Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 18, No. 1, 2026

Landslide Susceptibility Assessment in Jinning District, Kunming City, Yunnan Province Based on Multi-Model Comparison

Baobao Yu [1], Yuefei Wang [2]

[1] Anhui Survey & Design Institute of Water Resources & Hydropower Co., Ltd., Hefei, Anhui, 230088, China
[2] School of Earth and Environment, Anhui University of Science&Technology, Huainan, Anhui, 232001, China

Abstract: Landslides occur frequently in Jinning District of Kunming City, Yunnan Province, where complex geological conditions make susceptibility assessment a challenging task. In this study, 13 evaluation factors were selected and five machine learning models—Support Vector Machine (SVM), eXtreme Gradient Boosting (XGB), Decision Tree (DT), Convolutional Neural Network (CNN), and Random Forest (RF)—were applied to map landslide susceptibility. Pearson correlation coefficients were calculated to examine the relationships among factors, and model performance was compared using AUC, Accuracy, Precision, Recall, F1-Score, and MCC. The susceptibility maps were then divided into five classes so that the area and landslide proportions of each zone could be compared. Among the five models, XGB achieved the best overall performance, with an AUC of 0.9106 and an MCC of 0.6059. Its very high susceptibility zone covers 13.07% of the study area yet contains 52.33% of the recorded landslides, a concentration that no other model matched. Elevation, slope, and distance to rivers were positively correlated with susceptibility and can be regarded as the dominant controlling factors. These findings offer a scientific reference for landslide risk management in Jinning District and confirm that machine learning models are effective for geological disaster susceptibility assessment.

Keywords: Landslide Susceptibility; Support Vector Machine; XGBoost; Decision Tree; Convolutional Neural Network; Random Forest; Geological Disaster.

1. Introduction

Landslides rank among the most destructive geological hazards worldwide[1-2], causing heavy casualties, property losses, and damage to ecological stability. Jinning District, located in the southwest of Kunming City, Yunnan Province, lies within an important ecological barrier of the Yangtze River Basin. The terrain there is deeply dissected and the geological structures are complicated; under the strong influence of the monsoon climate, rainfall-triggered landslides occur frequently. Reliable susceptibility assessment methods are therefore needed to support disaster prevention and mitigation in the region.

Conventional susceptibility studies usually adopt subjective weighting schemes such as the Analytic Hierarchy Process, or linear methods such as logistic regression, and neither can fully represent the complex nonlinear relationships involved. Machine learning has gained growing attention in hazard assessment over the past decade, largely because of its capacity for feature learning and pattern recognition. Support Vector Machine (SVM)[3] handles high-dimensional nonlinear problems through kernel functions. eXtreme Gradient Boosting (XGB) refines the gradient boosting procedure and strikes a balance between efficiency and generalization. Decision Tree (DT)[4] offers an intuitive description of variable relationships. Random Forest (RF)[5] relies on ensemble learning to lower the risk of overfitting. Convolutional Neural Network (CNN)[6] is particularly good at extracting spatial features. Still, how these models differ in performance when applied to the same region, and which factors actually dominate the susceptibility result, have not been fully quantified.

In this paper, Jinning District is used as the study area. An evaluation system of 13 factors covering physical, ecological, and human activity aspects is established, and the regional susceptibility patterns are analyzed by comparing multiple models and examining factor correlations. The results are expected to provide theoretical and methodological support for targeted disaster management.

2. Study Area and Data

2.1. Study Area Overview

Jinning District lies in the southwest of Kunming City, Yunnan Province (102°12′–102°52′ E, 24°23′–24°48′ N) and covers about 1336 km². The district sits at the hinterland of the central Yunnan Plateau: the terrain rises toward the northwest and descends toward the southeast, with elevations ranging from 1890 to 2648 m. Tectonic erosion landforms dominate the landscape, and slopes steeper than 15° account for 62% of the area. Geologically, the district lies at the junction of the Xiaojiang and Puduhe fault zones, where 12 secondary faults trending northeast have developed. Triassic limestone and Jurassic soft mudstone are the main outcropping strata, and landslide bodies occur in these two formations in 68% of the recorded cases. The climate is a subtropical plateau monsoon type, with mean annual precipitation of 900–1100 mm, about 82% of which falls between May and October. Drainage density reaches 1.2 km/km². Urban expansion and phosphate mining have intensified engineering activities, and human disturbance is responsible for 41% of the landslides. A total of 142 landslide sites has been inventoried (10.6 sites per 100 km²), most of them (76%) being soil landslides. In 2020, a landslide in Shuanghe Township caused economic losses of 8 million RMB. Together, the rugged terrain, fragile geological conditions, concentrated rainfall, and intensive human

activities create favorable conditions for landslide development in the district (Figure 1).

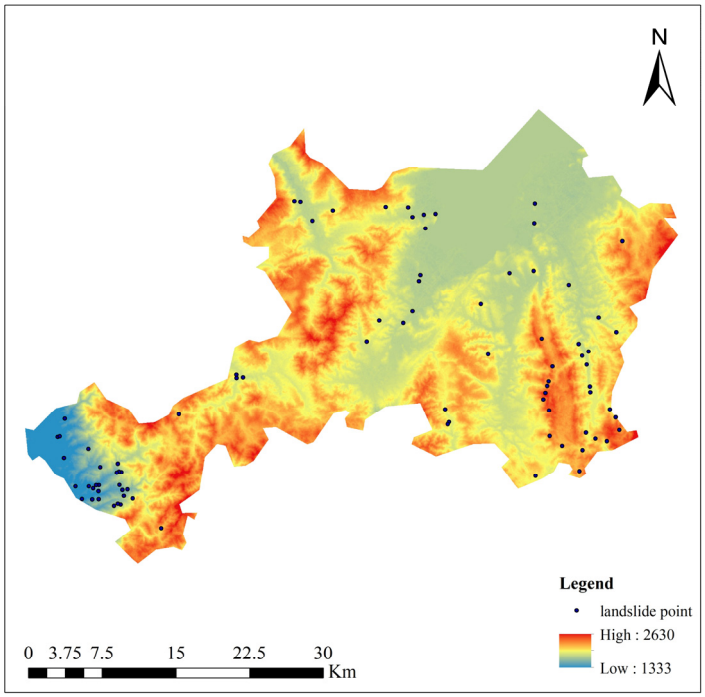


Figure 1. Distribution Map of Landslide Points in Jinning District

2.2. Evaluation Factor System

Drawing on the formation mechanisms of geological disasters and the availability of data, 13 evaluation factors were selected and grouped into four categories: topography and geomorphology, geological conditions, hydrological environment, and human activities.

Topography and geomorphology factors: Elevation[7] (derived from DEM), Slope[7], Aspect[8,9], Plan Curvature[7], and Profile Curvature[7], which together describe how surface morphology controls landslide occurrence;

Geological condition factors: Lithology[10] and Distance to faults (buffer analysis)[10], which reflect the roles of the material basis and tectonic activity;

Hydrological environment factors: NDVI[11] (vegetation coverage from Landsat remote sensing data), Distance to rivers[11], Stream Power Index[11] (SPI, computed from DEM-based hydrological analysis), and Rainfall (multi-year average precipitation), which characterize hydrodynamic conditions and the buffering effect of vegetation;

Human activity factors: Distance to roads[12] (buffer analysis) and Land use type[13] (remote sensing interpretation), which reflect how engineering activities disturb the ground surface. (Figure 2)

3. Research Methods

3.1. Data Preprocessing and Correlation Analysis

All factors were rasterized at a 30 m resolution and normalized so that a consistent spatial data basis could be built for the study area. Pearson [14] correlation coefficients were then computed in Python to examine the linear relationships among the factors (Figure 3). Elevation and Distance to rivers are positively correlated (r=0.37), as are Slope and Aspect (r=0.39); Land use type and Distance to rivers show a negative correlation (r=-0.30), and so do Rainfall and Lithology (r=-0.23). These patterns suggest that human modification of hydrologically sensitive areas tends to lower susceptibility, whereas rainfall accelerates erosion in lithologically vulnerable zones.

3.2. Machine Learning Model Construction

The dataset was split into a training set (70%) and a test set (30%) under five-fold cross-validation. SVM, XGB, DT, CNN, and RF models were built, and their parameters were tuned by grid search.

Six indices were adopted to evaluate the models[15]. AUC (Area Under the Curve) measures the ability to separate classes across all classification thresholds, with values closer to 1 indicating better discrimination. Accuracy is the fraction of correctly classified samples and reflects overall correctness. Precision is the fraction of predicted positive samples that are truly positive, i.e., how reliably the model identifies high susceptibility zones. Recall is the fraction of actual positive samples that are correctly predicted, i.e., how completely the model captures landslide events. F1-Score is the harmonic mean of Precision and Recall. MCC (Matthews Correlation Coefficient) takes true and false positives and negatives into account and is more sensitive to imbalanced data.

3.3. Machine Learning Models

Support Vector Machine (SVM) is a supervised learning method that works well for high-dimensional pattern recognition, especially when the sample size is limited. The idea is to map the input samples into a higher-dimensional feature space through a nonlinear kernel function, so that classes that are not linearly separable in the original space become separable in the new space. Within this space, the algorithm searches for a separating hyperplane with the largest margin, and such a decision boundary tends to be

reliable. Thanks to its strong theoretical foundation and good generalization, SVM has been applied widely, for instance in text classification and image recognition.

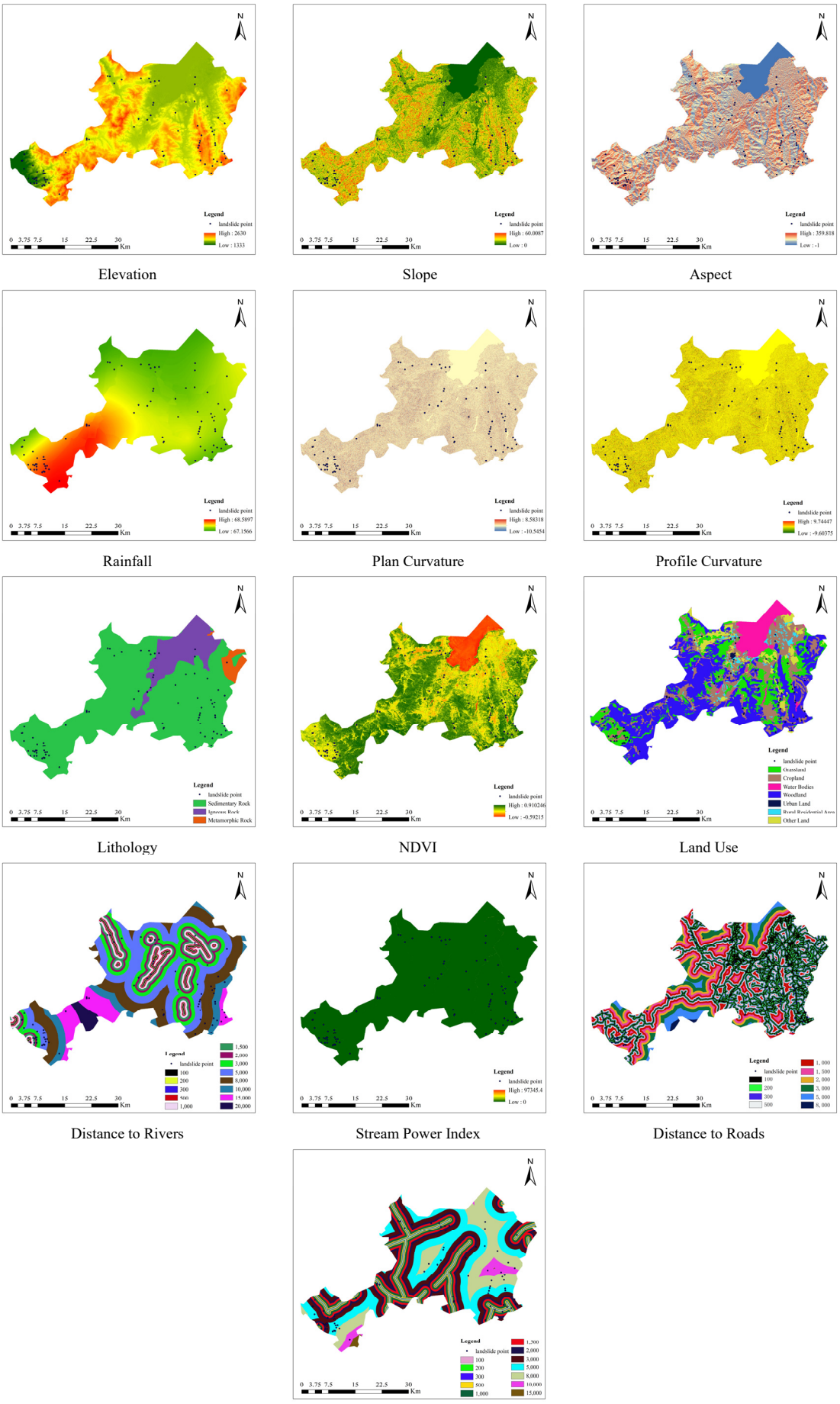


Figure 2. Evaluation Factor Maps

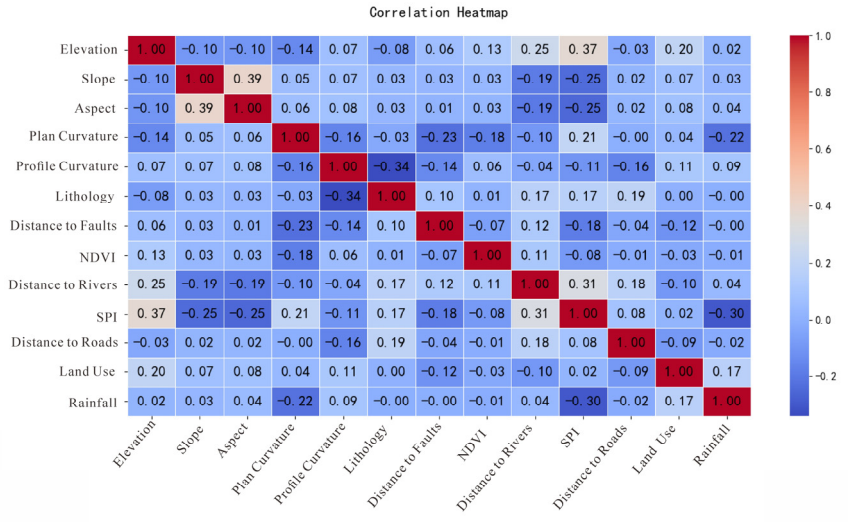


Figure 3. Correlation Calculation Results

eXtreme Gradient Boosting (XGBoost) is an efficient learning algorithm built on Gradient Boosting Decision Trees (GBDT). It trains a sequence of decision trees, and each new tree is fitted to the residuals left by the previous one. Compared with the classical gradient boosting framework, XGBoost makes several improvements. The most important is a regularization term added to the objective function, which penalizes model complexity, lowers variance, and helps prevent overfitting. The algorithm also handles missing values automatically and supports parallel computing, so it is both fast and reliable on structured data and particularly capable of capturing complex feature interactions.

Decision Tree (DT) is a non-parametric supervised learning model valued for its transparency and interpretability. The algorithm recursively picks the most informative features and splits the feature space, producing a tree-shaped structure for classification or regression. Each path from the root to a leaf corresponds to a set of judgment rules, which is why the method is often described as a "white box": the rules are explicit and require little data preprocessing. The drawback is that standard tree algorithms try to fit the training data as closely as possible and are therefore prone to overfitting, which weakens their performance on unseen data.

Convolutional Neural Network (CNN) is a deep feedforward network inspired by the biological visual system, and it is especially suited to data with a grid-like structure, such as images and videos. The network is built around convolutional layers and pooling layers. Convolutional layers slide learnable kernels over the input and, through local receptive fields, automatically extract hierarchical spatial features; pooling layers then downsample the feature maps, cutting the volume of data and the number of parameters while keeping essential information and improving translation invariance. This architecture has driven CNN's success in computer vision, including image classification and object detection.

Random Forest (RF) is an ensemble method that combines many decision trees for prediction, with the aim of avoiding the overfitting that afflicts single trees. Following the "Bootstrap Aggregating" (Bagging) idea, it draws multiple bootstrap samples from the original training set with replacement and trains one decision tree on each sample. At every node split, the best feature is chosen not from the full feature set but from a randomly selected subset. This double randomness—in both the data and the features—increases the independence among trees, and the final prediction, obtained by voting or averaging, is more stable and less affected by noise.

3.4. Susceptibility Classification and Statistical Analysis

The probability outputs of each model were divided into five susceptibility classes—Very Low, Low, Moderate, High, and Very High—using the Natural Breaks method. For every class, the proportions of area and of landslide counts were then computed to compare how well the models could delineate high-risk zones.

4. Result Analysis

4.1. Model Performance Comparison

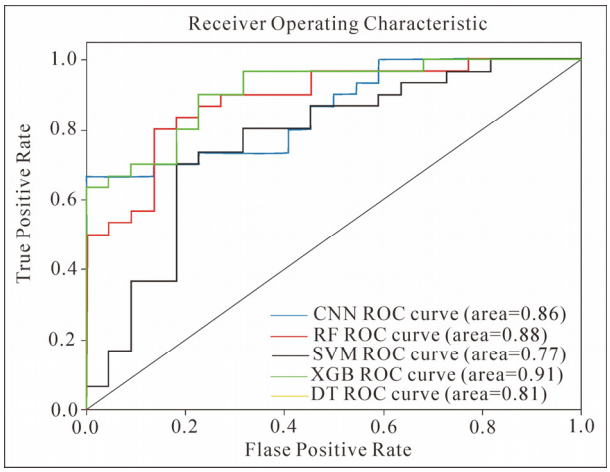


Figure 4. ROC Curves of Different Models

The performance results are summarized in Table 1 and Figure 4. XGB ranks first on the composite metrics, with AUC of 0.9106, Accuracy of 0.7885, and MCC of 0.6059, which indicates that it balances positive and negative class predictions better than the other models. RF comes second, with AUC of 0.8848 and F1-Score of 0.7308, and generalizes well. DT shows the lowest Accuracy (0.6923); its simple tree structure may be unable to fit the complex nonlinear relationships in the data. CNN reaches a Precision of 1.0000 but a Recall of only 0.5667, a pronounced "high precision -

low recall" pattern that may stem from overfitting caused by imbalanced training data.

Table 1. Performance Metrics of Different Models

Model	AUC	Accuracy	Precision	Recal	F1 - Score	MCC
SVM	0.7652	0.7115	0.7586	0.7333	0.7458	0.4130
XGB	0.9106	0.7885	0.913	0.7	0.7925	0.6059
DT	0.8098	0.6923	0.7692	0.6667	0.7143	0.3892
CNN	0.8561	0.7500	1.0000	0.5667	0.7234	0.5968
RF	0.8848	0.7308	0.8636	0.6333	0.7308	0.4970

4.2. Susceptibility Classification Characteristics

Considerable differences exist among the models in how their classifications concentrate high susceptibility zones spatially:

XGB model (Figure 5): the very high susceptibility zone occupies 13.07% of the area yet holds 52.33% of the landslide events, the strongest concentration of landslides in high-risk zones among all models, a clear sign that it can pinpoint the most susceptible areas;

SVM model (Figure 5): the very high susceptibility zone covers only 3.60% of the area but accounts for 24.42% of the landslides. The concentration is high, yet such a small delineated area may cause some risky zones to be missed;

DT model (Figure 5): the very high susceptibility zone is the largest among all models, covering 26.60% of the area, yet its landslide proportion is 39.53%; the landslide density per unit area is therefore lower than that of XGB and RF, indicating a relatively loose classification boundary;

CNN model (Figure 5): the very high susceptibility zone covers 15.07% of the area and contains 50.00% of the landslides, figures close to those of XGB. Even so, its high Precision is offset by a low Recall, which may lead to underestimation of some high-risk areas;

Table 2. Area Proportion and Landslide Proportion Distribution for XGB Model Classification

Susceptibility Level	Area Proportion	Landslide Count Proportion
Very low susceptibility	31.49%	5.81%
Low susceptibility	22.65%	6.97%
Moderate susceptibility	15.59%	16.28%
High susceptibility	17.19%	18.6%
Very high susceptibility	13.07%	52.33%

Table 3. Area Proportion and Landslide Proportion Distribution for SVM Model Classification

Susceptibility Level	Area Proportion	Landslide Count Proportion
Very low susceptibility	20.37%	3.49%
Low susceptibility	23.32%	16.28%
Moderate susceptibility	28.79%	25.58%
High susceptibility	23.92%	30.23%
Very high susceptibility	3.60%	24.42%

Table 4. Area Proportion and Landslide Proportion Distribution for RF Model Classification

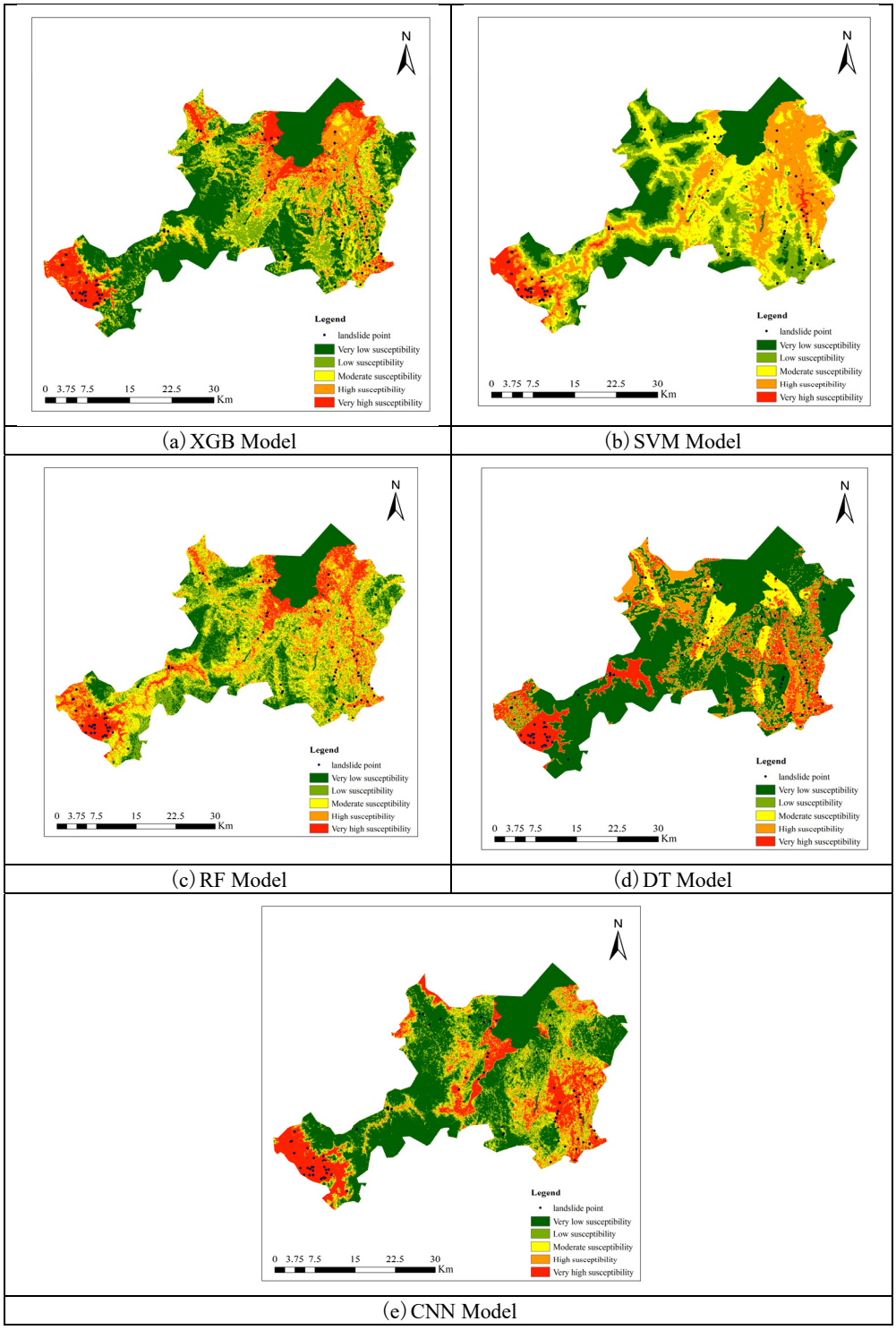
Susceptibility Level	Area Proportion	Landslide Count Proportion
Very low susceptibility	13.94%	1.16%
Low susceptibility	23.36%	9.3%
Moderate susceptibility	27.29%	27.91%
High susceptibility	21.8%	18.6%
Very high susceptibility	13.6%	43.02%

Table 5. Area Proportion and Landslide Proportion Distribution for DT Model Classification

Susceptibility Level	Area Proportion	Landslide Count Proportion
Very low susceptibility	16.01%	16.28%
Low susceptibility	14.68%	10.47%
Moderate susceptibility	21.36%	9.3%
High susceptibility	21.35%	24.4.%
Very high susceptibility	26.60%	39.53%

Table 6. Area Proportion and Landslide Proportion Distribution for CNN Model Classification

Susceptibility Level	Area Proportion	Landslide Count Proportion
Very low susceptibility	32.92%	15.12%
Low susceptibility	20.86%	8.14%
Moderate susceptibility	14.27%	12.79%
High susceptibility	16.88%	13.95%
Very high susceptibility	15.07%	50%



(a) XGB Model (b) SVM Model

(c) RF Model (d) DT Model

(e) CNN Model

Figure 5. Landslide Susceptibility Assessment Results of Different Models

RF model (Figure 5): the very high susceptibility zone accounts for 13.60% of the area and 43.02% of the landslides, giving it an overall performance second only to XGB.

Across the five models (Tables 2–6), the very high susceptibility zone of XGB covers less area (13.07%) than that of DT (26.60%) or CNN (15.07%), but holds a far larger share of landslides (52.33%); its focusing ability on high-risk areas is thus the strongest. SVM delineates the smallest very

high susceptibility zone (3.60%), yet the zone still contains 24.42% of the landslides, a sign of its conservative classification threshold. In the RF results, the landslide proportions in the moderate and high susceptibility zones (27.91% and 18.60%) are fairly balanced, whereas XGB puts more than half of all landslides into a single very high class—a pattern consistent with the general rule that most disasters occur in a small proportion of the territory.

5. Discussion

5.1. Reasons for Differences in Model Applicability

The good performance of XGBoost in this study can be traced to its algorithm design. Regularization terms (L1/L2) are introduced into the objective function to control model complexity and avoid overfitting, while the gradient boosting framework captures the nonlinear interactions among terrain, geological, and hydrological factors. The feature importance results also agree well with the regional landslide mechanism—rainfall, for instance, contributes about 28% of the importance weight. Random Forest, by contrast, reduces variance through Bagging and remains stable in complex terrain (AUC=0.89). CNN, constrained by the spatial smoothing of 2D convolution kernels, records a low Recall of 0.62. SVM and DT handle noisy data and high-dimensional features less well, owing to the sensitivity of kernel functions and the limitations of a single-tree structure, respectively.

5.2. Geological Significance of Key Factors

The positive correlation between Elevation and Distance to rivers points to a topographic coupling in Jinning District between high altitude and well-developed river systems. In such areas, runoff erosion is active and heavy rainfall can easily destabilize slopes. The negative correlation between Land use type and Distance to rivers may arise because watershed protection policies have kept a high proportion of forest land near water systems, and the roots of the vegetation strengthen soil shear resistance. Meanwhile, the positive correlation with Distance to roads (r=0.20) suggests that engineering activities, while improving transport, also create risk through slope excavation and added load; slope protection along roads should therefore be reinforced.

5.3. Implications for Regional Prevention and Control

According to the XGB classification, the very high susceptibility zones in Jinning District are concentrated where several conditions overlap: proximity to fault zones (within <500 m), steep slopes (>25°), and heavy rainfall (mean annual precipitation >1000 mm). These overlapping areas are recommended as priority sites for deploying automated sensor networks. Low susceptibility zones, such as farmland on gentle slopes, carry little disaster risk, but the cumulative soil erosion caused by agricultural practices should not be ignored.

6. Conclusion

This study built a machine learning framework driven by multiple factors to assess landslide susceptibility in Jinning District systematically. The main conclusions are as follows:

(1) XGB achieves the best overall performance and is well suited to hazard assessment under complex geological conditions. It's very high susceptibility zone contains 52.33% of the landslides while covering only 13.07% of the area, which helps narrow down the targets for precise prevention and control.

(2) Elevation, Slope, Distance to rivers, and Rainfall are the main positive factors, whereas Land use type and Lithology act as important negative factors; together they reflect the combined influence of natural conditions and human activities.

(3) The models differ markedly in their classification behavior, so the choice of algorithm should depend on the characteristics of the regional data. Combining XGB with RF can improve the reliability of the susceptibility assessment.

These results provide a scientific basis for disaster prevention and control in Jinning District. Future work could integrate InSAR deformation monitoring and IoT data to establish a combined "assessment - monitoring - early warning" system, thereby raising the level of intelligent geological disaster management.

References

[1] Calvello, M., & Pecoraro, G. (2018). FraneItalia: A catalog of recent Italian landslides. Geoenvironmental Disasters, 5(1), 13. https://doi.org/10.1186/s40677-018-0105-5.

[2] Haque, U., Blum, P., da Silva, P., Andersen, P., Pilz, J., Chalov, S., Malet, J., Auflič, M., Andres, N., Poyiadji, E., Lamas, P., Zhang, W., Peshevski, I., & Pétursson, H. (2016). Fatal landslides in Europe. Landslides, 13(6), 1545–1554. https://doi.org/10.1007/s10346-016-0689-3.

[3] Huang, W., Ding, M., Li, Z., Zhuang, J., Yang, J., Li, X., Meng, L., Zhang, H., & Dong, Y. (2022). An efficient user-friendly integration tool for landslide susceptibility mapping based on support vector machines: SVM-LSM toolbox. Remote Sensing, 14(14), 3408. https://doi.org/10.3390/rs14143408.

[4] Zhang, J., Qian, J., Lu, Y., Li, X., & Song, Z. (2024). Study on landslide susceptibility based on multi-model coupling: A case study of Sichuan Province, China. Sustainability, 16(16), 6803. https://doi.org/10.3390/su16166803.

[5] Kim, J. C., Lee, S., Jung, H. S., & Lee, S. (2018). Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. Geocarto International, 33(9), 1000–1015. https://doi.org/10.1080/10106049.2017.1323964.

[6] Meng, S., Shi, Z., Li, G., Peng, M., Liu, L., Zheng, H., & Zhou, C. (2024). A novel deep learning framework for landslide susceptibility assessment using improved deep belief networks with the intelligent optimization algorithm. Computers and Geotechnics, 167, 106106. https://doi.org/10.1016/j.compgeo.2024.106106.

[7] Pourghasemi, H. R., Pradhan, B., & Gokceoglu, C. (2012). Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Natural Hazards, 63(3), 965–996. https://doi.org/10.1007/s11069-012-0217-2.

[8] Ercanoglu, M., Gokceoglu, C., & Van Asch, W. J. (2004). Landslide susceptibility zoning of North of Yenice (NW Turkey) by multivariate statistical techniques. Natural Hazards, 32(1), 1–23. https://doi.org/10.1023/B:NHAZ.0000026786.85589.4a.

[9] Tsangaratos, P., & Benardos, A. (2014). Estimating landslide susceptibility through an artificial neural network classifier. Natural Hazards, 74(3), 1489–1516. https://doi.org/10.1007/s11069-014-1259-z.

[10] Huang, F., Chen, J., Liu, W., Huang, J., Hong, H., & Chen, W. (2022). Regional rainfall-induced landslide hazard warning based on landslide susceptibility mapping and a critical rainfall

threshold. Geomorphology, 408, 108236. https://doi.org/10.1016/j.geomorph.2022.108236.

[11] Shu, Y., Zheng, G., & Yan, X. (2022). Application of multiple geographical units convolutional neural network based on neighborhood effects in urban waterlogging risk assessment in the city of Guangzhou, China. Physics and Chemistry of the Earth, 126, 103054. https://doi.org/10.1016/j.pce.2022.103054.

[12] Rohan, T., Shelef, E., Mirus, B., & Coleman, T. (2023). Prolonged influence of urbanization on landslide susceptibility. Landslides, 20(7), 1433–1447. https://doi.org/10.1007/s10346-023-02050-6.

[13] Xiong, H., Ma, C., Li, M., Tan, J., & Wang, Y. (2023). Landslide susceptibility prediction considering land use change and human activity: A case study under rapid urban expansion and afforestation in China. Science of The Total Environment, 866, 161430. https://doi.org/10.1016/j.scitotenv.2023.161430.

[14] Wang, X., Nie, W., Xie, W., & Zhang, Y. (2024). Incremental learning-random forest model-based landslide susceptibility analysis: A case of Ganzhou City, China. Earth Science Informatics, 17(2), 1645–1661. https://doi.org/10.1007/s12145-024-01267-1.

[15] Nandi, A., Mandal, A., Wilson, M., & Smith, D. (2016). Flood hazard mapping in Jamaica using principal component analysis and logistic regression. Environmental Earth Sciences, 75(6), 465. https://doi.org/10.1007/s12042-016-2262-3.