for Predicting Forest Fire-Prone Zones. *Fire* **2024**, *7*, 440. [CrossRef]