

Research Article

Hybridization of Machine Learning Algorithms and an Empirical Regression Model for Predicting Debris-Flow-Endangered Areas

Xiang Wang,¹ Mi Tian ,^{2,3} Qiang Qin,⁴ and Jingwei Liang¹

¹Hunan Institute of Water Resources and Hydropower Research, 370 Shaoshan North Road, Changsha 410007, China

²School of Civil Engineering, Architecture and Environment, Hubei University of Technology, 28 Nanli Road, Wuhan 430068, China

³Innovation Demonstration Base of Ecological Environment Geotechnical and Ecological Restoration of Rivers and Lakes, Wuhan 430068, China

⁴Hunan Water Planning & Design Institute Co. Ltd., 158 Furong North Road, Changsha 410008, China

Correspondence should be addressed to Mi Tian; mitianhbut@hbut.edu.cn

Received 21 July 2023; Revised 17 August 2023; Accepted 10 October 2023; Published 6 November 2023

Academic Editor: Majid Ali

Copyright © 2023 Xiang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate delineation of debris-flow-endangered areas (e.g., the maximum runout distance) is a necessary prerequisite for the debris-flow risk assessment and countermeasures design. Recently, machine-learning models have been proved to be an effective tool in predicting debris-flow parameters. However, existing machine-learning models are generally developed based on a very limited number of observation data, which may result in the predictive model overfitting or underfitting. How to develop a robust model for accurate forecasting of debris-flow-endangered areas still remains a difficult task. This paper proposes a hybrid method for predicting debris-flow hazard zone by integrating machine-learning algorithms and an empirical regression model. The proposed method takes the calculated maximum runout distance obtained from the empirical model as supplementary inputs to increase the amount of training data to construct hybrid machine-learning models. Three commonly used machine-learning models (i.e., multivariate adaptive regression splines (MARS), random forest (RF), and support vector machine (SVM)) are developed based on the training datasets of a specific debris basin. Then, these three machine-learning models are combined with an empirical relationship developed using the same training datasets to generate corresponding hybrid models. Finally, the performance metrics (i.e., coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE)) of the proposed hybrid models are comprehensively investigated and compared with the single predictive model (i.e., MARS, RF, SVM, and the empirical model) under fivefold cross-validation. The proposed method is illustrated using 134 channelized debris-flow events in Sichuan province, China. Results show that compared with the three individual machine-learning models, hybridization of machine-learning algorithms and the empirical model results in R^2 , RMSE, and MAE improved by 70.5%, 32.9%, and 41.1%, respectively. In contrast to the empirical model, the R^2 , RMSE, and MAE value of the proposed hybrid models are improved by 29.6%, 22.3%, and 32.5%, respectively. The proposed hybrid models generally perform better than the single machine-learning and the empirical model, providing a promising tool for accurate forecasting of a debris-flow-endangered area.

1. Introduction

Debris flow is a common geological hazard in the mountainous areas, characterized by the rapid movements of saturated soil, rocks, and organic debris down steep mountain channels or slopes [1–6]. Debris flows can cause serious damages to the local residents, buildings, and infrastructures on the depositional area [7–13]. Therefore, it is crucial to delineate

accurate debris-flow-endangered areas in order to provide a practical guidance for the local authority in debris-flow-hazard assessment and control measure designs.

The maximum runout distance, L_f , on a depositional fan is one of the most important parameters for delineating the debris-flow-endangered areas [14–17]. Currently, statistical methods have been proved to be a simple and effective tool for predicting the maximum runout distance of debris flow

[18, 19]. The statistical methods generally relate the endangered area with geometric–morphological factors and the debris-flow volume through mathematical models based on the field investigation data of historical debris flows [20–23]. Both empirical regression analysis and machine-learning algorithms are widely employed to develop the mathematical models, such as univariate or multivariate nonlinear regression [16, 17], multivariate adaptive regression splines (MARS) [20], and support vector machine (SVM) [24–27]. The empirical models are generally developed by nonlinear regression analysis, where the underlying regression functions are preassumed by the experienced engineers or researchers. However, the assumed form of empirical models may not objectively capture the complex nonlinear relationship between the debris-flow runout distance and its influencing factors. Compared with the empirical models, machine-learning models have the powerful capacity in capturing the internal complex patterns from data and providing a rational forecasting as a guidance [28–30]. Machine-learning algorithms generally require sufficient data to guarantee their accuracy. Insufficient sample size of training datasets may affect the accuracy and generalization ability of the machine-learning models. However, the number of site-specific investigation data of a given debris basin is usually very limited. For example, Xu et al. [27] applied only 53 shallow landslides and 22 loess-bedrock landslides in the Heifangtai terrace to construct and validate the optimal machine-learning algorithms for runout distance prediction of shallow and loess-bedrock landslides, respectively. When a very limited number of data are applied, the machine-learning models may be overfitting or underfitting [28]. Thus, how to develop a robust model for accurate forecasting of debris-flow-endangered areas still remains a difficult task.

Different statistical models have their inherent mathematical theories and strengths. Currently, due to the ability to reasonably describe complex relationships, machine-learning models seem to have better accuracy than empirical models in debris-flow runout forecasting [20]. However, the empirical models derived from the historical debris flows still can provide useful information for the preliminary assessment of endangered areas. The empirical models show the main influencing factors of debris-flow runout from the perspective of expert experiences, thereby reducing data dimensions and avoiding overfitting. Therefore, empirical statistical models can be regarded as supplementary information for machine-learning models. In this case, the limited investigation data are expanded to increase the amount of training data to improve the robustness and accuracy of the predictive model. However, hybridizing the machine-learning algorithms and empirical models is scarcely seen in debris-flow runout prediction.

This paper proposes a hybrid method for predicting the endangered areas of debris flow by integrating machine-learning algorithms and an empirical regression model. Firstly, the collected investigation data is randomly divided into training and testing stages by k -fold cross-validation. Three commonly used machine-learning models (i.e., MARS, RF, and SVM, see the Appendix) and a nonlinear regression model (NLRM) are developed based on the

randomly selected training data, respectively. Then, the calculated maximum runout distance by the empirical model is considered as supplementary inputs and hybridized with the machine-learning models to generate a hybrid training model. Finally, the performance metrics (i.e., goodness-of-fit, root-mean-square error (RMSE), and mean absolute error (MAE)) of the hybrid models are comprehensively investigated and compared with the single machine-learning model and the empirical model under k -fold cross-validation. 134 datasets of channelized debris-flow events in Wenchuan earthquake zone are used to illustrate the applicability and reliability of the proposed approach.

The rest of this paper is organized as follows: Section 2 introduces the databases and data preparation. Section 3 elaborates the methodology. Section 4 presents the results and discussion, following by discussion on limitations of the paper, and Section 5 concludes the paper.

2. Databases and Data Preparation

2.1. Study Area. This study reanalyzed 134 datasets of channelized debris-flow events in Sichuan province, China [17]. The 134 debris flows occurred along the Yingxiu–Beichuan fault zone (e.g., Beichuan, Qingping, and Longchi area) between 2008 and 2012 [17]. The Yingxiu–Beichuan fault triggered the 8-magnitude Wenchuan earthquake and caused a number of landslides in Sichuan province. Loose landslide debris deposited on the slopes or channels were easily carried downstream to form debris flows by the torrential rain, resulting in serious damages to the local people on the depositional fan [12]. Previous studies in the study area show that the maximum runout distance of debris flow on a fan is mainly affected by the catchment internal relief (H), and the debris-flow volume (V_D) [17]. Therefore, these two parameters are taken as input variables to predict the endangered area of debris flow in the study area. More details can be referred to Zhou et al. [17]. Figure 1 shows the histograms of the maximum runout distance (L_f), and input variables (i.e., H and V_D) as well as their mean values and standard deviations. It can be seen that the distribution of these three parameters approximates a lognormal distribution.

2.2. K -Fold Cross-Validation. To construct and validate a predictive statistical model, the 134 debris-flow datasets in the study area are randomly divided into training data for model development and testing data for model validation. Different random combinations of training datasets and testing datasets may lead to fluctuations in model performance evaluation. To avoid the bias in data selection, k -fold cross-validation approach is employed for randomly selection of the training and testing data [31, 32]. k -fold cross-validation is a popular technique used in machine-learning and model evaluation to assess the performance and generalization ability of a model. It randomly divides the original dataset into k subsets D_1, D_2, \dots, D_k with an approximately equal size. The process is then repeated k times. One of the k subsets is used as testing data, and the other $k-1$ subsets are considered as training data. The model is trained k times with $k-1$ folds as the training data and the remaining fold as the validation

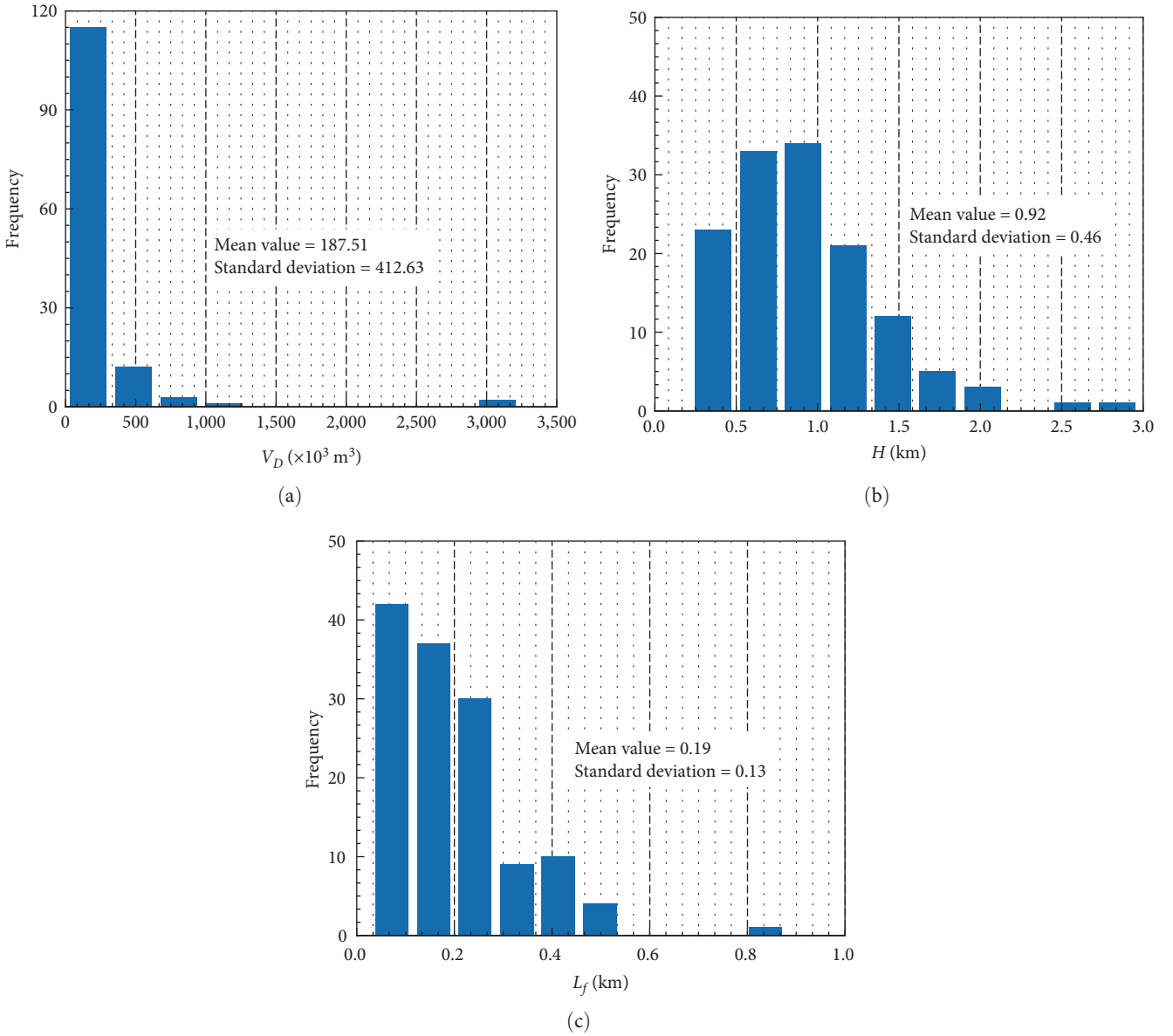


FIGURE 1: Histograms of L_f and input variables (i.e., H and V_D): (a) V_D , (b) H , and (c) L_f .

data. Then, the model’s performance is evaluated on each validation dataset. After the k iterations, the performance metrics obtained from each validation dataset are averaged to provide a single evaluation metric for the model’s performance. k -fold cross-validation ensures a predictive model is evaluated on different subsets of the data, providing a more comprehensive assessment of how well the model generalizes to unseen data.

Previous studies show that a value of $k = 5$ (or 10) is very common in the application of machine-learning algorithms [31]. By taking $k = 5$, the datasets are randomly divided into five equal parts, which are denoted as CR1, CR2, CR3, CR4, and CR5, respectively. 80% and 20% of the total datasets are training and testing data, respectively (as shown in Figure 2). Among 134 datasets of debris flow in Wenchuan area, 107 sets of data are selected for training the models and 27 sets of

data are selected to evaluate the performance of predictive models. Table 1 shows the ranges of H , V_D , and L_f for training and testing data under fivefold cross-validation. It can be seen that the data ranges of H , V_D , and L_f for the fivefold cross-validation are almost the same. The training data for the five different splits of CR1, CR2, CR3, CR4, and CR5 are then used to construct the machine-learning models and a NLRM. The testing data for the five different splits are used for validation.

3. Methodology

Various machine-learning models (e.g., C&RT, CHAID, boosting tree, MARS, RF, and SVM) can characterize the complex nonlinear relationship between input and output parameters [28–30]. In this paper, three commonly used

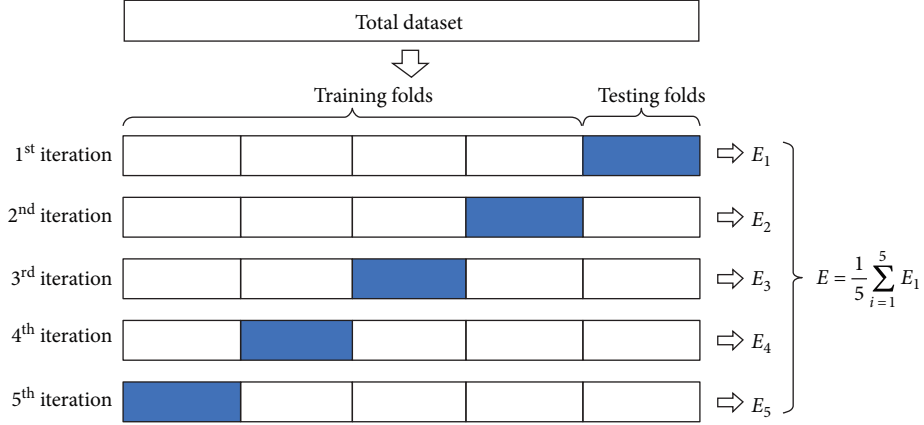


FIGURE 2: Schematic diagram of fivefold cross-validation.

TABLE 1: Ranges of V_D , H , and L_f at training and testing stages.

| Parameters | V_D ($\times 10^3$ m ³) | H (km) | L_f (km) |
|------------------|--|-----------|------------|
| Total data | 1.2–3238.4 | 0.22–2.98 | 0.03–0.88 |
| Cross-validation | Training | | |
| CR1 | 2.7–3238.4 | 0.22–2.98 | 0.03–0.88 |
| CR2 | 1.2–3238.4 | 0.22–2.48 | 0.03–0.88 |
| CR3 | 1.2–3238.4 | 0.22–2.98 | 0.03–0.88 |
| CR4 | 1.2–3238.4 | 0.26–2.98 | 0.03–0.88 |
| CR5 | 1.2–3100.0 | 0.22–2.98 | 0.038–0.51 |
| Cross-validation | Testing | | |
| CR1 | 1.2–711.0 | 0.34–2.48 | 0.04–0.39 |
| CR2 | 3.0–561.0 | 0.36–2.98 | 0.042–0.51 |
| CR3 | 2.9–1034.0 | 0.26–1.60 | 0.038–0.50 |
| CR4 | 4.5–3100.0 | 0.22–2.02 | 0.05–0.49 |
| CR5 | 2.7–3238.4 | 0.26–1.98 | 0.03–0.88 |

machine-learning models (i.e., MARS, RF, and SVM) are employed. The procedure for developing the hybrid prediction model of the maximum runout distance of debris flow is shown in Figure 3. First, the three machine-learning models and a NLRM are developed independently. The hybrid models (e.g., MARS–NLRM, RF–NLRM, and SVM–NLRM) are then generated by combining the NLRM with the machine-learning models. The proposed method is briefly presented as follows.

3.1. Multi-Nonlinear Regression Empirical Model. According to the previous studies, the maximum runout distance of debris flow is usually related to geometric–morphological factors and the debris-flow volume through statistical regression analysis. The existing empirical models are typically exponential formulas [20], which can be expressed as follows:

$$L_f = aV_D^b H^c, \quad (1)$$

where a , b , and c are the unknown parameters of the empirical model. By applying multi-nonlinear regression method,

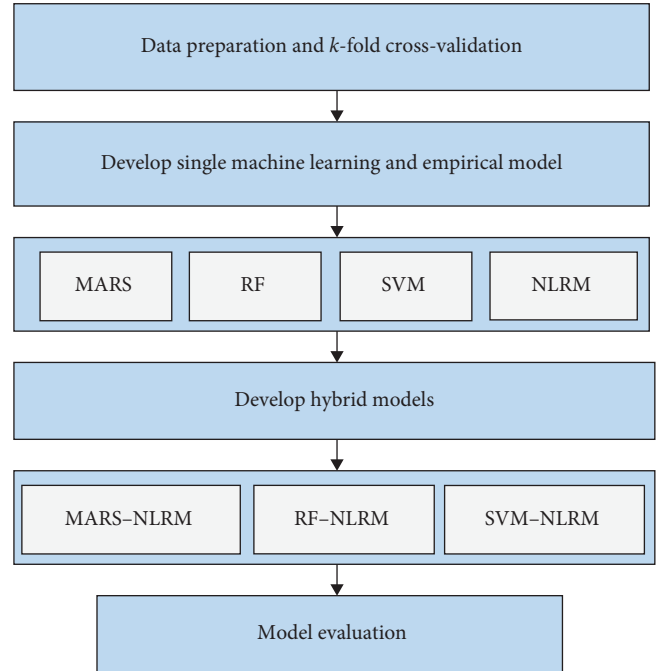


FIGURE 3: Flowchart of hybrid model development.

the model parameters in Equation (1) can be obtained for the five different splits of CR1, CR2, CR3, CR4, and CR5.

3.2. Machine-Learning Models. The core idea of machine-learning algorithms is to apply data-driven learning to construct models that can generalize well to the data, make accurate predictions, and improve their performance over time as they are exposed to more data. In this paper, three common machine-learning models are used to construct the training models, namely, MARS, RF, and SVM.

3.2.1. MARS Model. MARS is a nonparametric regression analysis method that can be used for modeling multidimensional nonlinear problems [33]. This method does not assume a specific functional relationship between the input variables and the output variable. Instead, it adaptively selects nodes to partition the training dataset into independent segments with

different gradients [34, 35]. Each segment of MARS is called a basis function, and the endpoints of each segment are called nodes. The basis functions are generated by MARS through a stepwise search, and the node positions are selected by the adaptive regression algorithm. By estimating the contribution of the basis functions, it allows determining the additive and interaction effects of the predictive variables. The MARS algorithm consists of two steps: forward selection and backward pruning. The forward selection process involves partitioning and fitting the sample data using spline functions to obtain new basis functions and a fitted model. As a basis function is added to the model space, interactions between basis functions that are already in the model are also considered. Basis functions are added until the model reaches some maximum specified number of terms leading to a purposely overfit model. The backward pruning process involves removing basis functions with smaller contributions to the model while maintaining the model's accuracy. The basis functions maintained in the final optimal model are selected from the set of all candidate basis functions used in the forward selection step. Model subsets are compared using the less computationally expensive method of generalized cross-validation (GCV). Finally, the optimal model is selected as the regression model to avoid overfitting.

Considering n independent or input variables $X = (x_1, x_2, \dots, x_n)$. The dependent or output variable, y can be estimated from a predefined function $g(X)$ with a model error ε , which can be given by

$$y = g(\mathbf{X}) + \varepsilon. \quad (2)$$

The predefined function $g(X)$ can be approximated by linear combination of basis functions and their interactions. The MARS model of $g(X)$ can be expressed as follows:

$$g(\mathbf{X}) = \beta_0 + \sum_{j=1}^M \beta_j \omega_j(\mathbf{X}), \quad (3)$$

where β_0 is a constant; M is the number of basis functions; $\omega_j(\mathbf{X})$ is the j -th basis function; β_j is the coefficient of the j -th basis function. The coefficients $\beta_0, \beta_1, \dots, \beta_M$ are determined by the least square method. The basis function $\omega_j(\mathbf{X})$ can be characterized by a piecewise linear function, which is written as follows:

$$\max(0, x - t) = \begin{cases} x - t & x \geq t \\ 0 & x < t \end{cases}, \quad (4)$$

where t is the value of knot.

The construction of MARS model is an adaptive process where basis functions and knots are all entirely "driven" from the training data. To obtain the MARS model in Equation (3), the forward building procedure is performed on the training data. Basis functions that produce the largest decrease in the training model error are added until the predefined

maximum number of terms is reached. This procedure can easily lead to an overfitting MARS model. Subsequently, the backward procedure prunes extraneous variables and basis functions with the least contributions based on the GCV method. The GCV index is an indicator that penalizes the complexity of large numbers of basis functions in the MARS model in order to reduce overfitting problems. With the N observations of training data, GCV for a model can be obtained by:

$$\text{GCV} = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - g(x_i)]^2}{\left[1 - \frac{M + c_0 \times (M-1)/2}{N}\right]^2}, \quad (5)$$

where c_0 is the penalizing parameter; $g(x_i)$ is the predicted values using the MARS model in Equation (3). The penalizing parameter c_0 is set as a default value of three according to Friedman [33]. Based on the GCV index, the MARS model with the minimum value of GCV is selected as the optimal MARS model.

In addition, it should be noted that the maximum number of basis functions needs to be predefined. The optimal preassumed M can be determined by comparing the evaluation metrics of MARS model (e.g., RMSE) with different preset values. The one with the lowest RMSE is considered as the optimal predefined maximum number of basis functions for the final MARS model.

3.2.2. RF Model. RF is a typical ensemble learning method based on classification and regression trees (CART) [36]. RF creates a collection of decision trees and combines their predictions through averaging to make the final prediction. Decision trees can be divided into classification trees and regression trees. Because the primary objective of this research is to predict the endangered area of debris flow, only the regression tree is discussed in this section. RF regression uses the random sampling with bootstrap resampling to extract multiple samples for the original sample, model the regression tree for each bootstrap sample and average the predictions of multiple decision trees to make the final forecasting [37, 38].

Considering $X = (x_1, x_2, \dots, x_n)$ is an n -dimension input vector that forms a forest. RF consists of a set of K trees $\{y_1(X), y_2(X), \dots, y_K(X)\}$. The ensemble produces K outputs corresponding to each tree y_k ($k = 1, 2, \dots, K$). The modeling procedure of RF is as follows: draw a bootstrap sample from the original data set. For each bootstrap sample, a total of two-thirds of the sample of the new training sample is utilized for deriving the regression function, and the remaining one-third constitutes the out-of-bag (OOB) sample. Each time, a regression tree is constructed using a randomized training sample drawn from the original data set. The OOB sample is utilized to validate accuracy. After the predictions of the k regression trees are collected, a regression model sequence $\{y_1(X), y_2(X), \dots, y_K(X)\}$ is obtained. Then, the final prediction is obtained by calculating an average of all tree predictions, which is given by:

$$f(x) = \frac{1}{K} \sum_{k=1}^K y_k(x), \quad (6)$$

where $f(x)$ represents the combined regression model, y_k represents an individual decision tree regression model, and K is the number of regression trees. The number of regression trees at RF model structure is a critical hyperparameter, which can be determined by Grid search method. Grid search is a hyperparameter tuning technique used in machine-learning to systematically search through a predefined hyperparameter space for the optimal combination that maximizes or minimizes a chosen evaluation metric. It involves exhaustively exploring various combinations of hyperparameters to identify the one that yields the best performance. In this work, the optimal value for the number of regression trees is selected using the Grid search method.

3.2.3. SVM Model. SVM is a powerful supervised machine-learning algorithm used for both classification and regression [39]. Based on the statistical learning theory and structural risk minimization principle, SVM optimizes a tradeoff between the complexity and learning ability of the model to obtain best generalization ability according to limited sample information. Its basic idea is to map input vectors into a high-dimension feature space via a kernel function and construct an optimal separating hyperplane that best separates or fits the data and maximizes the margin between the different classes or regression targets [40, 41]. The hyperplane effectively separates the data points and is supported by a small subset of critical data points, which is called support vectors. It can effectively handle high-dimensional data and nonlinearly separable problems by using the kernel trick to transform the data into a higher dimensional space.

Given a set of training data, (x_i, y_i) , $i = 1, 2, \dots, l$, $x_i \in R^n$, $y_i \in R$, a support vector machine for regression is to obtain a function in the following form:

$$f(x) = \omega \cdot \phi(x) + \beta, \quad (7)$$

where ω and β are the parameters to be determined from the training set; $\phi(x)$ represents a high-dimensional feature space that is nonlinearly mapped from the low dimensional space x . ω and β can be determined by minimizing the regularized risk function, which are defined as follows:

$$\frac{1}{2} \|\omega\|^2 + \frac{C}{l} \sum_{i=1}^l h_\varepsilon(y_i, f(x_i)), \quad (8)$$

where $\|\omega\|^2$ is the regularized term; C is called as penalty factor. $\frac{1}{l} \sum_{i=1}^l h_\varepsilon(y_i, f(x_i))$ is the empirical error measured by the ε -insensitive loss function, which is given by:

$$h_\varepsilon(y_i, f(x_i)) = \begin{cases} 0, & y_i - f(x_i) \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon, & \text{otherwise} \end{cases}. \quad (9)$$

In order to obtain ω and β , kernel functions are usually used to make computations performed directly in input space, without calculating the $\phi(x)$. At present, four basic kernels have been widely used, namely linear kernel, polynomial kernel, the radial-basis function and sigmoid kernel. In this paper, the Gaussian radial basis function (RBF) kernel function is used to construct the SVM model. More details of the RBF kernel function can be found in [40]. In SVM regression, two critical hyperparameters, i.e., gamma parameter γ and penalty factor C need to be preset before the learning process. Grid search method is employed to select the optimal values of gamma parameter and penalty factor.

3.3. Hybrid Models. To improve the predictive accuracy of the debris-flow runout distance, hybrid models are proposed to integrate the robustness of machine-learning models with the limitation of the empirical regression models. The maximum runout distance of debris flow obtained from the NLRM using the training data is taken as auxiliary input for the three machine-learning algorithms to construct the training models. Taken the MARS–NLRM hybrid model as an example. First, the 107 sets of training data for CR1 split in Table 1 is used to develop the empirical relationship between the maximum runout distance and its influencing factors (i.e., H and V_D) using Equation (1). Then, the maximum runout distance can be calculated by using the developed empirical model for CR1 and the training data. The calculated maximum runout distance and the corresponding catchment H and V_D is applied as supplementary input data for the MARS. To this end, a total of 214 training datasets are generated to construct the MARS–NLRM data-driven model. Finally, the performance of established hybrid MARS–NLRM model is evaluated by using the testing data for CR1 split. Similar method is utilized for CR2, CR3, CR4, and CR5, respectively. After completing the five iterations, the performance metrics of the hybrid MARS–NLRM model derived from each testing data are averaged to provide a final evaluation for the MARS–NLRM model's performance. As for other two hybrid models (i.e., RF–NLRM and SVM–NLRM), the same procedures are adopted to obtain their model evaluation indexes.

3.4. Performance Metrics of Different Models. It is essential to evaluate the accuracy and reliability of a predictive model. Hence, the predictive accuracy and robustness of the proposed hybrid models in the paper are assessed by three mathematical metrics, i.e., coefficient of determination (R^2), RMSE, and MAE. R^2 is a statistical metric used to assess the goodness-of-fit of a predictive model. It reflects the proportion of the variance between the predicted dependent variable and the measured value, which is given by:

$$R^2 = 1 - \frac{\sum_{i=1}^N [L_{f,i}^m - L_{f,i}^p]^2}{\sum_{i=1}^N [L_{f,i}^m - \bar{L}_f^m]^2}, \quad (10)$$

where N is the number of the data; $L_{f,i}^m$ represents the actual measured maximum runout distance; $L_{f,i}^p$ is the predicted maximum runout distance; L_f^m is the mean value of the actual measured maximum runout distance. The value of R^2 ranges from 0 to 1, where a higher value indicates a better fit of the model to the observed data.

RMSE quantifies the average discrepancy between the actual measured values and the predicted values produced by the predictive model, which is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N [L_{f,i}^m - L_{f,i}^p]^2}{N}}. \quad (11)$$

A lower RMSE value indicates a better predictive performance of the model, with smaller deviations between predictions and actual measurements. RMSE not only considers the relationship between predicted and actual values but also quantifies the magnitude of the errors.

MAE computes the absolute errors between the actual values and predicted values, which are calculated as follows:

$$\text{MAE} = \frac{\sum_i^N |L_{f,i}^m - L_{f,i}^p|}{N}. \quad (12)$$

The MAE can characterize the accuracy of the predictive model by considering the absolute size of the errors, which is not affected by the RMSE.

These three statistical metrics are used to evaluate the predictive models under k -fold cross-validation, where R^2 reflects how well the model fits the data, RMSE gauges the accuracy of the model, and MAE quantifies the absolute errors. A good predictive model should ideally have a high R^2 , a low RMSE, and a small MAE, demonstrating its ability to effectively model the data and make accurate predictions for unseen datasets.

4. Results and Discussion

4.1. Hyperparameters for Machine-Learning Models. The empirical models for CR1, CR2, CR3, CR4, and CR5 are developed based on the fivefold cross-validation by using Equation (1). The model parameters of multi-nonlinear empirical relationships for the five different splits are summarized in Table 2. It is found that the values of model parameters, i.e., a , b , and c , are very close to each other under the fivefold cross-validation. Then, the maximum runout distance is calculated by using the developed empirical models for the fivefold cross-validation, and used as supplementary inputs to establish the hybrid models. Totally, four single model (i.e., NLRM, MARS, RF, and SVM) and three hybrid models (i.e., MARS–NLRM, RF–NLRM, and SVM–NLRM) are generated for CR1, CR2, CR3, CR4, and CR5, respectively.

In machine learning, hyperparameter is the parameter that is needed to be predefined before the modeling procedure. For MARS, RF, and SVM, there are several

TABLE 2: Model parameters of multi-nonlinear regression empirical models.

| Cross-validation | a | b | c |
|------------------|-------|------|------|
| CR1 | 0.053 | 0.39 | 0.28 |
| CR2 | 0.056 | 0.33 | 0.27 |
| CR3 | 0.039 | 0.40 | 0.35 |
| CR4 | 0.054 | 0.44 | 0.27 |
| CR5 | 0.052 | 0.40 | 0.29 |

hyperparameters that have significant effects on the predictive accuracy. Reasonable selection of the optimal hyperparameters, is a necessary prerequisite for machine learning. To tune hyperparameters in these three machine-learning algorithms, Grid search is used to systematically search through a predefined hyperparameter space for the optimal combination that minimizes RMSE.

As for the MARS model, the predetermined maximum number of basis functions for CR1, CR2, CR3, CR4, and CR5 is determined with the minimum RMSE. The range of basis functions is adjusted as 60. The optimal MARS models adopt 30, 40, 40, 30, and 28 basis functions of linear spline functions for CR1, CR2, CR3, CR4, and CR5, respectively. As for the MARS–NLRM modeling procedure, the optimal maximum number of basis functions for CR1, CR2, CR3, CR4, and CR5 are 40, 30, 30, 30, and 40, respectively. As for the RF and RF–NLRM model, a maximum value of regression trees is defined as 500, i.e., $N_R = 500$ trees. Through trial and error, the tree numbers higher than 300 for the five different splits have no significant effect on the model performance. As for the SVM modeling procedure, the ranges of C and g are set as (0, 1,500) and (0, 10). The optimal C values of SVM model for the five different splits are 1,024, 1,024, 8, 256, and 512, respectively. The gamma parameter g equals to 0.177, 0.125, 0.177, 0.022, and 0.707, respectively. For the SVM–NLRM model, the optimal C values for CR1, CR2, CR3, CR4, and CR5 are 724, 512, 512, 0.707, and 0.5. Meanwhile, the optimal g is found as 0.354, 0.25, 0.125, 2.828, and 8, respectively. These optimal hyperparameters are applied in the machine-learning algorithms to generate the single training model and hybrid models.

4.2. Comparisons of Different Predictive Models. Table 3 shows the performance metrics of all the predictive models under the fivefold cross-validation. It is shown that except for MARS, NLRM, RF, and SVM have comparable predictive performance. Under the fivefold cross-validation, the MARS model shows the worst performance with the lowest average of R^2 , largest mean of RMSE, and maximum mean of MAE. Meanwhile, it can be seen that the hybrid models generally show better performance than individual models in terms of R^2 , RMSE, and MAE. From Table 3, the ranges of R^2 values of NLRM, MARS, RF, and SVM for the testing data are 0.292–0.840, 0.115–0.747, 0.252–0.722, and 0.233–0.619, respectively. However, the R^2 ranges of MARS–NLRM, RF–NLRM, and SVM–NLRM are 0.542–0.861, 0.543–0.763, and 0.540–0.769, respectively. Compared with NLRM, the R^2 values of MARS–NLRM, RF–NLRM,

TABLE 3: Performance metrics of all the predictive models under fivefold cross-validation.

| Model | | Training data | | | Testing data | | |
|-------|-----------|---------------|-------|-------|--------------|-------|-------|
| | | R^2 | RMSE | MAE | R^2 | RMSE | MAE |
| CR1 | NLRM | 0.804 | 0.086 | 0.057 | 0.575 | 0.057 | 0.047 |
| | MARS | 0.783 | 0.062 | 0.045 | 0.457 | 0.070 | 0.055 |
| | RF | 0.916 | 0.044 | 0.031 | 0.556 | 0.060 | 0.050 |
| | SVM | 0.669 | 0.079 | 0.051 | 0.573 | 0.059 | 0.049 |
| | MARS–NLRM | 0.818 | 0.052 | 0.034 | 0.715 | 0.048 | 0.036 |
| | RF–NLRM | 0.871 | 0.044 | 0.029 | 0.717 | 0.044 | 0.030 |
| | SVM–NLRM | 0.774 | 0.058 | 0.031 | 0.732 | 0.042 | 0.031 |
| CR2 | NLRM | 0.770 | 0.083 | 0.056 | 0.638 | 0.076 | 0.054 |
| | MARS | 0.696 | 0.071 | 0.052 | 0.747 | 0.061 | 0.048 |
| | RF | 0.913 | 0.044 | 0.032 | 0.722 | 0.066 | 0.049 |
| | SVM | 0.661 | 0.077 | 0.050 | 0.619 | 0.076 | 0.056 |
| | MARS–NLRM | 0.749 | 0.057 | 0.036 | 0.688 | 0.056 | 0.035 |
| | RF–NLRM | 0.859 | 0.043 | 0.028 | 0.727 | 0.051 | 0.035 |
| | SVM–NLRM | 0.746 | 0.057 | 0.036 | 0.702 | 0.054 | 0.033 |
| CR3 | NLRM | 0.768 | 0.078 | 0.053 | 0.292 | 0.104 | 0.069 |
| | MARS | 0.772 | 0.062 | 0.049 | 0.115 | 0.117 | 0.084 |
| | RF | 0.920 | 0.040 | 0.029 | 0.252 | 0.101 | 0.073 |
| | SVM | 0.633 | 0.080 | 0.056 | 0.233 | 0.110 | 0.072 |
| | MARS–NLRM | 0.850 | 0.050 | 0.033 | 0.542 | 0.074 | 0.045 |
| | RF–NLRM | 0.899 | 0.041 | 0.028 | 0.543 | 0.074 | 0.047 |
| | SVM–NLRM | 0.840 | 0.052 | 0.031 | 0.540 | 0.078 | 0.041 |
| CR4 | NLRM | 0.801 | 0.081 | 0.052 | 0.553 | 0.092 | 0.065 |
| | MARS | 0.670 | 0.073 | 0.055 | 0.476 | 0.103 | 0.070 |
| | RF | 0.913 | 0.044 | 0.032 | 0.590 | 0.082 | 0.061 |
| | SVM | 0.621 | 0.080 | 0.055 | 0.371 | 0.127 | 0.081 |
| | MARS–NLRM | 0.785 | 0.051 | 0.033 | 0.719 | 0.068 | 0.044 |
| | RF–NLRM | 0.857 | 0.042 | 0.027 | 0.744 | 0.061 | 0.041 |
| | SVM–NLRM | 0.738 | 0.056 | 0.032 | 0.715 | 0.065 | 0.039 |
| CR5 | NLRM | 0.660 | 0.065 | 0.056 | 0.840 | 0.076 | 0.051 |
| | MARS | 0.583 | 0.072 | 0.054 | 0.517 | 0.257 | 0.150 |
| | RF | 0.886 | 0.041 | 0.030 | 0.597 | 0.115 | 0.069 |
| | SVM | 0.570 | 0.075 | 0.049 | 0.494 | 0.125 | 0.072 |
| | MARS–NLRM | 0.713 | 0.055 | 0.034 | 0.861 | 0.060 | 0.035 |
| | RF–NLRM | 0.831 | 0.043 | 0.029 | 0.763 | 0.081 | 0.043 |
| | SVM–NLRM | 0.700 | 0.057 | 0.032 | 0.769 | 0.078 | 0.038 |

and SVM–NLRM are increased by 30%, 30%, and 29%, respectively. In contrast to MARS, RF, and SVM, R^2 values of MARS–NLRM, RF–NLRM, and SVM–NLRM are improved by 108%, 40%, and 64%, respectively. It is obvious that the performance metrics of hybrid models are significantly improved, indicating the proposed hybrid method has better goodness of data fitting. Figure 4 shows the R^2 curves of training and testing stages under fivefold cross-validation. It is clear that except for RF, the R^2 of MARS–NLRM, RF–NLRM, and SVM–NLRM at the training stages are generally larger than NLRM, MARS, and SVM. At the testing stages, the three hybrid models' R^2 are obviously larger than the other four stand-alone models' R^2 values. Although RF achieves a good training model, the prediction model has a large deviation for the testing data.

The main reason is that for a small sampling size, RF algorithm is prone to overfitting, which will lead to the degradation of model performance. However, after incorporating the supplementary datasets from NLRM, the RF–NLRM has better data-fitting ability for both training and testing stages, showing better predictive accuracy. Similar results are also observed for the MARS–NLRM and SVM–NLRM. This indicates that the hybrid models' performance is better than individual models in model evaluation.

Figure 5 plots the RMSE curves of training and testing stages under fivefold cross-validation. It is clear that at the training stages, the hybrid models have smaller RMSE values than stand-alone models except for RF. However, under fivefold cross-validation, the RMSE values of MARS–NLRM, RF–NLRM, and SVM–NLRM are all smaller than that of a

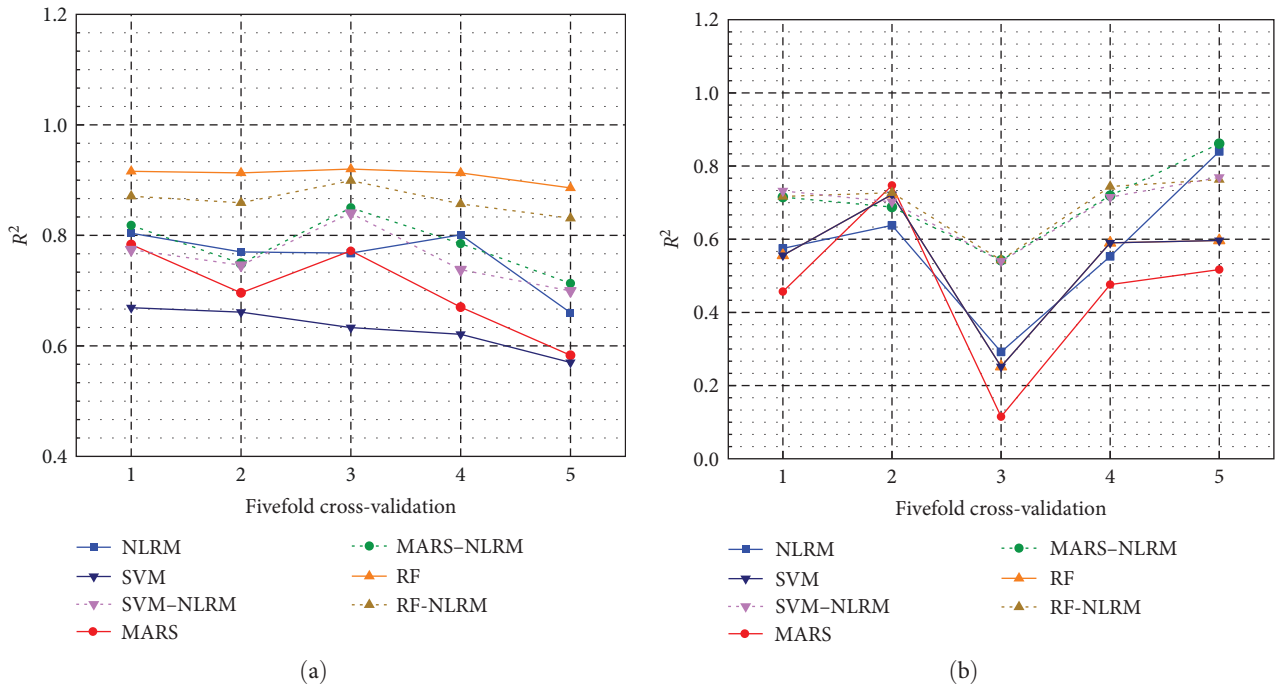


FIGURE 4: R^2 curves of training and testing stages under fivefold cross-validation: (a) training data and (b) testing data.

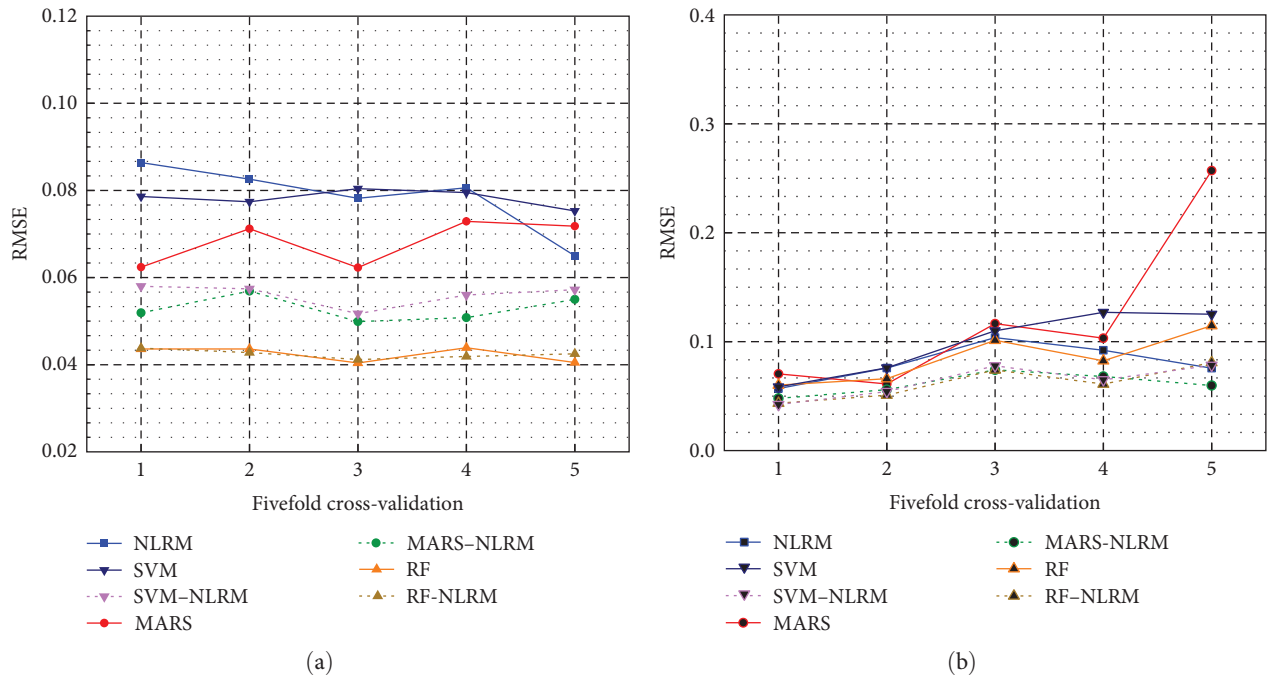


FIGURE 5: $RMSE$ curves of training and testing stages under fivefold cross-validation: (a) training data and (b) testing data.

single machine-learning model and NLRM at testing stages. From Table 3, the ranges of $RMSE$ values for NLRM, MARS, RF, and SVM at testing stages are 0.057–0.104, 0.061–0.257, 0.060–0.115, and 0.059–0.127, respectively, while the $RMSE$ for three hybrid models are 0.048–0.074, 0.044–0.074, and 0.042–0.078, respectively. Compared with NLRM, the $RMSE$ values of hybrid models decrease about 24%, 22%, and 21%,

respectively. In contrast to MARS, RF, and SVM, $RMSE$ values for MARS–NLRM, RF–NLRM, and SVM–NLRM are improved by 37.5%, 26.6%, and 34.4%, respectively. Obviously, hybridization of the empirical model and machine-learning algorithms can significantly reduce the prediction deviation of any individual model. The performance improvement of individual machine learning is higher than that of the

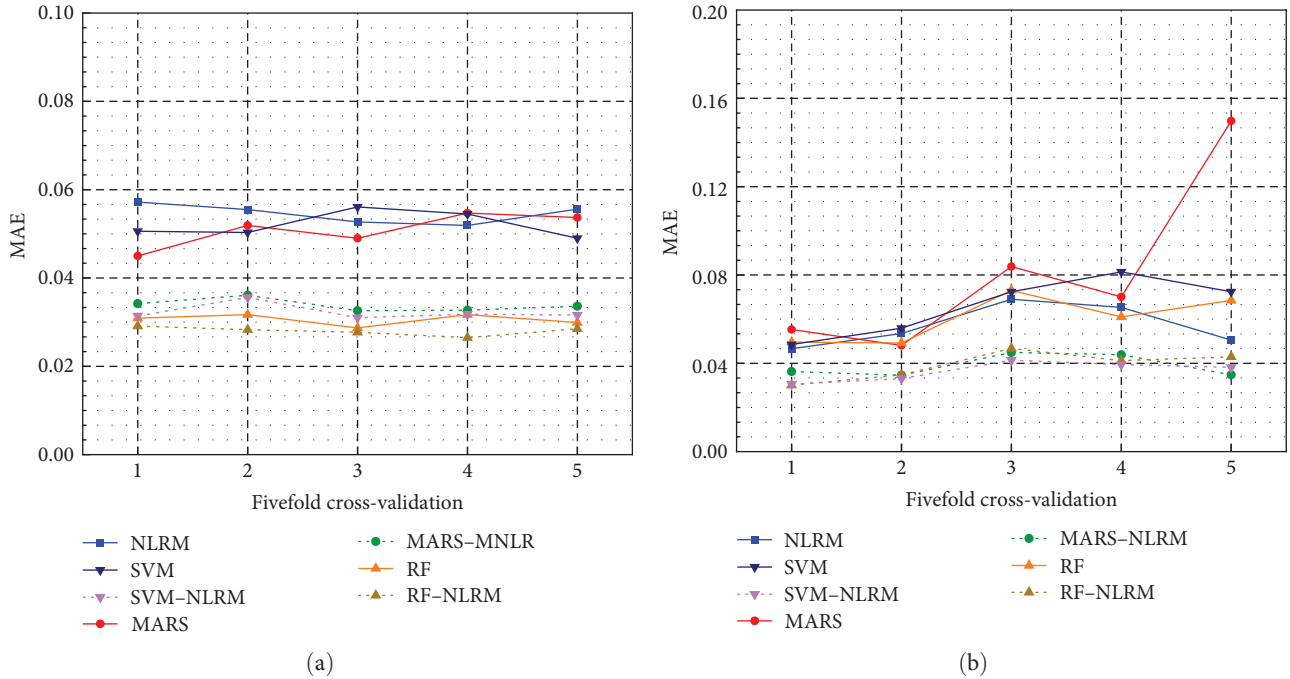


FIGURE 6: MAE curves of training and testing stages under fivefold cross-validation: (a) training data and (b) testing data.

TABLE 4: Comparisons of performance metrics of the predictive models.

| Model | Training data | | | Testing data | | |
|-------------------------|---------------|------|-------|--------------|-------|-------|
| | R^2 | RMSE | MAE | R^2 | RMSE | MAE |
| Single predictive model | NLRM | 0.76 | 0.079 | 0.58 | 0.081 | 0.057 |
| | MARS | 0.70 | 0.068 | 0.46 | 0.122 | 0.081 |
| | RF | 0.91 | 0.042 | 0.54 | 0.085 | 0.060 |
| | SVM | 0.63 | 0.078 | 0.46 | 0.099 | 0.066 |
| Hybrid predictive model | MARS-NLRM | 0.78 | 0.053 | 0.71 | 0.061 | 0.039 |
| | RF-NLRM | 0.86 | 0.042 | 0.70 | 0.062 | 0.039 |
| | SVM-NLRM | 0.76 | 0.056 | 0.69 | 0.063 | 0.037 |

empirical model. Totally, the hybrid predictive models have better performance and smaller deviations between predictions and actual measurements compared with an individual model.

Figure 6 plots the MAE curves of training and testing stages under fivefold cross-validation. At the training stages, the MAE values for the three hybrid models are lower than that of the single model except for RF. However, the MAE values for the hybrid models are significantly smaller than that of all the individual predictive model at testing stages, implying that hybridization of the machine-learning models and the empirical model has better predictive accuracy. For example, the MAE values of MARS for the testing dataset under fivefold cross-validation are 0.055, 0.048, 0.084, 0.070, and 0.150, respectively. While the MAE values of MARS-NLRM in the testing dataset for the five splits are 0.036, 0.035, 0.045, 0.044, and 0.035, respectively, which is obviously lower than that of MARS and NLRM (i.e., 0.047, 0.054, 0.069, 0.065, and 0.051). This shows that the integration of MARS with NLRM can greatly

reduce the absolute errors between the actual values and predicted values, and improve the predictive accuracy. Similar results can also be observed for the RF-NLRM and SVM-NLRM. It can be found that hybridizing empirical statistical models and machine-learning algorithms can expand the amount of training data to improve the robustness and accuracy of the predictive model.

The average values of R^2 , RMSE, and MAE under the fivefold cross-validation for all the predictive models are summarized in Table 4. As shown in Table 4, the mean values of R^2 for the testing dataset of MARS-NLRM, RF-NLRM, and SVM-NLRM are 0.71, 0.70, and 0.69, respectively. They all outperform the NLRM (i.e., 0.58), MARS (i.e., 0.46), RF (i.e., 0.54), and SVM (i.e., 0.46). R^2 is improved by an average of 70.5% compared to the three machine-learning algorithms. Compared with the NLRM, the total improvement of R^2 average values for the hybrid models is about 29.6%.

It can also be seen that the mean values of RMSE for the testing dataset of MARS-NLRM, RF-NLRM, and SVM-NLRM are 0.061, 0.062, and 0.063, respectively, which are lower than

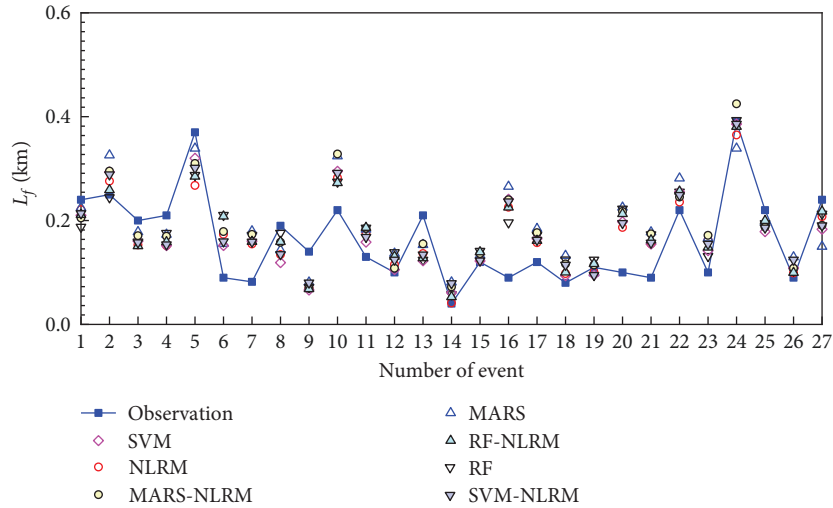


FIGURE 7: Predicted runout distance values for NLRM, MARS, RF, SVM, MARS–NLRM, RF–NLRM, and SVM–NLRM using testing data versus measured values ($k=1$).

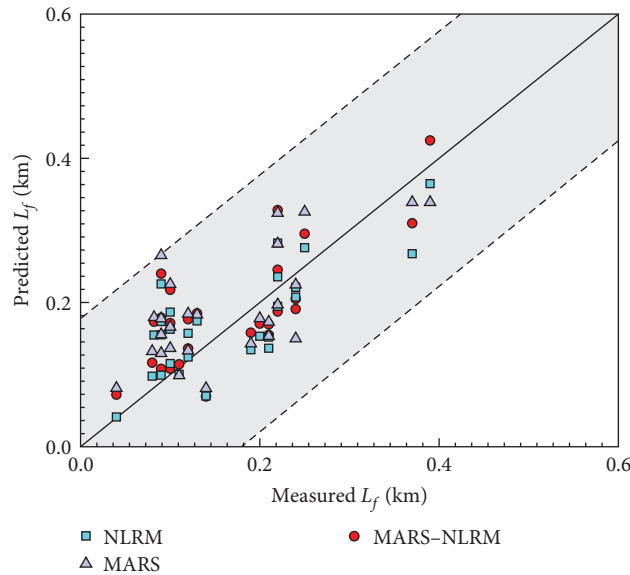


FIGURE 8: Comparisons of predictive performance of MARS–NLRM with single model for testing dataset ($k=1$).

that of NLRM (i.e., 0.081), MARS (i.e., 0.122), RF (i.e., 0.085), and SVM (i.e., 0.099). Compared with NLRM and the three machine-learning algorithms, the average improvement of RMSE values for the three hybrid models approximates 22.3% and 32.9%, respectively. Similar results are also observed in MAE. In contrast to the NLRM, the average MAE value of MARS–NLRM, RF–NLRM, and SVM–NLRM decreases about 32.5%. Compared with the three machine-learning algorithms, the average MAE value of corresponding hybrid models is reduced by 41.1%. Furthermore, the three hybrid models generally exhibit comparable predictive performance with similar indexes of R^2 , RMSE, and MAE. On the whole, the three hybrid models have higher prediction accuracy and lower errors. This implies that the proposed method can generate more accurate

and reliable predictions over the single machine-learning algorithm and the empirical relationship.

4.3. *Model Evaluations of Different Models at $k=1$.* The results of $k=1$ are selected as a representative to show the performance of the hybrid models. Figure 7 shows the predicted runout distance values for NLRM, MARS, RF, SVM, MARS–NLRM, RF–NLRM, and SVM–NLRM using testing data versus measured values. From Figure 7, the estimated values of the maximum runout distance using the single model and hybrid models are all close to the measured values. To further illustrate the performance of hybrid models, Figures 8–9 shows the comparisons of predictive performance of hybrid models (i.e., MARS–NLRM, RF–NLRM,

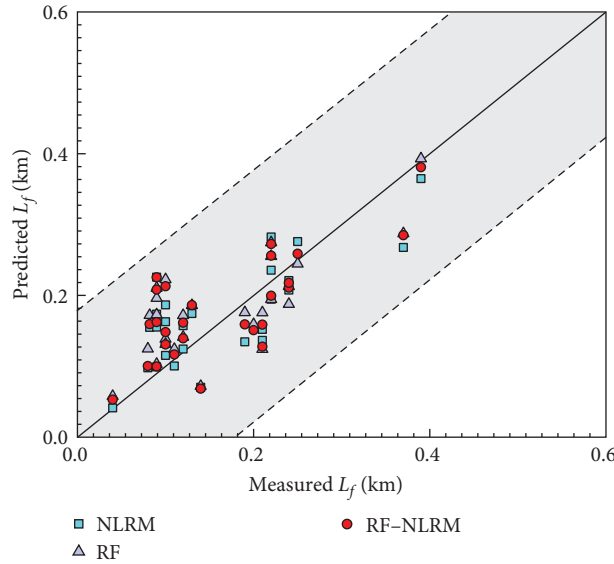


FIGURE 9: Comparisons of predictive performance of RF-NLRM with single model for testing dataset ($k = 1$).

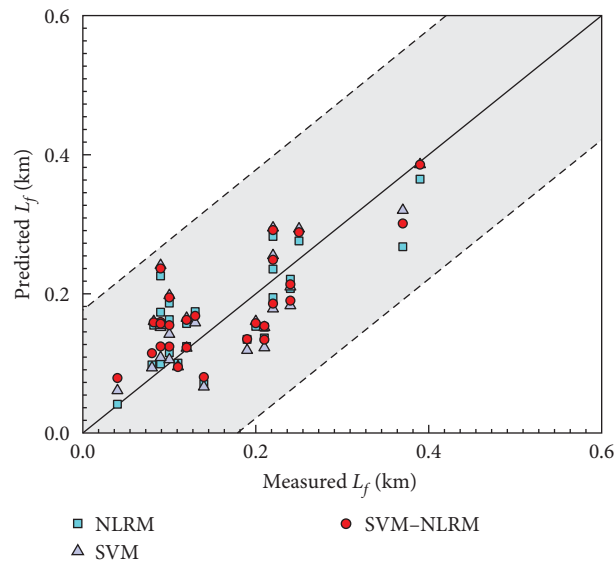


FIGURE 10: Comparisons of predictive performance of SVM-NLRM with single model for testing dataset ($k = 1$).

and SVM-NLRM) with single model ($k = 1$) for testing dataset. As shown in Figure 8, the gray area represents the 95% confidence interval for predictions of the maximum runout distance. The blue square, purple triangle, and red circle in Figure 8 represent the predicted L_f derived from the NLRM, MARS, and MARS-NLRM, respectively. It is clear that almost all the predicted data points fall into the 95% confidence interval. The predicted L_f by using MARS-NLRM is closer to the 1 : 1 line compared with NLRM and MARS. As for the performance of RF-NLRM, the predicted values of L_f

are also closer to the actual values compared with NLRM and RF, as shown in Figure 10. Similar results can also be found in Figure 9. The predicted values derived from SVM-NLRM are generally closer to the actual values than NLRM and SVM. Especially, when the actual measured L_f equals to 0.39 km, the predicted values from NLRM, RF-NLRM, and SVM-NLRM are 0.36, 0.38, and 0.39 km, respectively. It is clear that the RF-NLRM and SVM-NLRM model can provide more accurate predictions of the maximum runout distance of extreme debris-flow events.

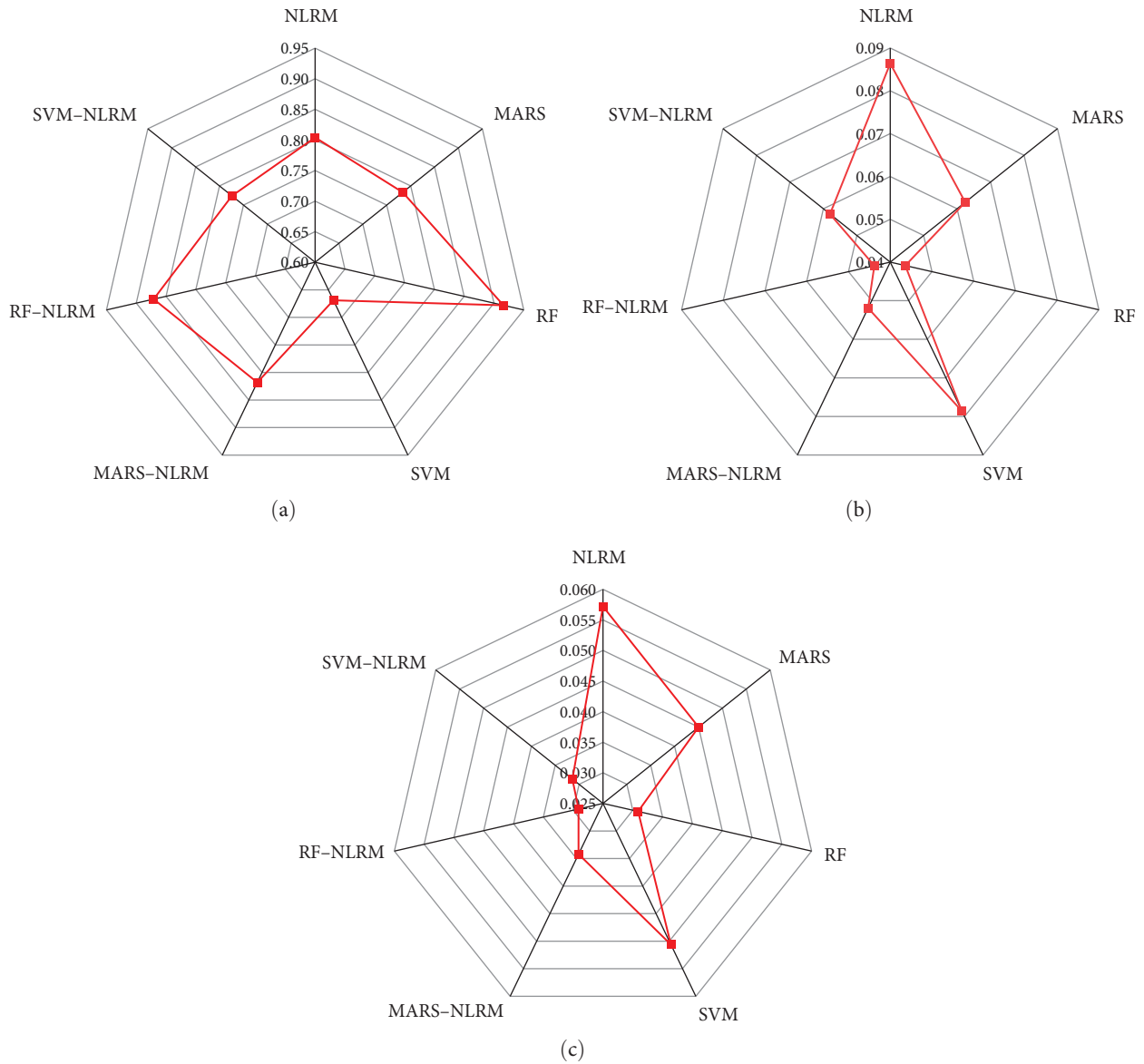


FIGURE 11: Performance metrics of four single predictive model and three hybrid models for training dataset ($k = 1$): (a) R^2 , (b) RMSE, and (c) MAE.

Figure 11 displays as radar diagrams the results for the R^2 , RMSE, and MAE statistical measures of accuracy for the four single predictive model and three hybrid models for training dataset under the first cross-validation. Figure 11 identifies that the MARS-NLRM, RF-NLRM, and SVM-NLRM models perform more accurately than the individual model in the calculation of the maximum runout distance of debris flow. Figure 12 plots the performance metrics of four single predictive model and three hybrid models for testing dataset. It is apparent that the three hybrid models have larger R^2 and smaller deviations

between predictions and actual measurements (i.e., lower RMSE and MAE) than the other models. Evaluation of the performance of the predictive models shows that MARS-NLRM, RF-NLRM, and SVM-NLRM outperform the empirical model and single machine-learning algorithm both for the training data and the validation data in terms of prediction accuracy. It is clear that combining the empirical model and machine-learning algorithms to predict the debris-flow runout zone can potentially overcome the overfitting or underfitting of machine-learning models due to the limited amount of sampling data.

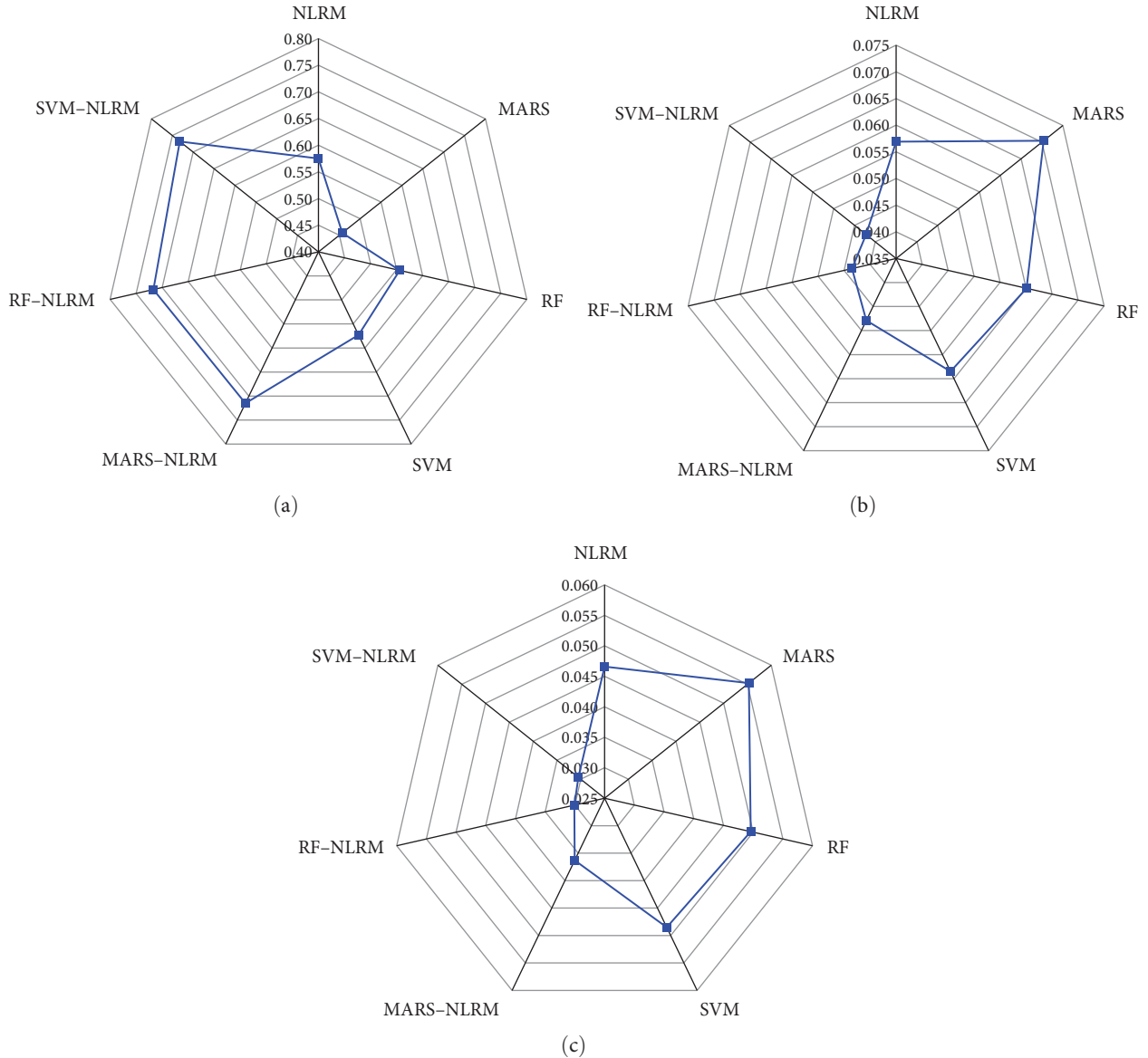


FIGURE 12: Performance metrics of four single predictive model and three hybrid models for testing dataset ($k = 1$): (a) R^2 , (b) RMSE, and (c) MAE.

5. Summary and Conclusions

Accurate delineation of debris-flow-endangered areas (e.g., the maximum runout distance) is a necessary prerequisite for the debris-flow risk assessment and countermeasures design. This paper proposes a hybrid method by integrating the machine-learning models and an empirical regression model to predict the maximum runout distance of debris flow. The proposed method takes the calculated maximum runout distance by the empirical model as supplementary inputs to increase the amount of training data of machine-learning models. The predictive performance of the proposed hybrid models is comprehensively evaluated by three statistical accuracy metrics (i.e., R^2 , RMSE, and MAE) and compared with the single predictive model (i.e., MARS, RF, SVM, and NLRM) under fivefold cross-validation. The proposed method is illustrated by using

134 datasets of channelized debris-flow events in Sichuan province, China. The following conclusions are drawn from the results and analysis:

- (1) For the individual predictive model, the MARS model shows the worst performance compared with the other single models (i.e., NLRM, RF, and SVM). The NLRM, RF, and SVM models have comparable predictive accuracy in estimating the maximum runout distance of debris flow in Wenchuan earthquake area.
- (2) For the study area, all the proposed hybrid MARS-NLRM, RF-NLRM, and SVM-NLRM models provide more accurate predicted values of the maximum runout distance than NLRM, MARS, RF, and SVM under fivefold cross-validation. After hybridizing

empirical statistical models and machine-learning algorithms, the amount of training data can be expanded to overcome the overfitting or underfitting of machine-learning models. Evaluation of the performance metrics of the predictive models shows that MARS–NLRM, RF–NLRM, and SVM–NLRM obviously outperform the empirical model and single machine-learning algorithm both for the training data and the validation data.

- (3) Compared with the three individual machine-learning model (i.e., MARS, RF, and SVM), hybridization of machine-learning algorithms and the empirical model results in R^2 , RMSE, and MAE improved by 70.5%, 32.9%, and 41.1%, respectively. In contrast to the empirical model, R^2 , RMSE and MAE values of the proposed hybrid models are improved by 29.6%, 22.3%, and 32.5%, respectively. The proposed hybrid method can obviously improve the predictive accuracy and robustness, providing a promising tool for predicting the debris-flow-endangered area and enhancing the model generalization.

Abbreviations

| | |
|------------|--|
| GCV: | Generalized cross-validation |
| MAE: | Mean absolute error |
| MARS: | Multivariate adaptive regression splines |
| MARS–NLRM: | Hybridization of MARS model and a nonlinear regression model |
| NLRM: | Nonlinear regression model |
| R^2 : | Coefficient of determination |
| RBF: | Radial basis function |
| RF: | Random forest |
| RF–NLRM: | Hybridization of RF model and a nonlinear regression model |
| RMSE: | Root-mean-square error |
| SVM: | Support vector machine |
| SVM–NLRM: | Hybridization of SVM model and a nonlinear regression model. |

Data Availability

All data in this study are available from the corresponding author upon reasonable request.

Additional Points

Limitations. In this paper, two crucial parameters (i.e., H and V_D) were considered as the input variables to develop the predictive models. H was derived from the digital elevation model (DEM). V_D was indirectly estimated by the field-observed thickness and depositional area. The quality and resolution of remote sensing images can significantly influence the accuracy of input variables, affecting the performance of the developed hybrid models. Moreover, various other factors (e.g., debris-flow velocity and fan topography) have great effects on the debris-flow runout. These factors are not considered in this paper. If more field data of these

factors are available, they can also be incorporated into the proposed method to reconstruct the hybrid training models to improve the predictive accuracy of the endangered area. These limitations do not affect the proposed method but shared by all the statistical methods. In this paper, only three commonly used machine-learning models (i.e., MARS, RF, and SVM) are employed to generate the hybrid models to predict the debris-flow runout. However, various machine-learning algorithms (e.g., C&RT, CHAID, and boosting tree) can also characterize the complex nonlinear relationship between input and output parameters. Other machine learning algorithms can also be applied in this study to investigate the performance of the hybrid models. This does not influence the feasibility of the proposed method in this paper. In addition, the developed hybrid models for predicting the maximum runout distance are suitable for the inputs within the parameter ranges in this study (as listed in Table 1). The performance of the proposed method largely relies on the quality of available data. If the input variables are beyond the parameter ranges in this study, the proposed hybrid models may exhibit poor performance. In this case, the proposed method should be used carefully.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Project No. 52009037), the Natural Science Foundation of Hubei Province of China (Project No. 2020CFB291), Hunan Water Science and Technology Project (Project No. XSKJ2021000-09), and the Wuhan Knowledge Innovation Special Project (Project No. 2022020801020268).

References

- [1] D. N. Swanston and F. J. Swanson, "Timber harvesting, mass erosion and steepland forest geomorphology in the Pacific Northwest," in *Geomorphology and Engineering*, D. R. Coates, Ed., pp. 199–221, Dowden Hutchinson and Ross Inc, Stroudsburg, PA, 1976.
- [2] O. Hungr, G. C. Morgan, and R. Kellerhals, "Quantitative analysis of debris torrent hazard for design of remedial measures," *Canadian Geotechnical Journal*, vol. 21, no. 4, pp. 663–677, 1984.
- [3] J.-E. Costa, "Physical geomorphology of debris flows," in *Developments and Applications of Geomorphology*, J. E. Costa and P. J. Fleischer, Eds., pp. 268–317, Springer, Berlin, Germany, 1984.
- [4] R. M. Iverson, "The physics of debris flows," *Reviews of Geophysics*, vol. 35, no. 3, pp. 245–296, 1997.
- [5] M. Jakob, "A size classification for debris flows," *Engineering Geology*, vol. 79, no. 3-4, pp. 151–161, 2005.
- [6] B.-Q. Luna, "Dynamic numerical run-out modeling for quantitative landslide risk assessment, Thesis of University of Twente," *ITC*, vol. 206, pp. 1–237, 2012.
- [7] M. Jakob and O. Hungr, *Debris Flow Hazards and Related Phenomena*, pp. 1–202, Praxis Publishing, Chichester, U.K, 2005.

- [8] M. Hürlimann, R. Copons, and J. Altimir, "Detailed debris flow hazard assessment in Andorra: a multidisciplinary approach," *Geomorphology*, vol. 78, no. 3-4, pp. 359-372, 2006.
- [9] R.-M. Iverson, M. Logan, R.-G. LaHusen, and M. Berti, "The perfect debris flow? Aggregated results from 28 large-scale experiments," *Journal of Geophysical Research*, vol. 115, no. F3, Article ID F03005, 2010.
- [10] R. Archetti and A. Lamberti, "Assessment of risk due to debris flow events," *Natural Hazards Review*, vol. 4, no. 3, pp. 115-125, 2003.
- [11] M. Berti and A. Simoni, "Prediction of debris flow inundation areas using empirical mobility relationships," *Geomorphology*, vol. 90, no. 1-2, pp. 144-161, 2007.
- [12] C. Tang, J. Zhu, M. Chang, J. Ding, and X. Qi, "An empirical-statistical model for predicting debris-flow runout zones in the Wenchuan earthquake area," *Quaternary International*, vol. 250, pp. 63-73, 2012.
- [13] L. M. Stancanelli and R. E. Musumeci, "Geometrical characterization of sediment deposits at the confluence of mountain streams," *Water*, vol. 10, no. 4, Article ID 401, 2018.
- [14] F. C. Dai, C. F. Lee, and Y. Y. Ngai, "Landslide risk assessment and management: an overview," *Engineering Geology*, vol. 64, no. 1, pp. 65-87, 2002.
- [15] V. D'Agostino, M. Cesca, and L. Marchi, "Field and laboratory investigations of runout distances of debris flows in the Dolomites (Eastern Italian Alps)," *Geomorphology*, vol. 115, no. 3-4, pp. 294-304, 2010.
- [16] S. Zhang, L.-M. Zhang, H.-X. Chen, Q. Yuan, and H. Pan, "Changes in runout distances of debris flows over time in the Wenchuan earthquake zone," *Journal of Mountain Science*, vol. 10, no. 2, pp. 281-292, 2013.
- [17] W. Zhou, J. Fang, C. Tang, and G. Yang, "Empirical relationships for the estimation of debris flow runout distances on depositional fans in the Wenchuan earthquake zone," *Journal of Hydrology*, vol. 577, Article ID 123932, 2019.
- [18] M. Hürlimann, D. Rickenmann, V. Medina, and A. Bateman, "Evaluation of approaches to calculate debris-flow parameters for hazard assessment," *Engineering Geology*, vol. 102, no. 3-4, pp. 152-163, 2008.
- [19] S. McDougall, "Canadian Geotechnical Colloquium: landslide runout analysis-current practice and challenges," *Canadian Geotechnical Journal*, vol. 54, no. 5, pp. 605-620, 2017, 2017.
- [20] M. Tian, L. Li, and Z. Xiong, "A data-driven method for predicting debris-flow runout zones by integrating multivariate adaptive regression splines and Akaike information criterion," *Bulletin of Engineering Geology and the Environment*, vol. 81, no. 6, Article ID 222, 2022.
- [21] A. B. Prochaska, P. M. Santi, J. D. Higgins, and S. H. Cannon, "Debris-flow runout predictions based on the average channel slope (ACS)," *Engineering Geology*, vol. 98, no. 1-2, pp. 29-40, 2008.
- [22] D. Rickenmann, "Empirical relationships for debris flows," *Natural Hazards*, vol. 19, no. 1, pp. 47-77, 1999.
- [23] H. Ikeya, "A method of designation for area in danger of debris flow," in *Erosion and Sediment Transport in Pacific Rim Steeplands*, T. R. H. Davies and A. J. Pearce, Eds., vol. 132, pp. 576-587, Christchurch, New Zealand, No., 1981.
- [24] A. S. Ansar, S. Sudha, S. Vinayamoorthi, M. M. Menachery, and S. Francis, "Landslide classification and prediction of debris flow using machine learning models," *IETE Journal of Research*, 2023.
- [25] W. Tang, H.-T. Ding, N.-S. Chen et al., "Artificial neural network-based prediction of glacial debris flows in the Parlung Zangbo Basin, Southeastern Tibetan Plateau, China," *Journal of Mountain Science*, vol. 18, no. 1, pp. 51-67, 2021.
- [26] D.-H. Lee, E. Cheon, H.-H. Lim, S.-K. Choi, Y.-T. Kim, and S.-R. Lee, "An artificial neural network model to predict debris-flow volumes caused by extreme rainfall in the central region of South Korea," *Engineering Geology*, vol. 281, Article ID 105979, 2021.
- [27] Q. Xu, H. Li, Y. He, F. Liu, and D. Peng, "Comparison of data-driven models of loess landslide runout distance estimation," *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 2, pp. 1281-1294, 2019.
- [28] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, 2006.
- [29] W. Zhang, X. Gu, L. Tang, Y. Yin, D. Liu, and Y. Zhang, "Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: comprehensive review and future challenge," *Gondwana Research*, vol. 109, no. 4, pp. 1-17, 2022.
- [30] W. Zhang and K.-K. Phoon, "Editorial for advances and applications of deep learning and soft computing in geotechnical underground engineering," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 14, no. 3, pp. 671-673, 2022.
- [31] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *International Joint Conference on Artificial Intelligence*, vol. 14, no. 2, pp. 1137-1145, 1995.
- [32] J. D. Rodriguez, A. Perez, and J. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 569-575, 2010.
- [33] J.-H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1-67, 1991.
- [34] A. T. C. Goh, W. Zhang, Y. Zhang, Y. Xiao, and Y. Xiang, "Determination of earth pressure balance tunnel-related maximum surface settlement: a multivariate adaptive regression splines approach," *Bulletin of Engineering Geology and the Environment*, vol. 77, no. 2, pp. 489-500, 2018.
- [35] W. Zhang, R. Zhang, W. Wang, F. Zhang, and A. T. C. Goh, "A multivariate adaptive regression splines model for determining horizontal wall deflection envelope for braced excavations in clays," *Tunnelling and Underground Space Technology*, vol. 84, pp. 461-471, 2019.
- [36] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [37] X. Wang, X. Gao, Y. Zhang et al., "Land-cover classification of coastal wetlands using the RF algorithm for worldview-2 and landsat 8 images," *Remote Sensing*, vol. 11, no. 16, Article ID 1927, 2019.
- [38] H. S. Barjoui, H. Ghorbani, N. Mohamadian et al., "Prediction performance advantages of deep machine learning algorithms for two-phase flow rates through wellhead chokes," *Journal of Petroleum Exploration and Production Technology*, vol. 11, no. 3, pp. 1233-1261, 2021.
- [39] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.
- [40] H. Ghorbani, D. A. Wood, A. Choubineh et al., "Prediction of oil flow rate through an orifice flow meter: artificial intelligence alternatives compared," *Petroleum*, vol. 6, no. 4, pp. 404-414, 2020.
- [41] F. Gu, J. Chen, X. Sun, Y. Li, Y. Zhang, and Q. Wang, "Comparison of machine learning and traditional statistical methods in debris flow susceptibility assessment: a case study of Changping district, Beijing," *Water*, vol. 15, no. 4, Article ID 705, 2023.