An Improved LSTM Time-Series Approach for Forecasting Debris Flow Events

*Hongtao Zhang, Qingguo Zhou**

School of Information Science and Engineering, Lanzhou University, China
13601156281@163.com, zhouqg@lzu.edu.cn

Abstract

Debris flows occur abruptly and possess strong destructive power, making them one of the most hazardous geomorphological processes capable of inflicting heavy losses in a short period of time. In view of the difficulty in obtaining timely and accurate predictions using existing warning techniques, this study introduces a novel time-series forecasting strategy for debris-flow hazards, built upon an enhanced LSTM model. Field surveys were first conducted to assemble the initial dataset, after which the data were carefully filtered and standardized. Principal Component Analysis was applied to reorganize and condense the numerous environmental and geological variables, allowing eight dominant factors related to debris-flow development to be extracted. Based on these refined inputs, an LSTM-driven prediction architecture was developed. The model was trained and validated using the prepared dataset, and its performance was benchmarked against several traditional forecasting methods. Experimental outcomes reveal that the proposed model maintains low prediction errors—below 0.06 for training, 0.12 for testing, and 0.19 for validation. These results highlight the method's superior predictive capability and robustness compared with conventional approaches. The framework presented here has the potential to significantly enhance debris-flow risk assessment and offers practical support for emergency management and early-warning operations.

Keywords: Debris flow, PCA, LSTM, Deep learning, Disaster prediction

1 Introduction

Debris flows are natural disasters that commonly occur in mountainous regions, triggered by heavy rainfall, floods, or other forms of water erosion that destabilize slopes and mobilize large volumes of soil and rock debris [1]. Due to their high fluidity and considerable destructive potential, debris flows can significantly affect both the surrounding environment and human activities, and in severe cases, may lead to substantial casualties and property damage [2].

The Longnan region of Gansu Province in China is located in the southern part of Gansu, at the junction of the Qinghai Tibet Plateau and the Loess Plateau. It belongs to the erosion erosion erosion tectonic mountain range of the Western Qinling Mountains, and the terrain is complex and variable. Here, there are overlapping mountains, dense vegetation, and soft soil, with relatively more rainfall. The rainfall period is concentrated, making it extremely prone to debris flow disasters. In addition, most cities and villages in the region are located in valley areas, and under the induction of factors such as rainfall, vibration, and human destruction, debris flow disasters are prone to cause significant damage to people and property, with great harm.

Accurately forecasting the timing and spatial extent of debris flow events and assessing their associated risks can assist relevant authorities in planning preventive measures, issuing timely warnings, and implementing effective control strategies, thereby playing a crucial role in disaster mitigation and reduction [3]. For decades, researchers have conducted a large amount of research work, and overall, they have summarized and formed models and methods for evaluating the risk of debris flows [4], such as qualitative analysis [5], statistical model analysis [6], and physical dynamics deterministic models [7]. In recent years, with the development of deep learning technology, deep learning model-based methods have gradually replaced traditional methods and become the mainstream of research due to their higher prediction accuracy and better generalization ability.

However, because there are many factors affecting debris flow disasters, it is difficult to obtain high-precision data, and a single model cannot make high-precision prediction [8]. For this reason, this paper first uses PCA algorithm to analyze the key influencing factors of flood disasters, and then uses the improved LSTM (Long Short Term Memory) algorithm to build a multi factor debris flow disaster time series prediction model.

The chapter arrangement of this paper is as follows: Section 1 introduces the research background and significance of the paper; The second section introduces the current research status of debris flow disaster prediction; The third section introduces the prediction model and implementation method proposed in this article; The fourth section introduces the experiment and analysis; The fifth section summarizes the entire article and provides prospects for future work.

*Corresponding Author: Qingguo Zhou; Email: zhouqg@lzu.edu.cn
DOI: https://doi.org/10.70003/160792642026072704011

2 Related Work

2.1 Overview of the Western Han River Basin

The water level of the Western Han Dynasty is located in the northern part of Longnan City, Gansu Province. The basin is adjacent to the main stream of the Jialing River to the east, connected to the Bailong River basin to the west and south, and bordered by the Yellow River basin to the west and north. The basin area is 10178 km^2. As shown in Figure 1.

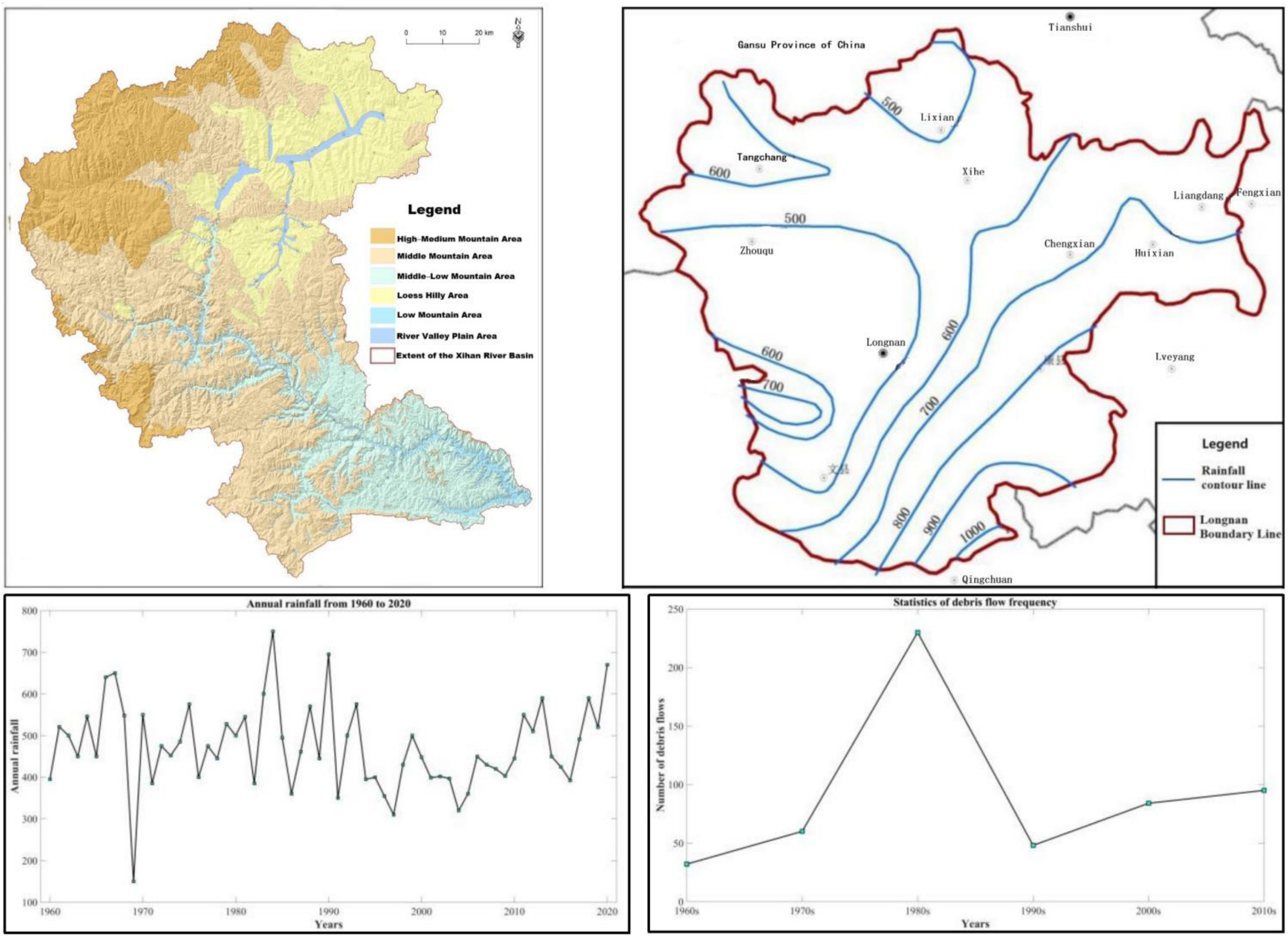

Figure 1. Annual rainfall and statistics of debris flow from 1960 to 2020

The water flow domain of the Western Han Dynasty belongs to a continental monsoon climate with abundant rainfall. The average annual precipitation ranges from 450 mm to 750 mm, and the distribution of precipitation within the year is extremely uneven. There are many heavy rains from June to September, which often lead to landslides, debris flows and other disasters.

The Western Han River Basin and its adjacent areas are located on the northeast edge of the Sichuan Qinghai Tibet Plateau, and are key areas for structural transformation between active faults with different directions and properties. The main characteristics of the development of structural features are the intersection, transformation, and superposition of large anticlines and synclines with large thrust faults and strike slip faults, forming various structural features of different orders and scales, strongly transforming the rock strata, and relatively low stratigraphic stability.

The upper reaches of the Western Han River Basin belong to the loess hilly and gully area, with loess covering the surface and poor vegetation. The rivers in the area are cut flat, the mountains are low and the gullies are crisscrossed, and the slope of the mountain is between 6° and 25°. During the inlaid period, the farmland is mostly distributed in a stepped shape, with poor soil quality, severe vegetation damage, and serious soil erosion; The middle reaches belong to the Huicheng Basin region. The downstream mountains are steep, the canyon is deep, and the bedrock is exposed. In recent decades, with the continuous increase of population, the reclamation area has been expanding, and a large amount of sloping farmland has been plowed and loosened after the summer fields mature, causing damage to the slope structure and easy soil erosion, creating conditions for the formation of debris flows.

2.2 The Influencing Factors of Debris Flow Disasters

Research has shown that the formation of debris flows typically requires three key conditions: fragile geological structures, steep slopes, and sufficient water flow [2]. In general, the geological structure of a region remains basically unchanged for a certain period of time, so areas with unstable geological structures have long been at risk of landslides and debris flow disasters. Due to the influence of earthquakes and human activities, the terrain and landforms of a region have time-varying characteristics. Areas with steep slopes, cliffs, and low vegetation coverage are prone to debris flow disasters, so terrain and landforms are a key factor in triggering debris flows. In addition, continuous rainfall in the short term in a region often makes the soil and rocks on the slopes moist, the surface load is heavy, the adhesion is poor, and it is easy to slide, thus forming debris flows. Therefore, rainfall is the most active factor in inducing debris flow activities. In addition, the risk of debris flow is also related to the blockage of debris flow channels.

Therefore, in the study of debris flow disaster evaluation models, it is necessary to focus on the main control factors among the three conditions mentioned above to evaluate the degree of danger of debris flow occurrence. Due to the complex geographical and environmental conditions in which debris flows occur, there are numerous potential influencing factors. Therefore, the selection and weight determination of influencing factors are key issues that researchers need to consider. It is necessary to comprehensively consider the interrelationships between various factors during the incubation and eruption process of debris flows, and extract several key factors with the highest correlation to establish an evaluation model.

2.3 Method for Predicting the Risk of Debris Flow Disasters

Mudslide disasters represent one of the most prevalent natural hazards, posing significant threats to human life and property. Numerous researchers, both domestically and internationally, have conducted studies to predict and evaluate the risks associated with debris flow in various regions. These risk assessment approaches can generally be classified into several categories, including expert judgment methods, physical modeling techniques, statistical analysis models, and artificial intelligence-based approaches.

The expert evaluation method [9] involves experts conducting on-site inspections and monitoring of the evaluation area to understand factors such as terrain, landforms, geological structures, meteorology, hydrology, etc., and to grasp the causes, characteristics, and laws of debris flow disasters. Based on experience, experts can determine the risk of debris flow disasters in the area. This method is highly subjective and mainly depends on the experience and knowledge of experts, with high costs.

The physical model evaluation [10] method is to study the causes, characteristics, and laws of debris flow disasters, establish mathematical models, simulate the occurrence process of debris flow disasters by inputting relevant parameters, predict the time, location, scale, and impact range of debris flow disasters. However, due to the complex causes of debris flow disasters, geological, topographical, meteorological, and hydrological data may not be complete, making it very difficult to establish accurate physical models.

The statistical model evaluation method [11] is to collect historical data of debris flow disasters, including time, location, scale, causes, and other information. Through statistical analysis methods, the law and trend of debris flow disasters are identified, and an early warning model is established. Then, various parameters of the research area are input into the model to obtain the probability of disaster occurrence. This method obtains prior knowledge by using a large amount of historical data, and forms a probability model of current geological disasters based on this prior knowledge, in order to make predictions and evaluations. However, due to the high dependence of this method on data volume and quality, it is difficult to establish models with high prediction accuracy when the data is incomplete or of low quality.

The AI method is a prediction method developed based on statistical model evaluation [11]. It combines remote sensing technology to interpret and process images and data of areas where debris flows occur, extract more complex information about debris flow disasters, and use AI technology to construct prediction models. The commonly used AI models include Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), BP Neural Network (BPNN), etc., all of which have achieved good results. However, debris flow disasters are often related to time series, and the risk of debris flow in a region often changes over time. The advantages of traditional AI algorithms in time series prediction are not obvious.

Recurrent neural networks (RNNs) are widely used in time series prediction due to their ability to remember long-term data features [12], Long Short Term Memory (LSTM) [13], ated Recurrent Units (GRU) [14], onvolutional eural Networks (CNN) [15], Transformer based Time Series Forecasting (TFT) [16], and so on. Deep learning models that leverage temporal dependencies have proven effective in automatically capturing salient features, yielding impressive results in fields such as speech and semantic recognition, and their applicability extends to a wide range of domains.

This study presents a case analysis of debris flow risk prediction in the Xihan River Basin, located in Longnan, Gansu Province, China. Initially, the raw data were subjected to preprocessing, including cleaning, missing value imputation, and normalization. Subsequently, Principal Component Analysis (PCA) was employed to reduce dimensionality and remove correlations among variables, providing a refined input set for temporal modeling. Long Short-Term Memory (LSTM) networks were then utilized to perform time-series predictions, and the model's predictive performance was rigorously assessed to validate its reliability.

3 Debris Flow Time Series Prediction Model

The prediction of debris flows involves numerous influencing factors, which makes constructing precise physical models a challenging task. In this study, the modeling process begins with comprehensive data acquisition, followed by thorough preprocessing steps such as handling missing values, rasterizing the data, normalization, temporal alignment, and annotation. To reduce complexity and highlight the most critical variables, Principal Component Analysis (PCA) was applied to extract the key factors from the full set of influencing variables. The refined dataset was subsequently split into training, testing, and validation portions. An improved Long Short-Term Memory (LSTM) network was then employed to perform time-series prediction of debris flow occurrences. Finally, the model's performance was systematically assessed. A schematic representation of the complete modeling workflow is provided in the Figure 2.

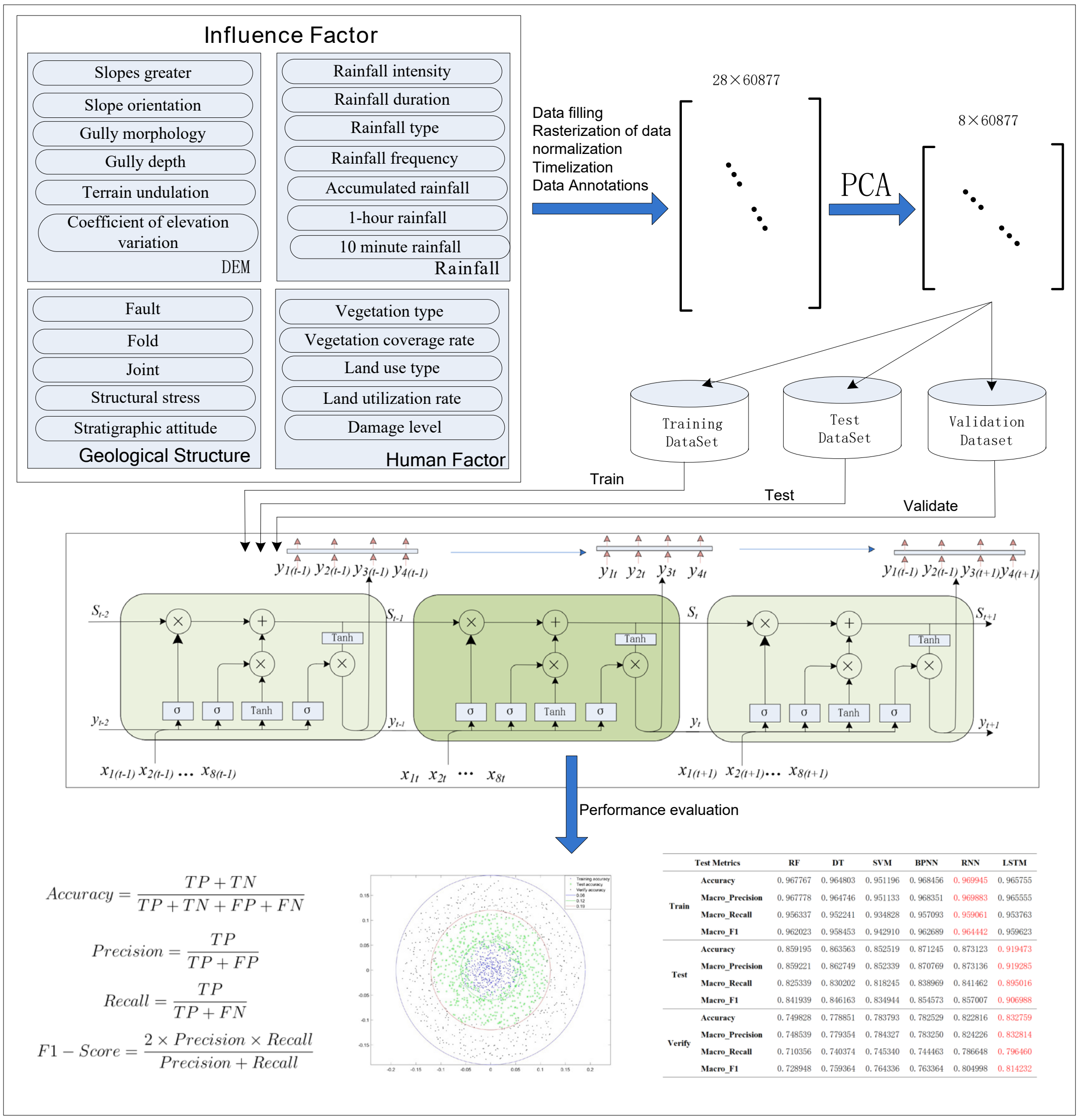

Figure 2. The technical roadmap of this experiment

3.1 Data Acquisition and Dataset

There are many factors that affect debris flows, which can be broadly divided into two categories: natural factors and human factors. Natural factors mainly include rainfall, geological structures, and topography. Rainfall is one of the main factors inducing debris flow, especially heavy rain, rainstorm and long-term continuous rainfall. Rainfall and intensity within 24 hours before debris flow outbreak are very critical factors. Geological structure is also an important factor affecting debris flows, and geological structural lines such as faults, folds, joints, and rock formations have a significant impact on the generation and development of debris flows. The terrain and topography can also affect the occurrence of debris flows, especially in terms of slope, height, and valley characteristics.

Human factors mainly include unreasonable development, indiscriminate logging of trees, construction of highways and railways that damage the surface of mountain slopes, and unreasonable quarrying and mining that damage the geological structure, all of which can induce and exacerbate the occurrence of debris flows.

This study investigates debris flows in the Western Han River basin, Longnan, Gansu Province, China. A total of 48 channels across seven key towns were analyzed, integrating a 1:100,000 geological map with a disaster point database to produce 2,869 investigation sites. From these, 870 sites with sufficiently complete data were selected for further study, including metrics such as watershed area, slope gradient, average gully bed slope, and watershed incision density. Among the selected sites, 220 corresponded to confirmed debris flow events, 242 were classified as high-risk locations, and 408 served as general observation points. Historical rainfall records spanning 50 years yielded 522,000 observations, forming the temporal prediction dataset. The survey data encompass geological, topographical, and land-use attributes. Each site was assigned a debris flow risk level based on historical occurrences. Detailed data sources are summarized in Table 1.

The data preprocessing stage includes multiple steps:

1) Data filling. For missing data, we interpolate and smooth the data from the previous year and the following year to obtain the current year's data, such as Rainfall, Human Factor, etc.

2) Data rasterization. Using GIS tools such as focus statistics and partition statistics, analyze the grid within the slope unit and estimate the results, such as DEM Geological Structure, Geological hazard susceptibility level.

3) Data normalization. Normalize the data of each influencing factor to [0,1].

Table 1. The source of basic data

Data category	Source description	Data representation
DEM (Digital Elevation Model)	Obtained from regional geospatial databases and used for deriving terrain parameters such as slope, aspect, and surface roughness of the study area	Raster grid with a spatial resolution of 10 m × 10 m
Geological Information	Provided by the national geological survey archives; utilized to extract lithology, structural features, and fault distribution	1:100000 geological compilation maps and associated vector layers
Precipitation Records	Sourced from the Lanzhou Meteorological Center and the Longnan debris-flow monitoring network; used to construct long-term rainfall sequences	Time-series precipitation datasets stored as vectorized records
Anthropogenic Factors	Generated through on-site surveys, interviews, and documented investigation materials to characterize human disturbances and land-use patterns	Vector-format thematic data collected from field investigations

3.2 PCA based Dimensionality Reduction of Impact Factors

There are many influencing factors of debris flow disasters. In order to accurately predict the susceptibility of debris flow disasters, we extracted 28 influencing factors from four aspects: rainfall, geological structure, topography, and human factors. However, it is evident that the correlation between these influencing factors and the susceptibility of debris flows is not the same, and some even have no correlation. To explore the applicability of regional geological hazard susceptibility evaluation models and the accuracy of evaluation results. In the selection of evaluation factors, it is necessary to be reasonable, flexible, and reflect the impact of factors on debris flow, while ensuring that each factor does not affect each other and does not have strong correlation.

In terms of correlation analysis, researchers have developed many methods. Such as Pearson Correlation Coefficient, Spearman's Rank Correlation Coefficient, Covariance, regression analysis, etc. Correlation analysis methods can analyze the correlation between two or more variables, which is intuitive and simple, and can clearly display the correlation and degree between variables. However, they can only reveal the linear relationship between variables, and may not accurately describe non-linear relationships. Furthermore, there is no clear guidance on how to select some relevant parameters.

The dataset parameters used in this paper mainly have 4 categories, totaling 28 parameters, the parameters and definitions of the vector expressions are shown in Table 2. Obviously, there is a strong correlation within each type of parameter, but the 28 parameters have different correlations with the susceptibility of debris flows. We not only need to identify the correlation between these parameters and the susceptibility of debris flows, but also need to select some strongly correlated parameters as the dataset for the research content of this article. So, based on the covariance formula, we calculated the covariance matrix of the parameter list, and obtained the eigenvalues and eigenvectors of the covariance matrix to identify strongly correlated parameters. In other words, it is to orthogonalize these 28 influencing factors and select the 8 vectors with the highest correlation from the orthogonal influencing factors as the selected principal components. This method is Principal Component Analysis (PCA). We used the PCA algorithm to identify the 8 parameters with the strongest correlation to the susceptibility of debris flows as further research objects.

We assume that the dataset X has m entries and the data dimension is n, then $X=[x_1,x_2,\ldots,x_n]^T$, $x_i=[x_{i1},x_{i2},\ldots,x_{in}]$. The PCA calculation process is divided into the following steps [17]:

(1) Firstly, sample centralization involves calculating the mean of each feature in the sample and calculating the difference between the sample and the mean:

$$\bar{v}_i = \frac{1}{M}\sum_{j=1}^{M} v_{ij} \tag{1}$$

$$\theta_{ij} = v_{ij} - \bar{v}_{ij} \quad \textbf{(2)}$$

(2) Further use θ Calculate the covariance matrix:

$$\text{cov} = \frac{1}{m}\sum_{i=1}^{M} \theta\theta^T \quad \textbf{(3)}$$

(3) Solve the eigenvectors and eigenvalues of the covariance matrix:

$$\text{cov} = P\Lambda P^T \quad \textbf{(4)}$$

Among them, P is the matrix formed by the eigenvectors, eigenvalues matrix Λ is a diagonal matrix. Arrange by value size, we can take the first s corresponding eigenvectors $p_1, p_2, ..., p_s$ in P, the projection matrix is obtained as:

$$P' = (p_1, p_2, ..., p_s) \quad \textbf{(5)}$$

P' is the sample matrix after dimensionality reduction.

3.3 Debris Flow Time Series Prediction Model Based on LSTM

LSTM have been widely recognized for their effectiveness in modeling temporal dependencies. Leveraging this property, we constructed a debris flow prediction model designed for time-series forecasting with real-time capabilities. Input features consist of an 8-dimensional vector obtained through Principal Component Analysis (PCA) for dimensionality reduction, while the model outputs a 4-dimensional vector representing four distinct debris flow risk levels. The network architecture includes a single hidden layer with 200 neurons, and a Dropout layer with a probability of 0.5 is positioned between the hidden and output layers to mitigate overfitting. Sequence data are processed with a temporal window of 12 steps, and a Softmax function is employed at the output layer to convert raw scores into categorical risk probabilities. A schematic of the simplified model is illustrated in Figure 3, with full hyperparameter settings listed in Table 3.

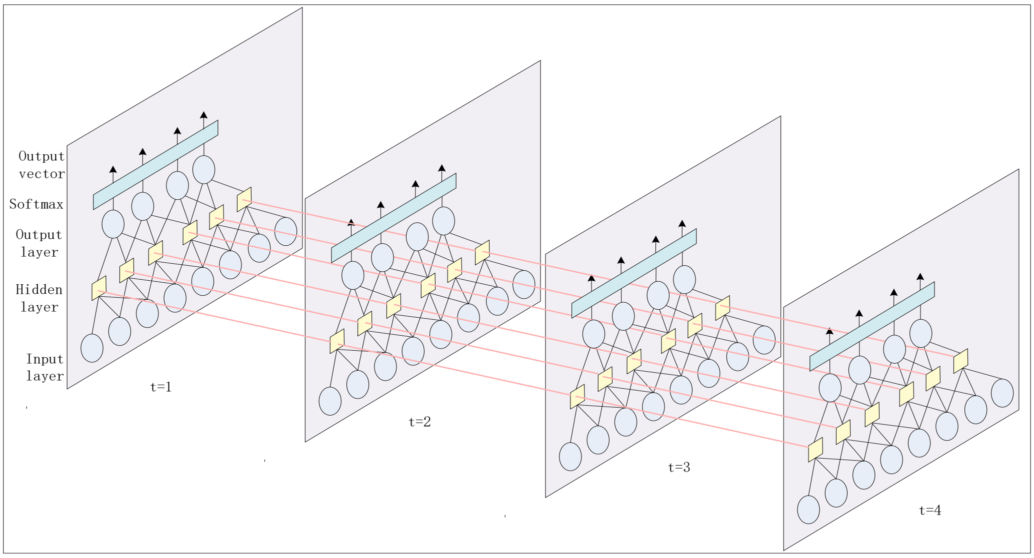

Figure 3. Simplified model diagram of the LSTM

Table 2. All parameters and descriptions of the dataset

Vector	Parameter	Definition
X1	Slope direction	The orientation of the terrain slope
X2	Gully morphology	Describe the spatial structure and appearance characteristics of the gully
X3	Gully depth	Vertical distance from the bottom of the gully to the surrounding ground
X4	Terrain undulations	The difference between the elevation of the highest point and the elevation of the lowest point in a specific area
X5	Height variation coefficient	The ratio of the elevation change of a point or area in the terrain to the horizontal distance
X6	Direction of rock strata	The straight line obtained from the intersection of the bedding plane and the horizontal plane is called the strike line, and the direction indicated at both ends of the strike line is the direction of the rock strata.
X7	Rock inclination	indicates the downward inclination of a rock layer.
X8	Rock inclination angle	Refers to the degree of inclination of a rock layer, that is, the angle between the inclined plane of the rock layer and the horizontal plane.
X9	Fold morphology	The shape of a previously nearly flat curved surface in rocks

X10	Wrinkle wavelength	The distance between adjacent peaks or valleys in a folded pattern
X11	Fold amplitude	Describe the degree or size of undulation in rock folds
X12	Fold axis plane	The central plane of the fold
X13	Fold hub	Describe the fold position
X14	Fault types	The relative displacement mode of two fault plates, such as normal fault, reverse fault, translational fault, etc.
X15	Fault displacement	Relative displacement of strata on both sides of the fault
X16	Fault plane orientation	The plane direction or inclination of a fault
X17	Rainfall intensity	The total amount of precipitation per unit time
X18	Rainfall duration	The time elapsed from the beginning to the end of rainfall
X19	Rainfall type	Classification based on the intensity of precipitation
X20	Rainfall frequency	The number or probability of rainfall occurring within a specific time period
X21	Accumulated rainfall	The total amount of rainfall received for six consecutive months
X22	1-hour rainfall	Precipitation within 1 hour
X23	10 minute rainfall	Precipitation within 10 minutes
X24	Vegetation type	The overall population of all plant species growing within a certain region
X25	Vegetation coverage rate	The ratio of vegetation area to total land area
X26	Land use type	It can be divided into cultivated land, forest land, grassland, and construction land
X27	Land utilization rate	The proportion of utilized land to total land area
X28	Damage level	The degree of land destruction

Table 3. Network hyperparameter settings

NO.	Hyperparameter	Explanation	Value
1	Input_size	Input vector dimension	8
2	Hidden_size	Number of hidden layer cells	200
3	Bias	Does the hidden layer have bias	True
4	Dropout	Is dropout necessary	True, Dropping probability is 0.5
5	Activation function	Activation function	Sigmoid
6	num_layers	The number of layers of RNN	3
7	Batch_size	Batch size	60
8	Output_size	Output vector dimension	4
9	L	Sequence length	12
10	Max_iterations	Max iterations	10000
11	Error	Error calculation method	$MSE = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(y-\hat{y}\right)^{2}$ **(6)**
12	Stop_condition	Training stop conditions	Iterations>Max_iterations Or Error<0.3
13	Rate	Learning rate	0.01

4 Experiments and Analysis

In this study, 870 time-series samples were utilized. A stratified sampling strategy was adopted, in which 30% of the samples (261 records) were assigned to the validation set. From the remaining 609 samples, 45 years of sequential data were employed for model training, while the most recent 5 years were reserved as an independent test set to rigorously evaluate generalization performance.

All experiments were conducted on a workstation equipped with an Intel Core i7-6700 CPU (3.40 GHz), 32 GB RAM, and an NVIDIA RTX 3090 GPU, running Windows 10 (64-bit), MATLAB R2021a.

4.1 Testing Methods and Indicators

The evaluation of a model's performance involves examining both its ability to learn from training data and its capacity to generalize to unseen data. In this study, the learning effectiveness of each model was quantified using Accuracy, while its generalization performance was assessed through multiple metrics, including *Accuracy, Macro-Precision*, *Macro-Recall*, and *Macro-F*1 score. The definitions and computational formulas for these indicators are provided below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{10}$$

Where TP (true positive) is the number of positive samples correctly classified, FP (false positive) is the number of negative samples misclassified as positive, *FN* (false negative) is the number of positive samples misclassified as negative, and *TN* (true negative) is the number of negative samples correctly classified. These formulas are generally used for binary classification problems. In the multi classification problem, we need to define positive and negative samples for each category to achieve classification for each category, and calculate their Precision (Pi), Recall (Ri), and F1 ($F1i$) respectively, and use the following equations to calculate *Macro-Precision*, *Macro-Recall*, *Macro-F*1, *Micro-Precision*.

$$macro\text{-}Precision = \frac{1}{k}\sum_{i=1}^{k} P_i \tag{11}$$

$$Macro\text{-}Recall = \frac{1}{k}\sum_{i=1}^{k} R_i \tag{12}$$

$$Macro - F1 = \frac{1}{k}\sum_{i=1}^{k} F1_i \tag{13}$$

$$Micro - Precicion = \frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k} TP_i + \sum_{i=1}^{k} FP_i} \tag{14}$$

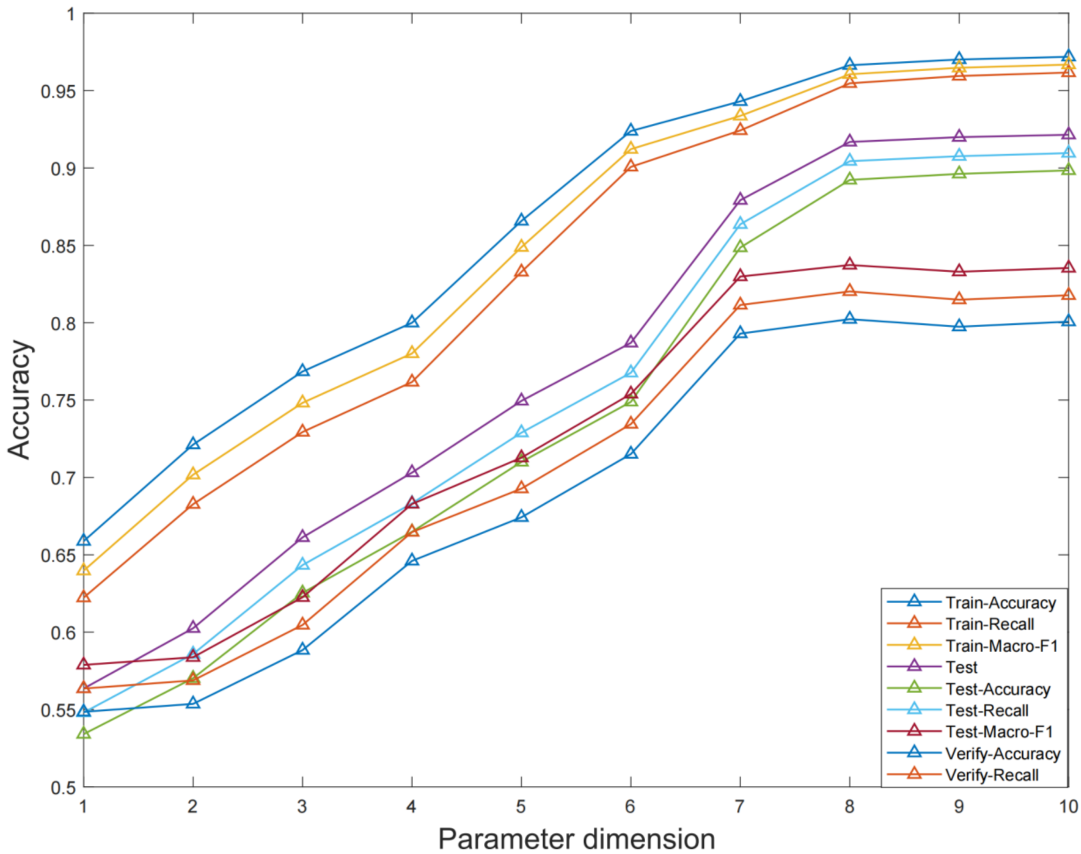

Figure 4. Testing accuracy of LSTM under different dimensional conditions

4.2 Test Results and Analysis

4.2.1 PCA Algorithm Performance Testing

We used PCA algorithm to reduce and sort the 28 dimensional dataset, and trained and tested the LSTM model using 1-dimensional principal components to 10 dimensional principal components as dataset samples. In the testing, after repeated validation and determination of the optimal hyperparameters, we used the same model. Each model was tested 10 times on a dataset, and the average value was calculated as the final result, as shown in Figure 4. From the results, it can be seen that as the dimensionality of the selected principal components increases, the training accuracy, testing accuracy, and validation accuracy all improve. However, as the dimensionality gradually increases, the magnitude of accuracy improvement gradually decreases. It should be noted that when the dimension is 8 dimensions, the performance improvement becomes slower and slower, while the performance indicators of the validation set still slightly decrease. Therefore, we chose an 8-dimensional vector as the principal component of the PCA algorithm after dimensionality reduction for further experiments.

We projected the raw data onto the principal components in the top 8 dimensions, with the first principal component as the X-axis, the second principal component as the Y-axis, and the other principal components as the Z-axis. We plotted a 3D scatter plot, and the visualization results are shown in Figure 5.

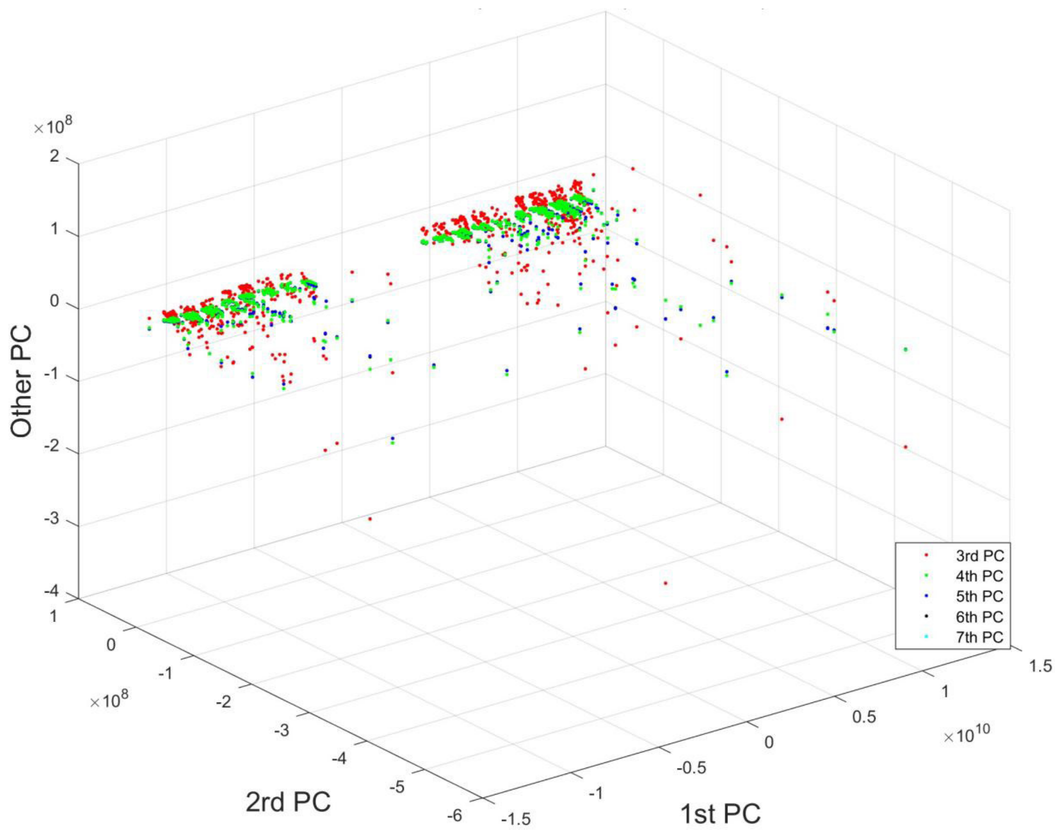

Figure 5. Visualization results of 8 principal components

4.2.2 Training and Testing of Predictive Models

The proposed model achieved a prediction accuracy of 96.6% on the training dataset, while the accuracy on the testing dataset reached 91.94%. Subsequently, an independent validation dataset was used to further assess model generalization, yielding an accuracy of 83.3%. To evaluate the distribution characteristics of prediction deviations, a total of 588 output residuals were converted into angular coordinates by mapping the residual magnitude to the radial direction of a polar coordinate system. Based on this transformation, polar error-distribution plots were generated for the training, testing, and validation subsets. The training and test results of the model are shown in Figure 6. The results show that training residuals are consistently below 0.06, and testing residuals remain under 0.12, indicating stable model convergence and reliable predictive performance on unseen but distribution-similar samples. Although the validation dataset is entirely independent and does not overlap with the training or testing samples, the maximum validation residual remains within 0.19, demonstrating that the model maintains a reasonable prediction consistency even under distribution shift. Overall, these results confirm that the proposed method exhibits strong robustness and favorable generalization capability for debris-flow risk prediction.

4.2.3 Comparative Experiments with Other Models

The predictive capability of the proposed algorithm for debris flow risk was evaluated through comparative experiments involving several machine learning techniques, including Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Backpropagation Neural Network (BPNN), and Recurrent Neural Network (RNN). For each model, ten independent rounds of training, testing, and validation were performed, with the iteration achieving the highest training accuracy selected as the representative result. Performance assessment was conducted using multiple metrics: Accuracy, Macro Precision, Macro Recall, and Macro-F1. The accuracy outcomes for the training, testing, and validation sets are provided in Table 4, while the associated confusion matrices are illustrated in Figure 7.

To ensure a fair comparison among the competing algorithms, all models were optimized until their training accuracy converged to approximately 96.5%. Despite comparable fitting performance on the training data, the subsequent evaluation revealed notable differences in predictive capability. The LSTM model consistently produced the highest test accuracy, with RNN following closely and BPNN showing performance similar to that of RNN. Moreover, LSTM achieved the strongest scores across the Macro-Precision, Macro-Recall, and Macro-F1 metrics, highlighting its superior ability to handle temporal dependencies inherent in debris-flow sequences.

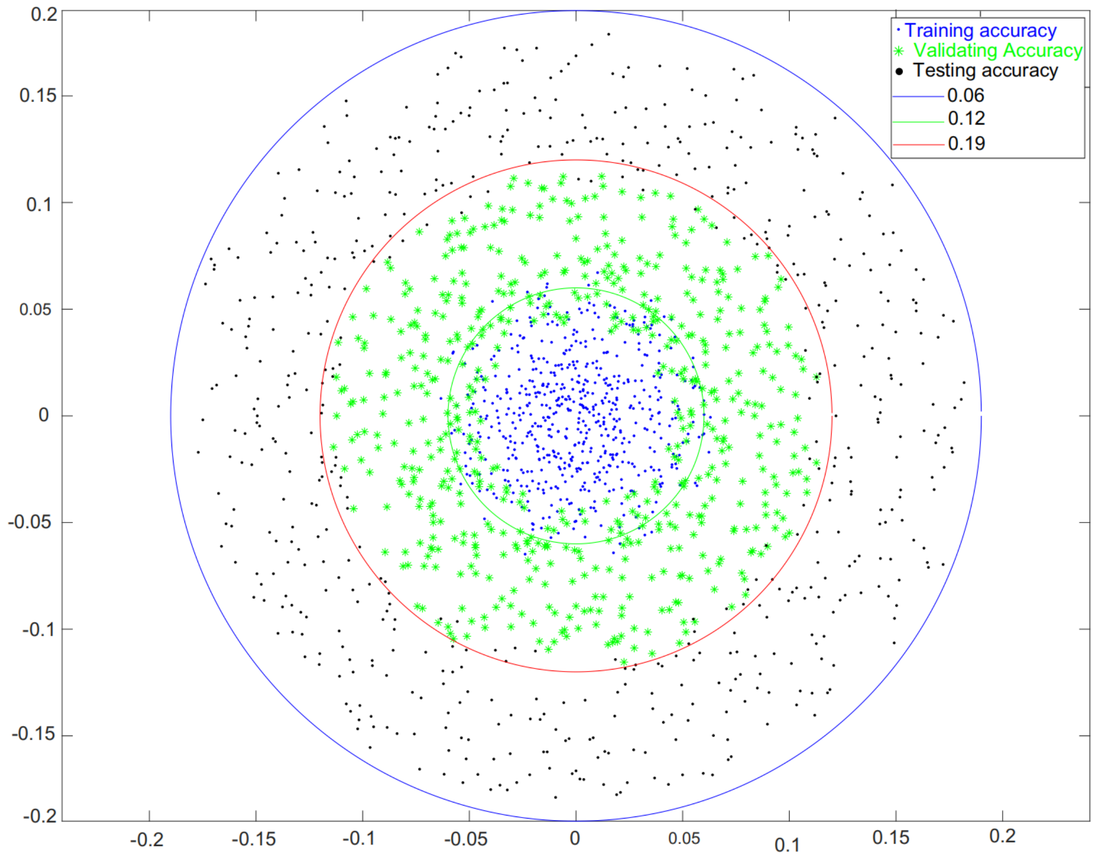

Figure 6. Training and test accuracy of LSTM

A marked decline in validation accuracy was observed across all methods when transferred to previously unseen samples, indicating diminished detection robustness when encountering distribution-shifted data. Nevertheless, LSTM retained the best generalization level, reaching 83.2759%, which exceeds the RNN and BPNN models by 0.9943% and 5.0233%, respectively. Its advantage is further reinforced by the consistently higher macro-level metrics, suggesting more balanced performance across all categories.

Collectively, the results demonstrate that among the evaluated models, LSTM provides the most reliable and stable predictive behavior. Its strong generalization and superior temporal-feature extraction make it particularly well suited for debris-flow hazard forecasting.

ROC curve (receiver operating characteristic curve) is commonly used to predict accuracy. It depicts the performance of different algorithms with False Positive Rate as the horizontal axis and True Positive Rate as the vertical axis. AUC (Area Under Curve) is the area under the ROC curve, which represents the algorithm's predictive ability. The range of AUC values is between 0 and 1. The closer the value is to 1, the better the performance of the classifier, and the closer the value is to 0.5, the worse the performance of the classifier. The significance of AUC is to measure the classifier's ability to correctly sort positive and negative samples. In binary classification problems, positive and negative examples represent different categories, respectively. AUC reflects the probability that the classifier ranks positive examples before negative examples. Therefore, the higher the value of AUC, the stronger the classifier's ability to rank positive examples before negative examples, that is, the better the performance of the classifier.

Therefore, in order to characterize the advantages and disadvantages of several algorithms, we binarized the classification results, plotted the ROC curves of training samples, test samples, and validation samples, and calculated the AUC values, as shown in Figure 8. From the graph, it can be seen that in the ROC curve of the training samples, RNN has the highest AUC, reaching 0.9770, followed by BPNN with an AUC of 0.9757, while LSTM's AUC is not optimal; However, in the ROC curve of the test samples, the AUC of LSTM is 0.9391, which is much higher than other algorithms, and the AUC of RNN ranks second; In the ROC curve of the validation sample, LSTM still has the highest value of 0.87409, followed by RNN. It can be seen that under the same training conditions, LSTM has the best generalization ability and is an ideal model for temporal prediction and evaluation of debris flow disasters.

From the above experimental results analysis, it can be concluded that PCA obtains the key influencing factors of the original samples by orthogonalizing the collected samples. The LSTM algorithm consciously preserves or discards information in the time series, maximizing the integrity of the information and ensuring the accuracy and generalization ability of the entire prediction result.

Table 4. Training and prediction results of several algorithms

Test metrics		RF	DT	SVM	BPNN	RNN	LSTM
Train	Accuracy	0.967767	0.964803	0.951196	0.968456	**0.969945**	0.965755
	Macro_Precision	0.967778	0.964746	0.951133	0.968351	**0.969883**	0.965555
	Macro_Recall	0.956337	0.952241	0.934828	0.957093	**0.959061**	0.953763
	Macro_F1	0.962023	0.958453	0.942910	0.962689	**0.964442**	0.959623
Test	Accuracy	0.859195	0.863563	0.852519	0.871245	0.873123	**0.919473**
	Macro_Precision	0.859221	0.862749	0.852339	0.870769	0.873136	**0.919285**
	Macro_Recall	0.825339	0.830202	0.818245	0.838969	0.841462	**0.895016**
	Macro_F1	0.841939	0.846163	0.834944	0.854573	0.857007	**0.906988**
Verify	Accuracy	0.749828	0.778851	0.783793	0.782529	0.822816	**0.832759**
	Macro_Precision	0.748539	0.779354	0.784327	0.783250	0.824226	**0.832814**
	Macro_Recall	0.710356	0.740374	0.745340	0.744463	0.786648	**0.796460**
	Macro_F1	0.728948	0.759364	0.764336	0.763364	0.804998	**0.814232**

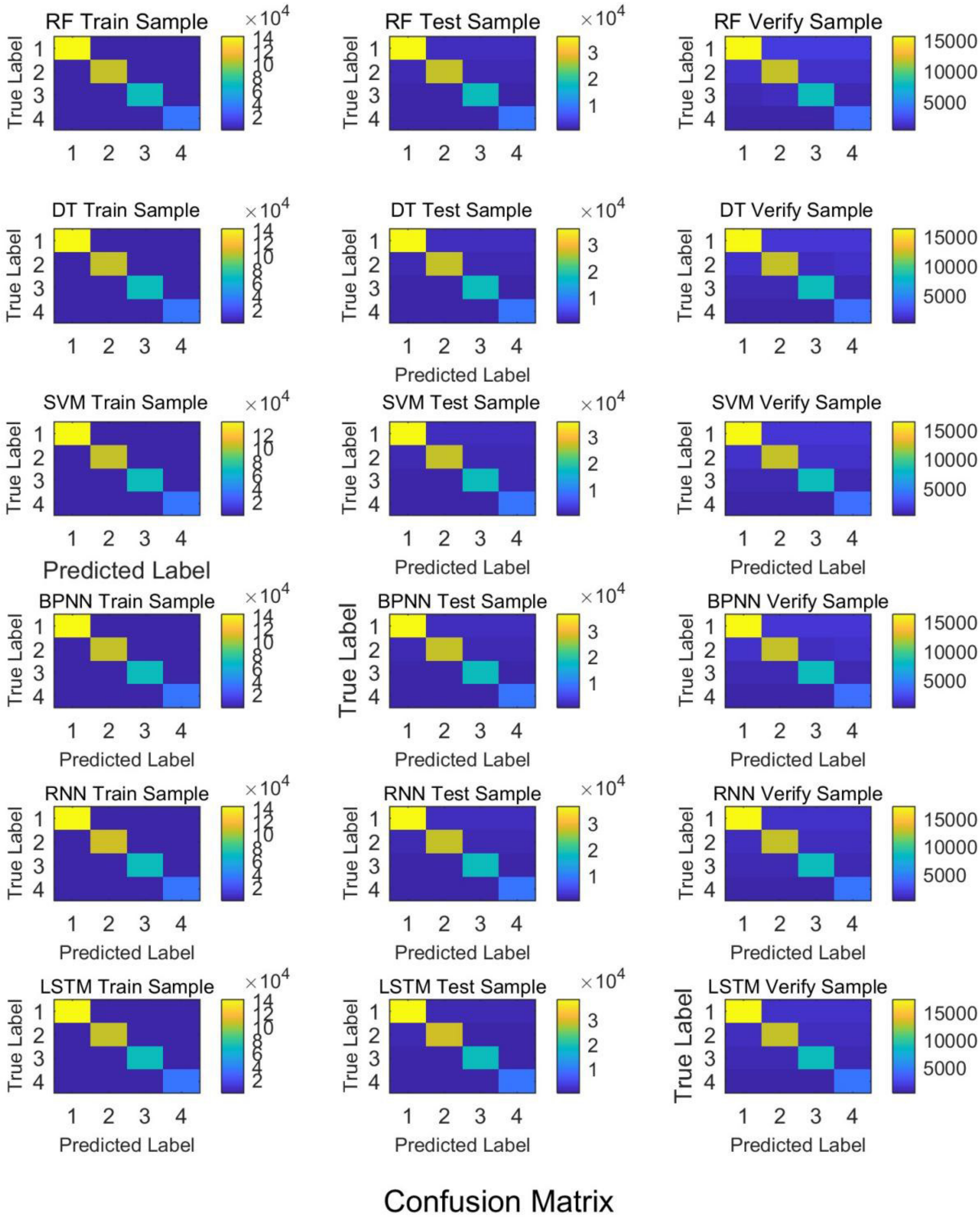

Figure 7. Confusion matrix of the models

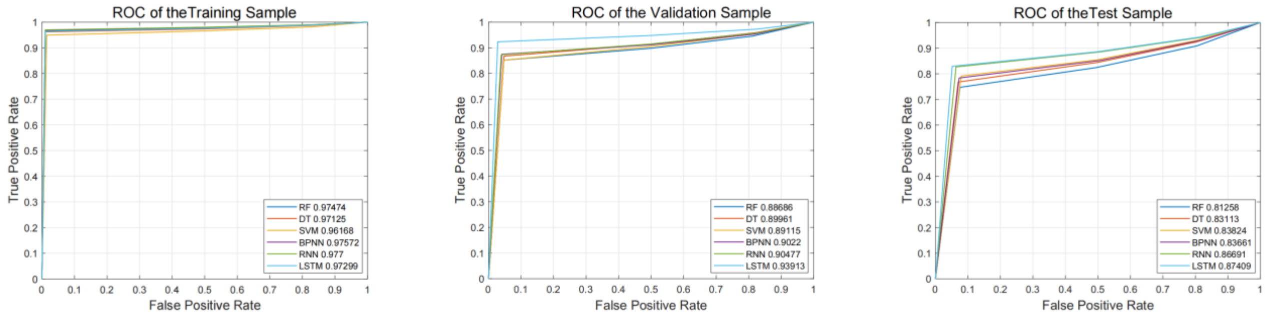

Figure 8. ROC curves and AUC values of several algorithms

5 Conclusion

Mudslides are among the most common geological hazards worldwide, posing significant threats to human life and property. Accurate assessment of debris flow risk and reliable prediction based on real-time monitoring data are essential prerequisites for effective early warning and disaster prevention. Constructing a robust risk prediction model is therefore critical for mitigating debris flow impacts. This study focuses on developing and evaluating a temporal prediction model for debris flow risk in the Western Han River basin of the Longnan region, an area particularly susceptible to debris flows in northwest China. Initially, a machine learning dataset was prepared by cleaning and labeling existing survey and historical data. Principal Component Analysis (PCA) was then applied to reduce dimensionality and remove correlations, resulting in 8-dimensional feature vectors representing the most relevant factors. Using these inputs, a temporal prediction model based on the Long Short-Term Memory (LSTM) network was constructed. Following training and testing, the model's predictive performance was validated, achieving a testing accuracy of 91.95% and a validation accuracy of 83.28%. Comparative analysis indicates that the LSTM model demonstrates superior generalization performance relative to alternative approaches. These results provide a technical basis for precise debris flow forecasting in the Longnan Mountains and offer valuable insights for establishing real-time early-warning systems for small- to medium-scale debris flows.

However, due to the imbalance and high data noise in the dataset studied in this article, the prediction accuracy of the model has not yet reached the expected goal. In future work, we will combine some optimization algorithms to improve our prediction model and further tune the parameters of the model, in order to design a model with stronger prediction ability and higher generalization ability, make more accurate predictions of evaluation units, and design a digital twin system that can provide accurate decision support to serve relevant departments.

6 Acknowledgements

This work was partially supported by National Key R&D Program of China under Grant No. 2023YFB4503903 and 2020YFC0832500, National Natural Science Foundation of China under Grant No. U22A20261 and 61402210, Gansu Province Science and Technology Major Project - Industrial Project under Grant No. 22ZD6GA048, Gansu Province Key Research and Development Plan - Industrial Project under Grant No. 22YF7GA004, the Fundamental Research Funds for the Central Universities under Grant No. lzujbky-2022-kb12, lzujbky-2021-sp43, lzujbky-2020-sp02, lzujbky-2019-kb51 and lzujbky-2018-k12, Science and Technology Plan of Qinghai Province under Grant No. 2020-GX-164, Supercomputing Center of Lanzhou University.

This work was also partially supported by the Gansu Provincial Science and Technology Major Special Innovation Consortium Project (21ZD3GA002), the name of the innovation consortium is Gansu Province Green and Smart Highway Transportation Innovation Consortium.

References

[1] T. Takahashi, Debris flow, *Annual Review of Fluid Mechanics*, Vol. 13, No. 1, pp. 57–77, January, 1981. https://doi.org/10.1146/annurev.fl.13.010181.000421

[2] K. Hutter, B. Svendsen, D. Rickenmann, Debris flow modeling: A review, *Continuum Mechanics and Thermodynamics*, Vol. 8, No. 1, pp. 1–35, February, 1994. https://doi.org/10.1007/BF01175749

[3] M. Hürlimann, D. Rickenmann, V. Medina, A. Bateman, Evaluation of approaches to calculate debris-flow parameters for hazard assessment, *Engineering Geology*, Vol. 102, Nos. 3–4, pp. 152–163, December, 2008. https://doi.org/10.1016/j.enggeo.2008.03.012

[4] K. Long, S. Zhang, F. Wei, K. Hu, Q. Zhang, Y. Luo, A hydrology-process based method for correlating debris flow density to rainfall parameters and its application on debris flow prediction, *Journal of Hydrology*, Vol. 589, Article No. 125124, October, 2020. https://doi.org/10.1016/j.jhydrol.2020.125124

[5] A. Blais-Stevens, P. Behnia, Debris flow susceptibility mapping using a qualitative heuristic method and Flow-R along the Yukon Alaska Highway Corridor, Canada, *Natural Hazards and Earth System Sciences*, Vol. 16, No. 2, pp. 449–462, February, 2016. https://doi.org/10.5194/nhess-16-449-2016

[6] C. Tang, J. Zhu, M. Chang, J. Ding, X. Qi, An empirical–statistical model for predicting debris-flow runout zones in the Wenchuan earthquake area, *Quaternary International*, Vol. 250, pp. 63–73, February, 2012.

https://doi.org/10.1016/j.quaint.2010.11.020

[7] B. Q. Luna, J. Blahut, C. Camera, C. van Westen, T. Apuani, V. Jetten, S. Sterlacchini, Physically based dynamic run-out modelling for quantitative debris flow risk assessment: A case study in Tresenda, northern Italy, *Environmental Earth Sciences*, Vol. 72, No. 3, pp. 645–661, August, 2014. https://doi.org/10.1007/s12665-013-2986-7

[8] Y. Zhao, X. Meng, T. Qi, F. Qing, M. Xiong, Y. Li, P. Guo, G. Chen, AI-based identification of low-frequency debris flow catchments in the Bailong River basin, China, *Geomorphology*, Vol. 359, Article No. 107125, June, 2020. https://doi.org/10.1016/j.geomorph.2020.107125

[9] M. G. Winter, J. T. Smith, S. Fotopoulou, K. Pitilakis, O. Mavrouli, J. Corominas, S. Argyroudis, An expert judgement approach to determining the physical vulnerability of roads to debris flow, *Bulletin of Engineering Geology and the Environment*, Vol. 73, No. 2, pp. 291–305, May, 2014. https://doi.org/10.1007/s10064-014-0570-3

[10] J. Sun, S. Qin, S. Qiao, Y. Chen, G. Su, Q. Cheng, Y. Zhang, X. Guo, Exploring the impact of introducing a physical model into statistical methods on the evaluation of regional scale debris flow susceptibility, *Natural Hazards*, Vol. 106, No. 1, pp. 881–912, March, 2021. https://doi.org/10.1007/s11069-020-04498-4

[11] Z. Liang, C.-M. Wang, Z.-M. Zhang, K.-U.-J. Khan, A comparison of statistical and machine learning methods for debris flow susceptibility mapping, *Stochastic Environmental Research and Risk Assessment*, Vol. 34, No. 11, pp. 1887–1907, November, 2020. https://doi.org/10.1007/s00477-020-01851-8

[12] W. Yin, K. Kann, M. Yu, H. Schütze, Comparative study of CNN and RNN for natural language processing, *arXiv preprint*, arXiv:1702.01923, February, 2017. https://doi.org/10.48550/arXiv.1702.01923

[13] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, November, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[14] T. Lei, Y. Zhang, S. I. Wang, H. Dai, Y. Artzi, Simple recurrent units for highly parallelizable recurrence, *arXiv preprint*, arXiv:1709.02755, September, 2017. https://doi.org/10.48550/arXiv.1709.02755

[15] K. O'Shea, R. Nash, An introduction to convolutional neural networks, *arXiv preprint*, arXiv:1511.08458, November, 2015. https://doi.org/10.48550/arXiv.1511.08458

[16] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J.Yan, L. Sun, Transformers in time series: A survey, *arXiv preprint*, arXiv:2202.07125, February, 2022. https://doi.org/10.48550/arXiv.2202.07125

[17] H. Abdi, L. J. Williams, Principal component analysis, *Wiley Interdisciplinary Reviews: Computational Statistics*, Vol. 2, No. 4, pp. 433–459, July–August, 2010. https://doi.org/10.1002/wics.101

Biographies

Hongtao Zhang received the master degree from Beijing University of Chemical Technology, China in 2012. He is currently pursuing the PhD degree at School of Information Science and Engineering, Lanzhou University, Lanzhou, China. His research interests include water conservancy and hydropower, disaster prevention and reduction, machine learning, and optimization algorithms.

Qingguo Zhou received the BS and MS degrees in physics from Lanzhou University, Lanzhou, China in 1996 and 2001, respectively, and the PhD degree in theoretical physics from Lanzhou University, Lanzhou, China in 2005. Now he is a professor at School of Information Science and Engineering, Lanzhou University, Lanzhou, China. He is also a fellow of IET. His research interests include machine learning, optimization algorithm, and computer application technology.