Debris Flow Early Warning Model Based on Support Vector Machine

Shihai Ma

School of River and Ocean Engineering, Chongqing Jiaotong University, Chongqing 400074, China

Email: 652442618@qq.com

Abstract. Debris flow is a significant geological hazard in mountainous regions, threatening human life and property. This study applies the Support Vector Machine (SVM) algorithm to predict debris flows in small watersheds, using Liangshan Yi Autonomous Prefecture as a case study. Eight key factors influencing debris flows were identified to construct a dataset, with each factor's weight calculated. Model performance, evaluated using the Receiver Operating Characteristic (ROC) curve, showed an AUC of 0.84, indicating strong interpretability and efficiency. The model provides valuable technical support for debris flow disaster prevention in southwestern mountainous regions.

Keywords: Machine Learning · Support Vector Machine · Debris Flow Early Warning

1 Introduction

Debris flows, common in mountainous regions covering two-thirds of China, are frequent and pose serious threats to life and property. Effective prediction and early warning are essential for disaster prevention, requiring ongoing methodological improvements[1].

Key determinants like topography, geological structure, soil properties, and rainfall intensity are crucial for accurate early warning systems, supporting disaster management through targeted responses and risk mitigation[2].

Recent advancements in remote sensing, GIS, and machine learning have improved debris flow susceptibility assessment. Researchers have developed diverse machine learning models to enhance prediction accuracy, using large-scale environmental data[3,4]. For instance, Liu et al. achieved an AUC of 0.84[5] with a random forest model for the Wenchuan earthquake region, and Wang et al. combined LDA with XGBoost, optimized via Grey Wolf Optimization, to enhance model performance[6]. Techniques like BP neural networks and improved cloud models further showcase machine learning's potential in refining predictive accuracy[7].

Support Vector Machine (SVM) is an effective tool for debris flow prediction, handling high-dimensional, non-linear data and small samples. This study applies SVM to

G. Zhao et al. (eds.), *Proceedings of the 2024 6th International Conference on Civil Architecture and Urban Engineering (ICCAUE 2024)*, Advances in Engineering Research 262,
https://doi.org/10.2991/978-94-6463-688-8_33

develop a predictive model for debris flow early warning in Liangshan Yi Autonomous Prefecture, a high-risk area with steep terrain, low vegetation, and heavy seasonal rainfall. Key features such as terrain slope, vegetation, and rainfall were selected based on historical analysis to create a region-specific warning system for improved disaster prevention.

2 Study Area Overview

The Liangshan Yi Autonomous Prefecture, located on the eastern edge of the Hengduan Mountains in southwestern Sichuan Province (Figure 1), features complex topography with steep mountains, deep valleys, and frequent geological activity due to notable fault development. Liangshan's monsoon-driven climate brings heavy rainfall during the rainy season (June to September), with concentrated downpours frequently triggering debris flows. Given these conditions, assessing debris flow susceptibility in Liangshan is crucial. Comprehensive analysis of geology, topography, and precipitation patterns provides essential data for predicting debris flows, forming a scientific basis for disaster prevention and emergency response. Liangshan thus serves as an ideal study area for developing early warning models, with practical insights for disaster management and policy formulation.

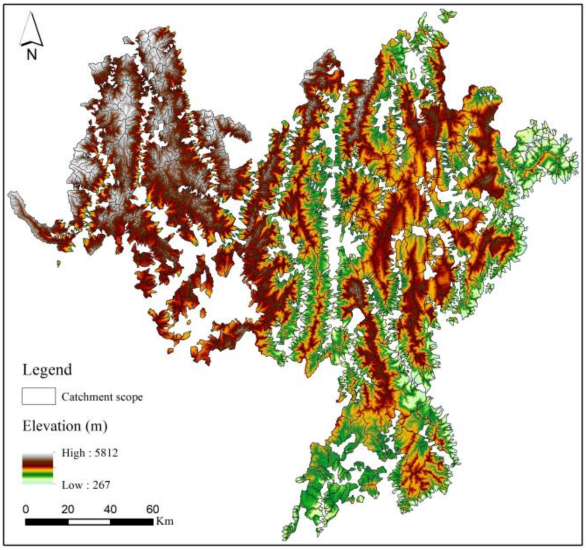

Fig. 1. Liangshan Yi Autonomous Prefecture study area

2.1 Debris Flow Dataset and Characteristic Factors

Debris flow prediction relies on datasets that reveal relationships between triggering factors and event occurrence. This study uses data from Liangshan Prefecture (2023-

2024), focusing on key factors such as topography, rainfall, geological conditions, and vegetation. Topographical factors include watershed area, slope, aspect, elevation difference, and NDVI, derived from an 8-meter DEM. Rainfall data,The common approach is to establish rainfall thresholds based on the relationship between intensity and duration, enabling early warning and prediction[8-11]. Geological factors, calculated as a composite indicator based on rock hardness and correction coefficients[12], reflect seismic and weathering effects on debris flow potential. Finally, vegetation cover, measured by NDVI, highlights areas with low vegetation where debris flows are more likely to occur.

3 The Principle of Support Vector Machines

Support Vector Machines (SVM) are a powerful supervised learning model used for classification and regression tasks[13]. SVM adapts to a limited set of training samples by applying a nonlinear transformation, mapping the input space to a high-dimensional feature space. In this space, SVM seeks the optimal hyperplane to separate data that is non-separable in the original space, making it effective for both linear and nonlinear classification problems, especially in high-dimensional spaces. The goal is to find a hyperplane that separates different classes, represented by the following equation:

$$f(x) = \omega^T \phi(x) + b \tag{1}$$

Here, ω represents the normal vector of the hyperplane, and b is the bias term. D represents the sample space, and $x_i \in R^n$ denotes the feature vector.

3.1 For Linearly Separable Cases

For linearly separable datasets, SVM requires the following constraints to be satisfied:

$$y_i(w \cdot x_i + b) \geq 1 \tag{2}$$

The goal of SVM is to maximize the margin $\frac{2}{\|w\|}$, which transforms the optimization problem into minimizing the following objective function:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{3}$$

$$\text{Subject to } y_i(w^T x_i + b) \geq 1, i = 1, 2, \ldots, n \tag{4}$$

While satisfying the constraints.

4 Data Preprocessing

Data preprocessing and feature selection are among the key steps for the successful application of machine learning[14]. Proper preprocessing can enhance learning performance, while feature selection can reduce data dimensionality and improve model interpretability. The input variables are standardized using the following formula:

$$Z = \frac{(X - \mu)}{\sigma} \tag{5}$$

Here, X represents the raw data value, while μ and σ denote the sample mean and standard deviation, respectively.

4.1 Feature Selection

This study uses the Pearson feature selection method, which calculates the linear correlation between features and the target variable to identify the most influential factors for classification or regression models. The results show that rainfall is the most significant factor contributing to debris flow occurrence, followed by slope, average rainfall intensity, and slope aspect. Factors such as watershed area contribute the least. Based on these contributions, factors are eliminated from low to high to obtain the optimal combination. Five key indicators are selected: rainfall (28.9%), slope (25.8%), average rainfall intensity (12.6%), slope aspect (5.6%), and normalized vegetation index (0.8%) (Figure 2).

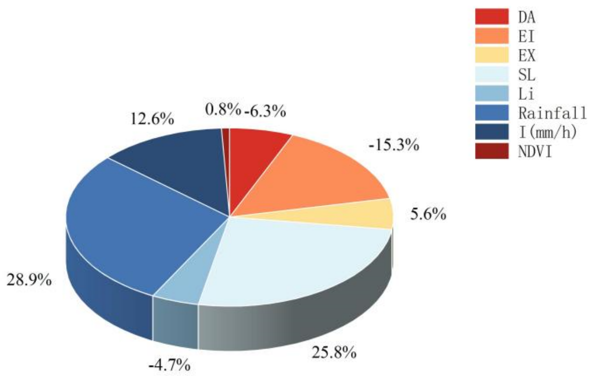

Fig. 2. Susceptibility Factors for Debris Flow

5 Model Performance Evaluation Metrics

This study evaluates the early warning capability of debris flow using the Support Vector Machine (SVM) model and quantifies the model's performance with the Receiver

Operating Characteristic (ROC) curve [15]. As shown in the figure, the ROC curve illustrates the relationship between the true positive rate and the false positive rate, reflecting the model's classification ability at various thresholds. The area under the curve (AUC) is 0.84, indicating shows that the model has good prediction performance and can effectively distinguish between positive (debris flow occurrence) and negative (no debris flow occurrence) categories(Figure 3).

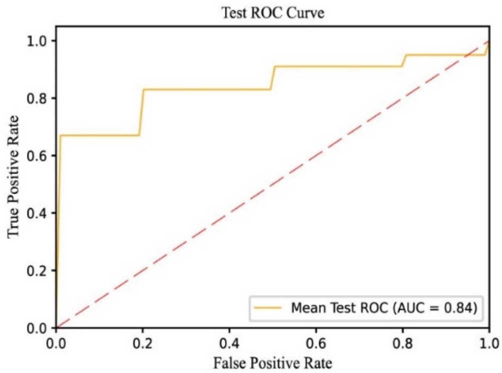

Fig. 3. The ROC curve of the model using the test dataset.

Table 1. Model Prediction Metrics

Model Evaluation Metrics	Value
F1 Score	0.84
Recall	0.87
Precision	0.81
Accuracy	0.83

Additionally, other performance metrics of the model indicate its strong predictive capability (Table 1). This analysis suggests that the SVM model is feasible for early warning of debris flow and provides a solid theoretical foundation for further research.

6 Conclusion

This study systematically assesses debris flow susceptibility in Liangshan Prefecture using an SVM model, highlighting key influencing factors. Based on 2023–2024 data, the conclusions are as follows:

1. Key Factor Identification: Precipitation, slope, average rainfall intensity, aspect, and vegetation coverage are identified as critical indicators of debris flow risk. Increased precipitation, in particular, significantly elevates this risk.
2. Model Performance: The optimized SVM model achieves high accuracy with an AUC of 0.821 and F1 score of 0.753, affirming SVM's potential for debris flow risk assessment.

To enhance model accuracy, future studies could explore additional machine learning models integrate long-term climate data, and utilize real-time hydrological data. These approaches would provide valuable insights for refining debris flow prediction and disaster management practices.

Reference

1. Cui, Peng. (2009). Progress in Landslide Prevention and Control in China. *Journal of Soil and Water Conservation*, 7(05): 7–13+31. doi:10.16843/j.sswc.2009.05.002.
2. Meng, Xingmin, Chen, Guan, Guo, Peng, Xiong, Muqi, & Wasowski, Janusz. (2013). Research Progress and Prospects of Landslide and Debris Flow Disasters in the Bailong River Basin. *Marine Geology and Quaternary Geology*, 04: 1–15.Zhang, Shuhao, & Wu, Guang. (2019). Debris Flow Susceptibility and Reliability Using Random Forest and GIS. *Earth Sciences*, 09: 3115–3134.
3. Zhang, Shuhao, & Wu, Guang. (2019). Debris Flow Susceptibility and Reliability Using Random Forest and GIS. *Earth Sciences*, 09: 3115–3134.
4. Dong, Aijia, & Wu, Chunxue. (2022). Evaluation of Debris Flow Susceptibility in Leibo County Based on GIS and FNN-SGD. *Software Guide*, 11: 58–68.
5. Liu, Yongyao, Di, Baofeng, Zhan, Yu, & Stamatopoulos, Constantine A. (2018). Debris Flow Susceptibility Assessment Based on Random Forest Model: A Case Study of the Wenchuan Earthquake Devastated Area. *Journal of Mountain Science*, 05: 765–773. doi:10.16089/j.cnki.1008-2786.000372.
6. Wang, Zhiyong, Li, Limin, Wen, Zongzhou, Shang, Yanfang, & Wang, Lianxia. (2023) Debris Flow Disaster Prediction Based on GWO-XGBoost. *Electronic Measurement Technology*, 03: 92–99. doi:10.19651/j.cnki.emt.2210581.
7. Wang, Xin, Zhao, Kang, Jiang, Yelin, & Zhu, Wen. (2022). Research on Dynamic Zoning Model of Debris Flow Susceptibility in the Nu River Basin Based on BP Neural Network. *Chemical Industry Minerals and Processing*, 01: 39–43+48. doi:10.16283/j.cnki.hgkwyjg.2022.01.008.
8. Papa, M. N., Medina, V., Ciervo, F., & Bateman, A. (2013). Derivation of critical rainfall thresholds for shallow landslides as a tool for debris flow early warning systems. *Hydrology and Earth System Sciences*, 17: 4095–4107. https://doi.org/10.5194/hess-17-4095-2013.
9. Guzzetti, F., Peruccacci, S., Rossi, M., & et al. (2008). The rainfall intensity–duration control of shallow landslides and debris flows: an update. *Landslides*, 5: 3–17. https://doi.org/10.1007/s10346-007-0112-1.
10. Giannecchini, R., Galanti, Y., Avanzi, G. D. A., & Barsanti, M. (2016). Probabilistic rainfall thresholds for triggering debris flows in a human-modified landscape. *Geomorphology*, 257: 94–107. https://doi.org/10.1016/j.geomorph.2015.12.012.
11. van Asch, T. W. J., Tang, C., Alkema, D., & et al. (2014). An integrated model to assess critical rainfall thresholds for run-out distances of debris flows. *Natural Hazards*, 70: 299–311. https://doi.org/10.1007/s11069-013-0810-z.
12. Yu, B., Chu, S., Lu, K., Han, L., & Xie, H. (2012). A study about the relationship between the frequency of debris flows and lithology. In *Landslides and Engineered Slopes: Protecting Society Through Improved Understanding* (pp. 757–761). Taylor & Francis Group, London.
13. Amarappa, S., & Sathyanarayana, S. V. (2014). Data classification using Support Vector Machine (SVM): a simplified approach. *International Journal of Electronics and Computer Science Engineering*, 3: 435–445.

14. Ahmad, T., & Aziz, M. N. (2019). Data preprocessing and feature selection for machine learning intrusion detection systems. *ICIC Express Letters*, 13(2): 93–101.
15. Wang, Yunsheng, Xie, Bingyan, Wan, Fanghao, Xiao, Qiming, & Dai, Liangying. (2007). Application of ROC Curve Analysis in Evaluating Invasive Species Distribution Models. *Biodiversity Science*, 04: 365–372.