

Original Research

Prediction of Landslide Susceptibility Based on Neural Network Model and Negative Sample Selected by Information Value Model

Zixuan Wang¹, Jingru Tang^{1,2*}

¹School of Earth Sciences, Guilin University of Technology, Guilin, 541006, China

²Collaborative Innovation Center for Exploration of Nonferrous Metal Deposits and Efficient Utilization of Resources by the Province and Ministry, Guilin University of Technology, Guilin, 541006, China

Received: 21 September 2023

Accepted: 18 April 2024

Abstract

Landslides occur frequently in the Chishui River Basin under the interaction of the geological environment and local human activities, negatively impacting the safety of people and properties, and social order; thus, landslide-prone areas must be analyzed. Here, based on field research and data collection performed in the Chishui River Basin, we identify 13 landslide conditioning factors to construct a landslide susceptibility identification system through principal component analysis by comprehensively considering the geological environment, topography and geomorphology, climate and hydrology, human engineering activities, vegetation cover, and other factors. The information volume model was used to select non-landslide points, and the back-propagation (BP), long- and short-term memory (LSTM), and convolutional neural network (CNN) models were selected to predict the landslide susceptibility zoning in the study area; the area under the curve values of the three models were 0.981, 0.984, and 0.997, respectively. The CNN was significantly more valid in predicting landslide zones than BP and LSTM and could better predict landslide susceptibility. CNNs have a promising future in landslide susceptibility analysis. These findings provide a basis for landslide susceptibility assessment, which can aid in developing appropriate pre-disaster prevention and post-disaster relief programs to decrease the threat posed by existing or future landslides.

Keywords: landslide, susceptibility, convolutional neural networks, back propagation network, long short term memory, non-landslide, Chishui river basin

Introduction

Landslides are common in mountainous regions and have become more frequent because of climate

change, frequent human engineering activities, and rapid urbanization. These changes pose significant risks to human life and property, and hinder such regions' economic development [1-5]. A landslide is a physical and geological phenomenon in which unstable soil and rock flow downward onto a soft surface (or sliding area) under the influence of gravity. Factors inducing landslides are categorized as natural or anthropogenic natural factors [6,

*e-mail: tjr118@glut.edu.cn

7], such as rainfall, vegetation, and river erosion, can lead to a decrease in shear strength and an increase in slope shear stress, which ultimately induces landslides [8, 9]. The construction of roads and bridges, and changes in land use can also trigger landslides; furthermore, if they are not adequately constructed, they may result in more significant and dangerous landslides [8-10]. Geological hazards in China, especially landslides, are prevalent in the Yunnan, Guizhou, and Sichuan Provinces, where the Chishui River flows, as well as a small area in the southwest of Chongqing Municipality. Conducting a landslide susceptibility assessment is a critical step in pre-disaster warning and post-disaster relief. Through vulnerability assessment, landslide-susceptible areas can be identified and a corresponding disaster prevention system can be set up. Further, through vulnerability zone mapping, a post-disaster rescue program can be quickly formulated. Therefore, it is necessary to conduct landslide susceptibility assessments in the Chishui River Basin.

In 1984, Brabb et al. introduced the concept of landslide sensitivity to characterize the likelihood of landslides under the control of regional environmental variables. Susceptibility analysis methods can be classified into two broad categories: knowledge-driven strategies (e.g., the frequency-ratio method, analytic hierarchy process, fuzzy comprehensive appraisal, and information volume model) and data-driven approaches (e.g., artificial neural networks). Knowledge-driven schemes, some of the earliest methods used for landslide susceptibility assessment, rely on the experience of experts, to a certain extent; therefore, they have more stringent requirements for the expertise of experimenters [11]. With increasing access to data, however, landslide assessment has evolved to include more complex knowledge areas. To compensate for the shortcomings of knowledge-driven methods, such as a large workload, high subjectivity, and low accuracy of prediction results, landslide susceptibility evaluation using traditional machine learning methods has gradually become a mainstream method with wide application. However, owing to the shallow structure of traditional machine learning methods, these methods are limited in their ability to extract in-depth information; meanwhile, neural network models have a more complex structure, leading to better generalization ability than general algorithmic models. They also have an advantage in handling voluminous data sets and can quickly obtain global information [12]. For example, Shahri et al. accurately predicted landslide susceptibility in Sweden using artificial neural network modeling [13]. Han et al. trained and validated a landslide susceptibility model for the Boshan District of Zibo City using a convolutional neural network (CNN) optimized for the region [14]. Azarafza et al. proposed combining CNN and deep neural networks (DNNs) for landslide susceptibility assessment in Isfahan province, Iran [12]. Zhang et al. used a CNN to analyze landslide susceptibility and compared it with logistic regression (LR), support vector machine

(SVM), and random forest (RF) models [14]. The CNN outperformed the other models in feature extraction and multidimensional data processing. Ghorbanzadeh et al. focused on exploring the interrelationships among landslide factors when assessing landslide susceptibility based on CNNs, which resulted in better model performance [15].

The accuracy of landslide susceptibility analysis is closely related to the quality of the selection of non-landslide points [16]. According to existing data, there are four main ways to select non-landslide points at home and abroad.

(1) Random method: With this method, researchers randomly select points from outside the known landslide area; the points identified may be similar to the geological background of the landslide area [17], and, therefore, be areas of potential landslides;

(2) Buffer method: Researchers randomly select points outside the appropriate buffer distance from the historical landslide site; however, determining the buffer distance is highly subjective [18];

(3) Slope method: Points are randomly selected from non-landslide areas with slopes below a certain threshold [19]; the points selected by the susceptibility and slope methods in particular are insufficient to reflect the overall characteristics of points.

(4) Susceptibility method: Random selection of points from the initial very-low-susceptibility area can help to reduce the limitations imposed by anthropogenic factors [20].

Therefore, despite these different approaches, the accurate selection of high-quality non-landslide points remains a significant problem that must be solved for landslide susceptibility modeling using neural networks. Landslides have become an obstacle to sustainable social development, causing irreversible damage to daily life, road network construction, and economic development [21, 22]. Landslide susceptibility analysis is particularly important for pre-disaster prevention and relief.

In this study, we used the Chishui River Basin as the study area and aimed to: (1) analyze the corresponding landslide conditioning factors based on principal component analysis (PCA); (2) conduct preliminary landslide susceptibility zoning using the information degree model and randomly select 724 non-landslide points in the very low susceptibility zone; (3) perform landslide susceptibility prediction in the study area using the back-propagation model (BP), long- and short-term memory (LSTM) model, and CNN, and prepare a corresponding landslide susceptibility zoning map (Fig. 1). The methodology of this study is not the same as that when using other single models, aiming to analyze the landslide susceptibility assessment by combining an informative model of the selected non-landslide sites with a neural network model, and to provide an important scientific basis for the establishment of a landslide prevention and control system in the Chishui River Basin.

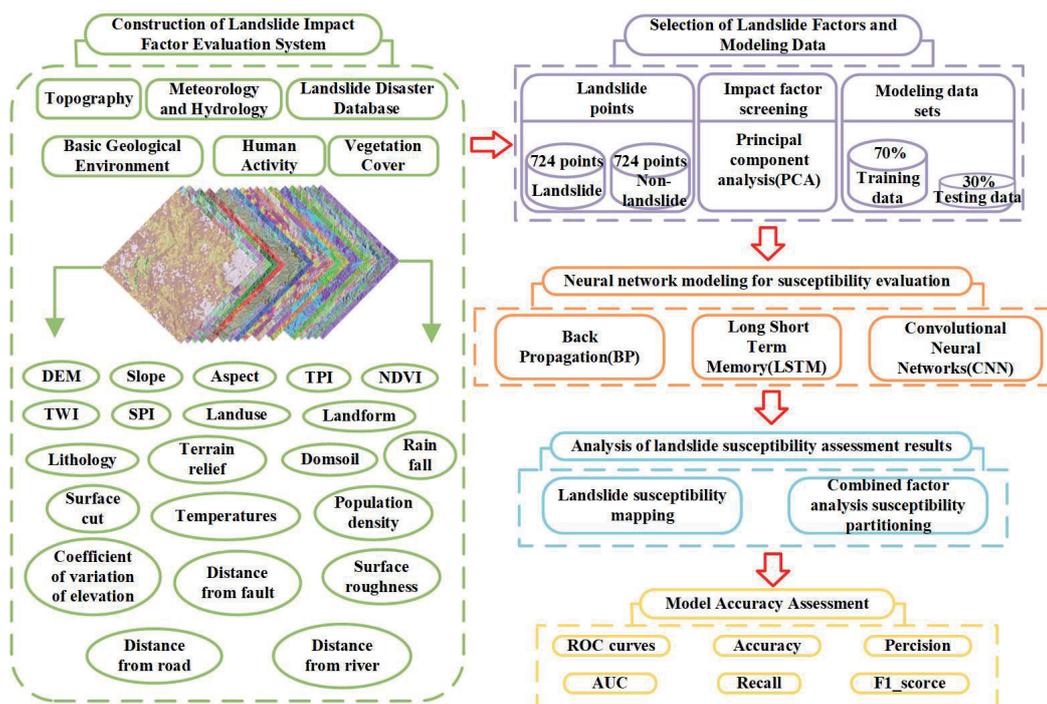


Fig. 1. Landslide susceptibility analysis flow chart.

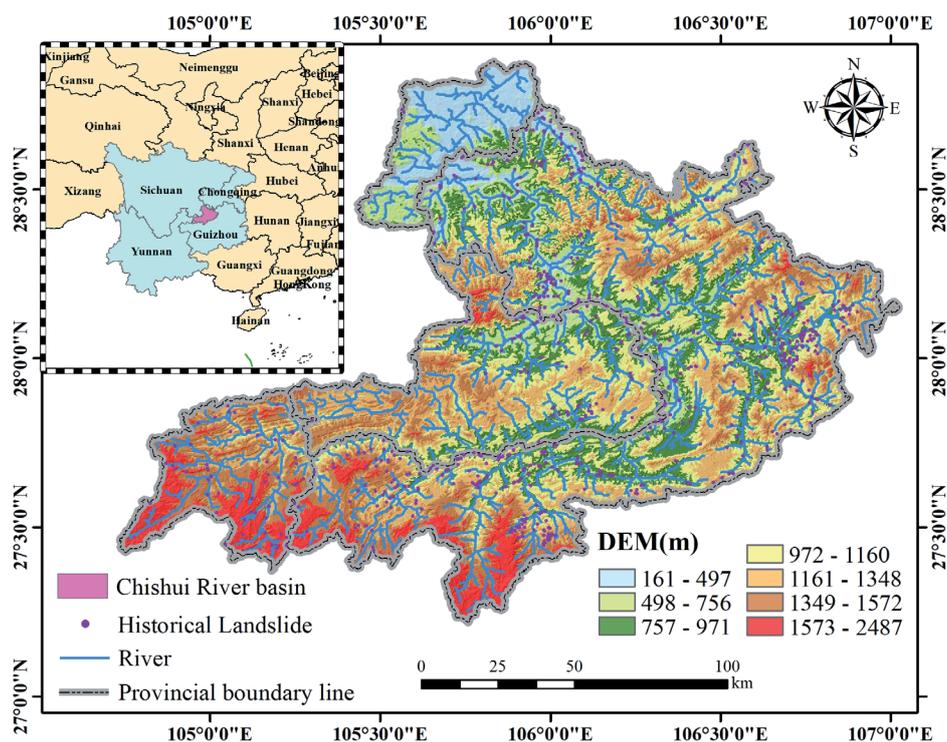


Fig. 2. Location of the study area and the occurred landslides.

Study Area and Data

Study Area

The study area is located in the border region of the Yunnan, Guizhou, and Sichuan Provinces, in addition to

a minimal number of rivers flowing through Chongqing Municipality (Fig. 2). The Chishui River Basin originates in Fangba Township, southwestern Zhenxiong County, Yunnan Province. Its main tributaries are the Erdao, Tongzi, and Xishui Rivers, with the rest spread over many counties and cities, including Zhenxiong

and Weixin in Yunnan Province; Bijie, Dafang, Jinsha, Zunyi, Rhenhuai, Chishui, Xishui, and Tongzi in Guizhou Province; and Xuyong, Gulin, and Hejiang in Sichuan Province. The Chishui River Basin is located at 104°45'–106°51'E, 27°20'–28°50'N. The basin has a total area of 20,440 km². The terrain of the Chishui River Basin is higher in the northeast and is on the slope from the Yunnan–Guizhou Plateau to the Sichuan Basin. In the Chishui River Basin, the Paleozoic stratum dominates the section above Taipingdu and Sandy Beach, with widely distributed limestone interspersed with shale and coal; meanwhile, the Mesozoic stratum, with sandstone, mudstone, conglomerate, and coal, dominates the section below Sandy Beach. The basin experiences southeastern Gulin area mountain tectonics along with regional tectonics. The section of the river below Unk Indu crosses the shield of the mountain tectonic plate, and the inconspicuous spine part is composed mainly of a series of broad and gentle east–west comb folds with fewer ruptures, where regional tectonic plates are single and stable. The Chishui River Basin is adjacent to a plateau and valley, characterized by a continental climate. Most of the basin is dry and cold in winter, and wet and hot in summer, with an average temperature of 15–20°C. The atmosphere in the middle and lower reaches of the basin shows a pattern of hot summers and mild winters. The rainy season is from June to September, and the annual rainfall is 900–1000 mm in the upper part of the Basin and 1000–1500 mm in the middle and lower parts.

Data

Landslide Survey

As the first step in producing a landslide susceptibility zone map, a landslide survey is necessary for modeling [23]. We used the China Disaster Database of the Global Disaster Data Platform as a data source, which is a landslide catalog that includes 724 landslides in the Chishui River Basin, including the latitude and longitude locations, geographic locations, scale size, and area attributes of each landslide site. Fig. 2 illustrates the layout of historical landslide sites in the Chishui River Basin.

Data Sources

Landslide susceptibility analysis predicts the magnitude of the likelihood of landslides in a study area based on historical landslide point data and the geo-environmental background of the site [23–26]. In this landslide susceptibility assessment of the Chishui River Basin, we used elevation data, introductory geology, climate and hydrology, geomorphology and topography, human engineering activities, and historical landslide points to select appropriate factors that induce landslides [27, 28]. The Sentinel-2 data have a resolution of up to 10 m, include 13 bands, and are primarily based on

Sentinel-2 images. Remote sensing image data were extracted from the Normalized Difference Vegetation Index (NDVI) using ENVI software, and road and river data were obtained from the Open Street Map website. Land use types were obtained from LandGlobe30, and fault distances were obtained from a 1:250,000 geological map in the National Geological Database. Digital elevation model (DEM) data were obtained from ASTER GDEM digital elevation data, with a ground resolution of 30 m, and the slope direction, slope gradient, Stream Power Index (SPI), Topographic Wetness Index (TWI), Topographic Position Index (TPI), surface cut, surface roughness, terrain relief, and elevation coefficient of variation were extracted using ArcGIS software. Temperature and precipitation data were obtained from the National Earth System Science Data Center, and lithology, soil, and geomorphology data were obtained from the ISRIC database. The population distribution density was obtained from data from the Seventh National Census. The raster size of each influencing factor was uniformly 30 m.

Research Method

Selection and Analysis of Evaluation Factors

Initial Selection of Evaluation Factors

Landslides can be caused by both natural and human factors. Natural factors include the geological characteristics of the area, the amount of vegetation, temperature, and rainfall. On the other hand, human factors include land use practices such as deforestation, farming, road and river construction, and population density.

Therefore, in the study of landslide susceptibility in the Chishui River Basin, reference was made to historical literature and the analysis of the geological environment characterization of the basin area. Preliminarily, we selected 17 natural factors of landslide susceptibility in the Chishui River Basin area, such as elevation, slope, slope direction, topographic relief, surface roughness, surface incision, coefficient of variation of elevation, landform, lithology, soil, TPI, TWI, SPI, NDVI, air temperature, rainfall, and distance from faults, etc.; in addition to the 4 variables of landslide occurrence due to human behaviors: distance from rivers, distance from roads, land use type, and population density. distance, distance to roads, land use type, and population density.

Analysis of Conditioning Factors

While there is no uniform standard for the number of factors to include, the inclusion of too many factors can lead to an increase in the amount of data and the running time for the evaluation becomes longer; the selection of fewer factors may be inadequate to reveal potential correlations with landslide-triggering factors.

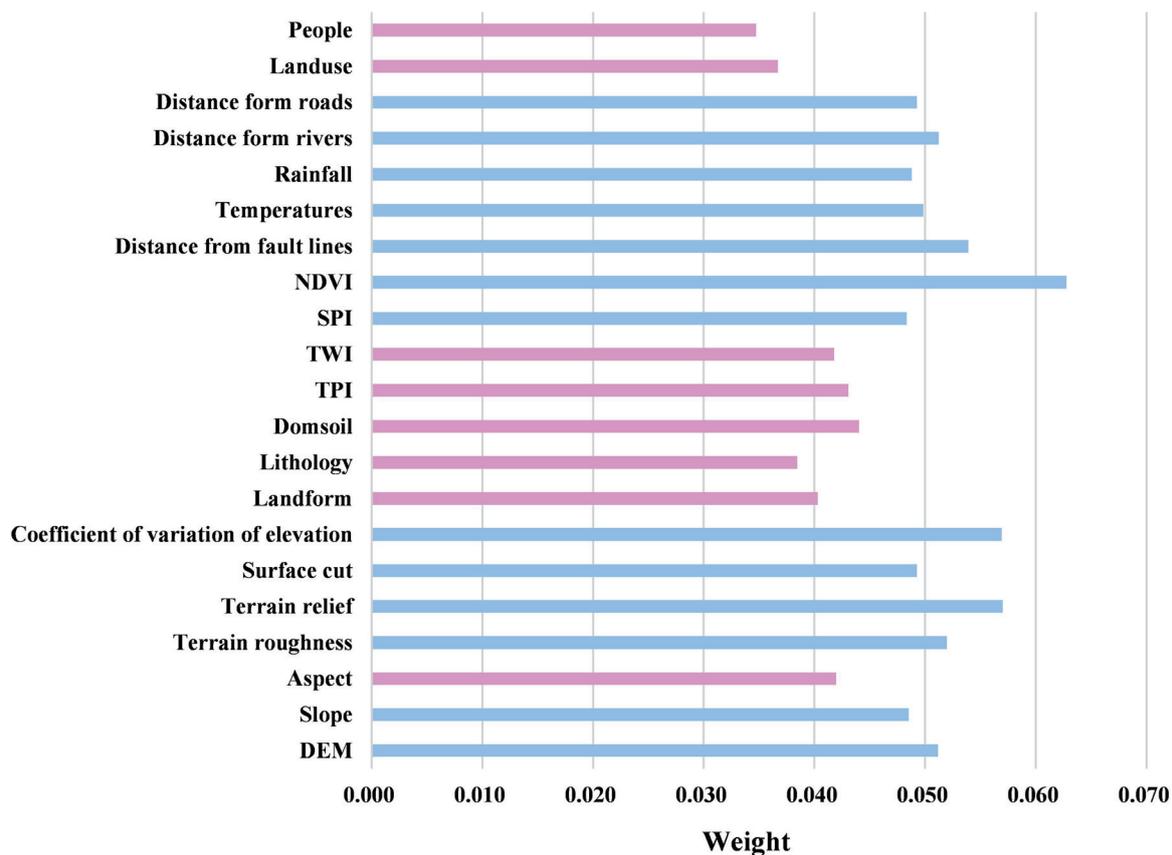


Fig. 3. Landslide impact factor weighting values.

Therefore, it is crucial to conduct a preliminary investigation and analysis of the mechanisms that trigger landslides in the study area. The correct selection of applicable evaluation factors can significantly improve the accuracy of vulnerability evaluations for effective disaster prevention and mitigation in disaster-affected areas.

This study used PCA to screen for landslide conditioning factors in the Chishui River Basin. The attribute data of landslide evaluation factors were extracted based on field investigations and a review of the literature, and a landslide attribute database was established. The attribute database of historical landslides was used as the primary dataset for evaluating landslide conditioning factors. The PCA conducted in SPSS software was used to evaluate the weights and correlations of the factors. The weights of the factors were used to screen appropriate landslide conditioning factors so that the results of landslide susceptibility zoning were as accurate as possible.

The relationships between the weights of the 21 evaluation factors are shown in Fig. 3. The weights of the slope direction, geomorphological type, lithology, soil, TPI, TWI, land use type, and population were 0.042, 0.040, 0.038, 0.044, 0.043, 0.042, 0.037, and 0.035, respectively, with a weight value below 0.045. Therefore, the remaining 13 landslide evaluation factors (DEM; slope; NDVI; temperature; precipitation; distance

from roads, rivers, and faults; SPI; surface cut; surface roughness; terrain relief; and elevation coefficient of variation) were applied to landslide susceptibility evaluation and susceptibility mapping in the Chishui River Basin.

Methods for Selecting Non-Landslide Points

To avoid errors caused by anthropogenic factors when selecting non-landslide points, we used the information volume model to take the historical landslide points as the dataset and used the 13 landslide conditioning factors identified by PCA to calculate the information quantity of each factor (Table 1).

Utilizing the natural breakpoint method through the raster superposition value of the 13 information quantities to divide the landslide susceptibility zones into five grades: very low, low, medium, high, and very high susceptibility (Fig 4). In this study, we used an information volume model to predict landslide susceptibility zones and select 724 non-landslide points in very-low-susceptibility zones.

Neural Network Models

(1) BP model

The BP model is a multi-layer feedforward neural network model similar to human neurons for signaling;

Table 1. Number of rasters, number of landslides, and amount of information for each category of the landslide impact factor.

Factor classification	Evaluation factor	Number of rasters/pcs	Percentage of graded area/%	Number of landslides/pcs	Percentage of landslides/%	Volume of information
161 – 585	DEM(m)	2402819	11.69	133	18.27	0.452
586 – 909		3888609	18.92	216	29.83	0.455
910 – 1179		5684817	27.66	207	28.59	0.033
1180 – 1465		5813374	28.28	133	18.27	-0.432
1466 – 2487		2763515	13.45	35	4.83	-1.023
0 - 9.94	Slope (°)	4150048	20.23	113	15.61	-0.259
9.95 - 17.49		6170762	30.08	227	31.35	0.042
17.5 - 26.06		5441219	26.52	252	34.81	0.272
26.07 - 36.69		3456391	16.85	103	14.23	-0.169
36.7 - 87.45		1298223	6.33	29	4.01	-0.457
-9.21 - -4.26	SPI	1413663	6.89	52	7.18	0.042
-4.25 - 0.68		5293318	25.80	184	25.28	-0.021
0.69 - 3.95		5529353	26.95	212	29.28	0.083
3.96 - 7.31		4522926	22.05	160	22.24	0.009
7.32 - 13.31		3757383	18.31	116	16.02	-0.134
0.29 - 0.6	NDVI	167	0.91	9	1.24	0.317
0.61 - 0.74		1024	5.55	64	8.84	0.465
0.75 - 0.81		4381	23.74	215	29.70	0.224
0.82 - 0.86		5605	30.38	223	30.80	0.014
0.87 - 0.9		7274	39.42	213	29.42	-0.293
0 - 0.028	Coefficient of variation of elevation	8749810	41.88	165	22.79	-0.608
0.029 - 0.059		8324925	39.84	354	48.90	0.205
0.06 - 0.104		3046186	14.58	168	23.20	0.465
0.105 - 0.347		773466	3.70	37	5.11	0.322
0.348 - 0.714		385	0.00	0	0.00	0.000
1	Surface roughness (m)	1028492	5.01	43	5.94	0.170
1.001 - 1.336		19001884	92.62	672	92.82	0.002
1.337 - 5.539		486157	2.37	9	1.24	-0.645
5.54 - 12.767		69	0.00	0	0.00	0.000
12.768 - 22.432		41	0.00	0	0.00	0.000
0.001 - 118.471	Terrain relief	8087071	38.70	176	24.31	-0.465
118.472 - 221.145		9625309	46.07	456	62.98	0.313
221.146 - 789.804		3181918	15.23	92	12.71	-0.181
789.805 - 1619.098		120	0.00	0	0.00	0.000
1619.099 - 2,014		354	0.00	0	0.00	0.000
0 - 45.117	Surface cut	5383391	25.76	141	19.48	-0.280
45.118 - 76.503		6759079	32.35	287	39.64	0.203
76.504 - 111.811		5172204	24.75	204	28.18	0.130
111.812 - 160.851		2756221	13.19	81	11.19	-0.165
160.852 - 500.209		823877	3.94	11	1.52	-0.954

Table 1. Continued.

11.20 – 13.80	Temperatures (°)	3033	12.42	30	4.14	-1.098
13.80 – 14.90		5652	23.15	107	14.78	-0.449
14.90 – 16.10		6371	26.09	179	24.72	-0.054
16.10 – 17.50		5776	23.66	258	35.64	0.410
17.50 – 19.30		3585	14.68	150	20.72	0.344
792 – 815	Rainfall (mm)	3652	14.93	187	25.83	0.548
815 – 829		5056	20.66	144	19.89	-0.038
829 – 842		8436	34.48	243	33.56	-0.027
842 – 858		5133	20.98	137	18.92	-0.103
858 – 889		2190	8.95	13	1.80	-1.606
0	Distance from rivers(m)	7191	21.13	235	32.46	0.429
0.01 - 1036.75		13210	38.82	306	42.27	0.085
1036.76 - 1647.69		7051	20.72	114	15.75	-0.275
1647.7 - 2647.4		5224	15.35	62	8.56	-0.584
2647.41 - 4720.9		1354	3.98	7	0.97	-1.415
0 - 0.006	Distance from roads(m)	20816	55.27	436	60.22	0.086
0.007 - 0.02		10498	27.87	177	24.45	-0.131
0.021 - 0.036		3818	10.14	77	10.64	0.048
0.037 - 0.054		1842	4.89	31	4.28	-0.133
0.055 - 0.091		689	1.83	3	0.41	-1.485
0 - 7200.529	Distance from fault lines(m)	10634073	51.74	392	54.14	0.045
7200.53 - 20292.399		5437198	26.45	226	31.22	0.165
20292.4 - 38621.017		1673234	8.14	39	5.39	-0.413
38621.018 - 57931.525		1561593	7.60	55	7.60	0.000
57931.526 - 83460.672		1247043	6.07	12	1.66	-1.298

it consists of an input layer–implicit layer–output layer structure, which has a remarkable ability to process nonlinear data [29]. The construction of the BP model first requires determining the number of neuron nodes in the input layer (n), the number of neuron nodes in the implicit layer (w_{ij}), and the number of neuron nodes in the output layer (m). Next, the hidden layer threshold (a_j), output layer threshold (b_k), learning rate, number of training iterations, minimum error of training objectives, and excitation function were determined, where the input layer was connected to the hidden layer with weight w_{ij} , and the hidden layer was connected to the output layer with weight w_{jk} .

In this study, a BP neural network with the structure of “13-6-2” was constructed in MATLAB. The number of neuron nodes in the input layer was 13, the number of neuron nodes in the hidden layer was 6, and the number of neuron nodes in the output layer was 2. The excitation function of the hidden layer was tansig, the function of the output layer was pure-line, and the training function was trainlm, with randomly determined w_{ij} , w_{jk} , a_j , b_k .

The number of training times was 1000, the learning rate was 0.01, and the minimum error of the training target was 0.000001.

(2) LSTM model

The LSTM model is a deep learning model optimized for RNN models that effectively overcomes the vanishing gradient problem and achieves long- and short-term information memory [30]. With these advantages, scholars in various countries have begun to gradually apply LSTM models to regional landslide susceptibility predictions, allowing information to flow continuously through the point nodes of the model [31]. The structure of LSTM is shown in Fig 5.

The input gate (i_t) determines whether the new input landslide information features are allowed to be updated and saved to memory, the forget gate (f_t) controls whether the memory remembers or forgets the previous landslide information features, the output gate (O_t) determines whether the landslide feature information is allowed to be output, the cellular state (C_t) stores the long-term landslide displacement information,

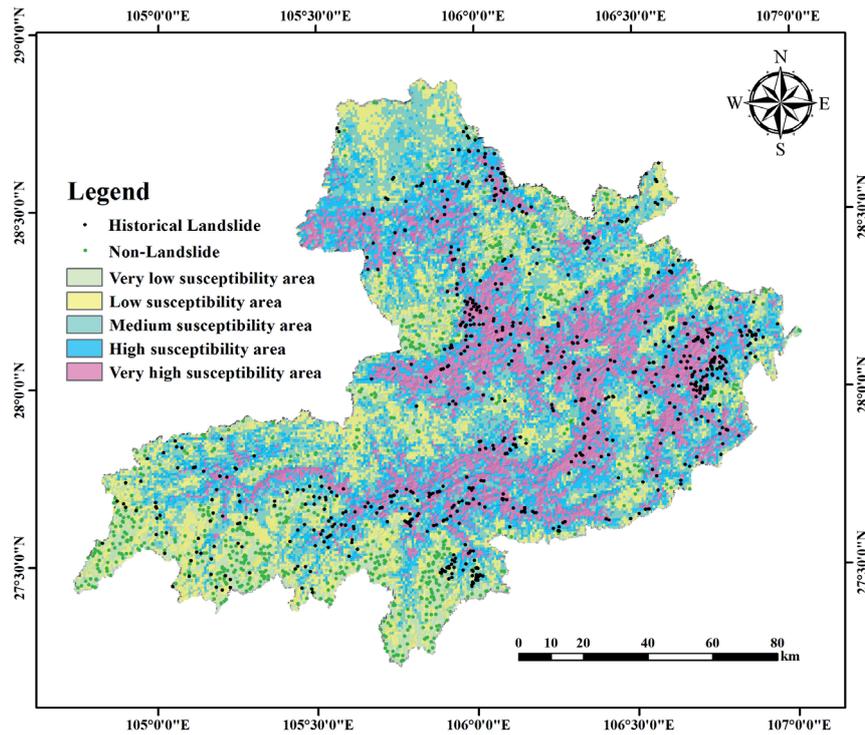


Fig. 4. Landslide susceptibility zoning map based on informativeness modeling.

memristors (h_t) is responsible for storing the short-term landslide displacement information, and the candidate state (\tilde{C}_t) represents the new displacements of landslides summarized by the recurrent neural network [32].

In this research work an LSTM model was constructed using MATLAB software, its specific network structure contains an input layer with 13 dimensions, six hidden units, a ReLu layer, a fully connected layer, a Softmax function layer, and finally a binary classification layer. To optimize the prediction accuracy of this model, the MaxEpochs were set to 1000, and the MiniBatchSize parameter was also added with a value of 100, as well as the Adam gradient descent algorithm, which resulted in a learning rate of 0.0001 after 700 training sessions.

(3) CNN model

The CNN model is a neural network model initially proposed by LeCun for digital recognition. It differs from other models mainly in its convolutional operation, and has been widely used for image extraction and classification processing [33, 34]. CNN mainly comprises feed-forward neural networks consisting of one or more convolutional, pooling, and fully connected layers [35, 36]. Their local perceptions and weight sharing make it easy to optimize and reduce model complexity [34, 37-41]. The convolutional layer extracts the features of the data, and the convolutional kernel performs convolutional computations to obtain the feature layer of the input image. Assuming that the output layer is an $M \times 2D$ image, the convolution formula is:

$$y_{mn} = f\left\{\sum_{j=0}^{j-1} \sum_{i=0}^{i-1} x_{m+i,n+j} w_{ij} + b\right\} \quad (1)$$

In Equation (1), y_{mn} is the output feature map of the convolutional layer, f is the activation function, w is the convolutional kernel of $j \times i$, and b is the bias [34]. Assuming that there are n landslide conditioning factors in the input data and that the convolutional layer filters m convolutional kernels, the layer has $n \cdot m + 1$ feature vectors. Regularization prevents overfitting, which is followed by classification learning using a fully connected layer. The last layer outputs the probability of landslide category a from the normalized exponential function, also known as the softmax function, assuming that there is an array V . V_i denotes the i^{th} element in V , then the softmax value of V_i is

$$S_i = \frac{e^i}{\sum_{j=1}^T e^j}, j = 1, 2, \dots, T \quad (2)$$

In Equation (2), ε is a natural constant and S_i denotes the ratio of index I to the sum of the indices of V_i .

In this study, a CNN was built using MATLAB software, and the specific network model structure was analyzed, which contains a $13 \times 1 \times 1$ input layer, two convolutional layers, two maximal pooling layers, two ReLu layers, a Softmax's function layer, and finally a binary classification layer. In addition, based on the results of multiple training and prediction, MaxEpochs was set to 500. Moreover, to make model prediction more accurate, the parameter MiniBatchSize was added, with a value of 128. To avoid overfitting, the regularization

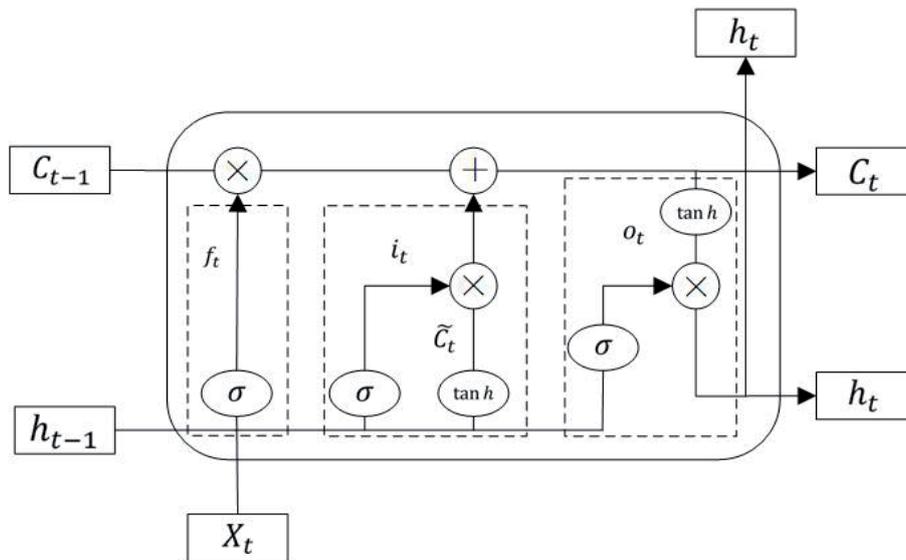


Fig. 5. LSTM model structure.

parameter $1e-04$ was used, and the learning rate was 0.0005.

Results and Discussion

Landslide Susceptibility Zoning Results

The ArcGIS application randomly generated 724 non-landslide points with a proportion of 1:1 to the number of landslide points in the very-low-susceptibility zone under the information quantity model and constructed a point dataset with 724 historical landslide points. The study area was separated into 20604472 rasters, and ArcGIS was utilized for the raster-to-point operation using a tool for multi-value extraction to points in ArcGIS and extracting the values of the 13 landslide conditioning factors for all raster points in the study area. We randomly selected 70% of the dataset as a training set and the remainder as a test set, and imported the divided training set and test set into MATLAB software to complete the training and testing of BP, LSTM, and CNN. The three trained neural

network models predicted the probability of a landslide occurring at all grid points in the study area.

After training the neural network models (i.e., BP, LSTM, and CNN), each raster point resulted in a corresponding landslide susceptibility probability value, and the same natural breakpoint method was used to classify the landslide susceptibility probability value into five categories of susceptibility zones (Fig. 6). According to the susceptibility results of BP, LSTM, and CNN models, the historical landslide point map was “spatially connected” in ArcGIS, and the partitioned area of each susceptibility partition and the distribution of historical landslides were obtained respectively (Table 2). The distribution of historical landslide points in landslide susceptibility zones predicted by different models shows that there are 520 historical landslide points in the very high susceptibility zone predicted by the BP model, and there are 20 historical landslide points in the very low susceptibility zone; there are 599 historical landslide points in the very high susceptibility zone predicted by the LSTM model, and there are 25 historical landslide points in the very low susceptibility zone; the number of historical landslide points distributed in the very

Table 2. The size of landslide-prone areas and the distribution of historical landslide sites in the prone areas under different models.

Susceptibility Class	BP		LSTM		CNN	
	Area (km ²)	Landslide (pcs)	Area (km ²)	Landslide (pcs)	Area (km ²)	Landslide (pcs)
Very low	2524.1382	20	3398.7087	25	9090.6651	16
Low	1876.2858	32	1308.2499	22	879.7734	38
Medium	1837.0503	53	1137.1365	35	668.9259	33
High	2943.0297	99	1541.4921	43	828.3303	33
Very high	9363.5208	520	11158.4376	599	7076.3301	604

high susceptibility zone predicted by the CNN model is the number of historical landslides distributed in the very high susceptibility zone predicted by the CNN model is 604, accounting for 83.43% of the total number of historical landslides, and the number of historical landslides distributed in the very low susceptibility zone predicted by the CNN model is 16, which is only 2.21% of the number of historical landslides. Therefore, the prediction result of the CNN model is closest to the actual situation of the number of landslides in the Chishui River Basin and is effective.

Model Accuracy Evaluation

Model accuracy assessment is crucial for landslide susceptibility evaluation. A confusion matrix is commonly used for the prediction accuracy analysis of classification problems. The accuracy, precision, recall, and F1 scores obtained from this matrix are important quantitative metrics for assessing classification results [42-45]. As shown in Table 5, the evaluation indices of the three neural network models were all greater than 0.9, indicating that the evaluation

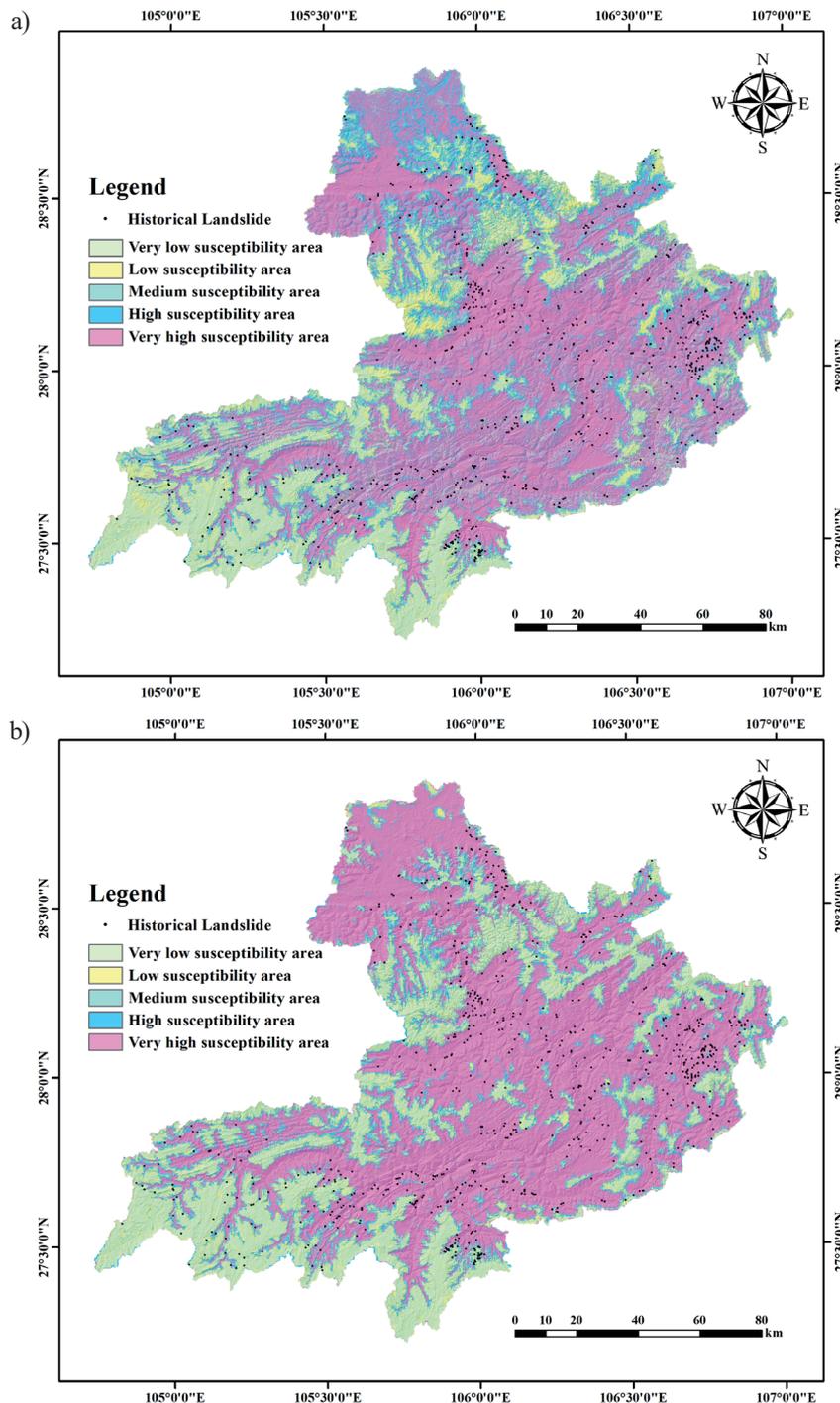


Fig. 6. Landslide susceptibility zoning map under a) BP; b) LSTM; c) CNN models.

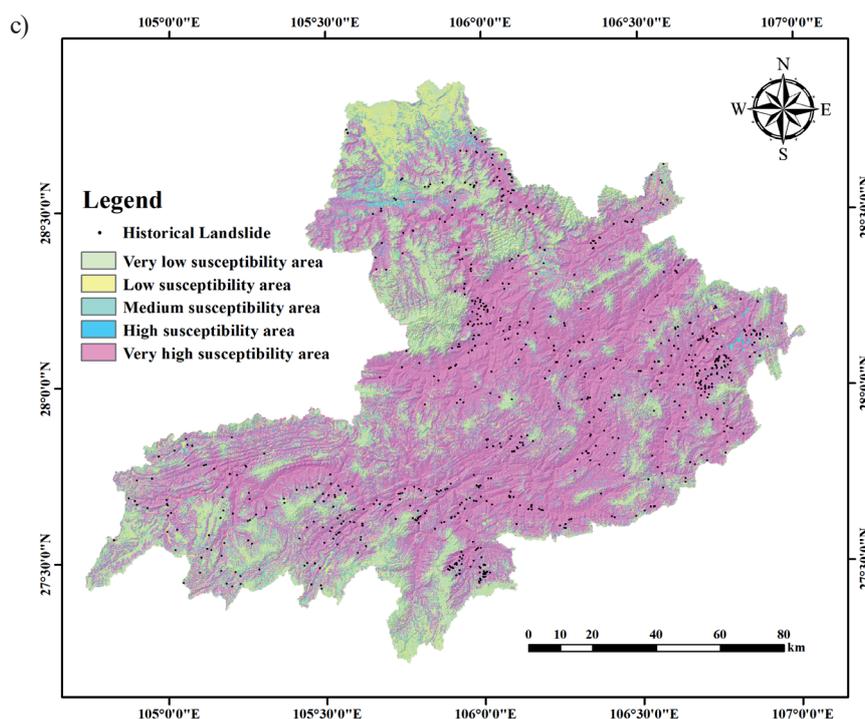


Fig. 6. Landslide susceptibility zoning map under c) CNN models.

results of these three models had a certain degree of credibility.

In landslide susceptibility evaluation, previous authors have often used a receiver operating characteristic (ROC) curve in combination with the first four test indicators to validate the susceptibility zoning results and then evaluated the accuracy of the zoning results by calculating the area under the curve (AUC) value [43-45].

Based on the prediction results obtained from training the three neural network models, SPSS was used to generate the ROC curves. As shown in Fig. 7, the AUC values of BP, LSTM, and CNN were all greater than 0.9. In particular, the AUC value of 0.997 for CNN was significantly greater than that of BP (AUC=0.981) and LSTM (AUC =0.984), indicating that this model had the best fitting and prediction ability for this study area.

Validity of the Information Model for Selecting Non-Landslide Points

In landslide susceptibility assessments, the importance of the correct selection of non-landslide

points cannot be underestimated. In addition to the informative model used in this study to select non-landslide points, most recent studies have relied on randomly selected non-landslide sites in areas where no landslides have occurred [46, 47]. In general, the possibility of landslides occurring in low-slope areas, such as rivers and gullies, is slight [19, 48], so low-slope areas derived from the DEM data can be utilized to randomly select non-landslide sites within the area. A buffer zone was established at a certain distance from the historical landslide point, and non-landslide points were selected outside the buffer zone [49-52]. To verify the reasonableness of selecting non-landslide points with the information volume model in this study, four non-landslide point-selection methods were used: random selection of non-landslide areas [44], random selection of low-slope areas with slopes less than 3°, random selection of landslide sites outside the buffer of 1000 m [51], and selection of very-low-landslide-susceptibility zones with the informative model [52]. The selected non-landslide points were input into the CNN model with the best-known susceptibility zoning effect.

Table 3. Comparison of evaluation indicators of different models.

Model	Accuracy	Precision	Recall	F1_Score
BP	0.922	0.908	0.941	0.924
LSTM	0.933	0.915	0.960	0.937
CNN	0.956	0.941	0.943	0.960

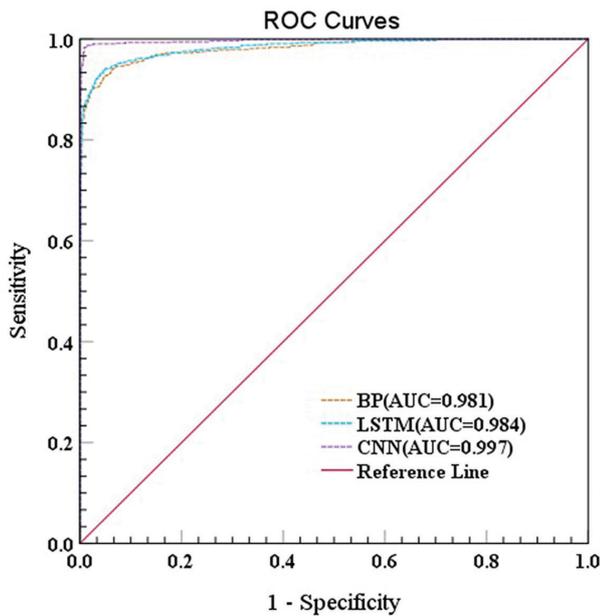


Fig. 7. ROC curves for different model performance analyses.

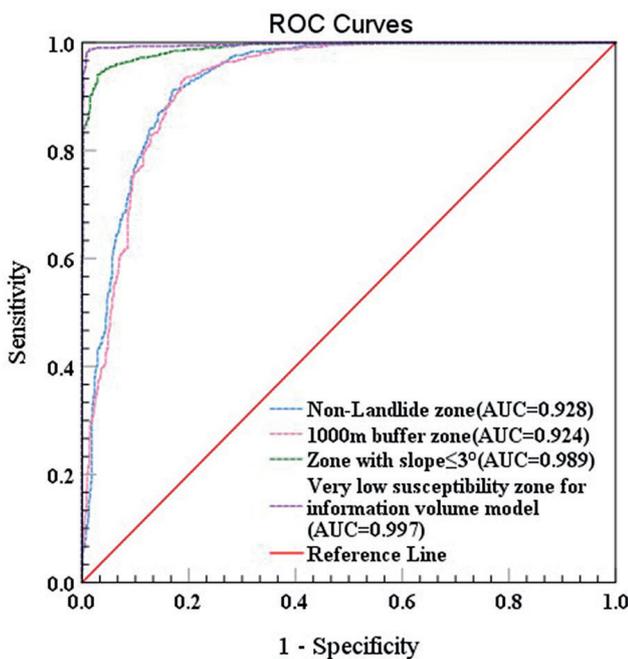


Fig. 8. ROC curves for four methods of selecting non-landslide samples.

Moreover, based on the ROC curves (Fig. 8), the AUC value of selecting non-landslide points in non-landslide areas was 0.928 and that of selecting non-landslide points in buffer zones 1000 m away from historical landslide sites was 0.924. Setting the slope equal to 3° as a threshold, the AUC value of selecting non-landslide points in areas with $\leq 3^\circ$ was 0.989; using the information volume model for the landslide susceptibility zoning, the AUC value of non-landslide

points selected in the very-low-susceptibility zone was 0.997, which was the highest AUC value of those found with the four methods. It is clear, feasible, and practical to analyze the landslide susceptibility of the entire Chishui River Basin by randomly selecting the same number of non-landslide points in the very low susceptibility zones of landslides using the information volume model.

In this study, we discarded the previous traditional models and successfully utilized CNN, LSTM, BP, and other neural network models to predict the landslide-prone zones in the Chishui River Basin. Although these three neural network models are more advanced than traditional machine learning models, they may not be able to solve the difficulty when using a large sample size in future research, because the network structure level is not yet deep. Basic neural networks can easily result in lower efficiency of models when training a large amount of data, so we could optimize the model by combining a basic neural network model with an attention module, which only focuses attention on the region of interest and can disregard the unimportant information, which would significantly improve the efficiency of the whole model.

Conclusions

The Chishui River Basin was considered as the study area in this research. We selected 21 influencing factors related to topography and geomorphology, including the essential geological environment, meteorology and hydrology, anthropogenic activities, and vegetation cover to construct a landslide susceptibility evaluation system after field investigation and data collection. In total, 13 landslide factors were used for training and prediction after PCA and ranking factor weights, non-landslide points were selected based on the information volume model, and BP, LSTM, CNN, and other models were employed to evaluate landslide susceptibility and produce landslide susceptibility zoning maps for the study area. These results led to the following important conclusions:

(1) The results of CNN susceptibility zoning showed that the areas of very-high, high-, medium-, low-, and very-low-susceptibility areas of landslides in the Chishui River Basin were 2034.0180 km², 2838.4950 km², 1251.0520 km², 1338.1200 km², and 11082.3400 km², respectively. The very-high- and high-susceptibility areas of landslides were mainly located in hilly and mountainous regions that were at a short distance from roads, rivers, and faults, and where the air temperature, precipitation, and soil moisture content were suitable.

(2) The accuracy of the evaluation indices of the BP, LSTM, and CNN models was greater than 0.9, with specific AUC values of 0.981, 0.984, and 0.997, respectively. The evaluation accuracy of the CNN model was better, with better landslide prediction ability, than those of the other two models. These

methods showed a significant advantage when dealing with multidimensional factor data from large areas and landslide disaster data.

(3) Of the four methods for selecting non-landslide points in non-landslide areas, 1000-m buffer zones, low-slope zones with slopes below 3°, and very-low-susceptibility zones from the information volume model, the information volume model selected non-slip points as the optimal method with an AUC of 0.997.

(4) The top six landslide conditioning factors were NDVI, elevation coefficient of variation, terrain relief, distance from faults, DEM, and distance from the river, reflecting the extensive landslides in the Chishui River Basin. The complexity and variety of causes were closely related to the geological structure, topography, vegetation cover, meteorology, and hydrology.

In conclusion, this study demonstrated the reliability of CNN in analyzing landslide susceptibility and explored an effective way to select high-quality non-landslide points using an information volume model. The research presented in this study can provide engineers and technicians with a basis for landslide susceptibility evaluations, which can be used to develop appropriate pre-disaster prevention and post-disaster relief programs to reduce the threats posed by existing or future landslides.

Acknowledgments

This work was supported by The National Natural Science Foundation of China (grant number 416620009).

Conflict of Interest

The authors declare no conflict of interest.

References

- ASSILZADEH H., LEVY J.K., WANG X. Landslide Catastrophes and Disaster Risk Reduction: A GIS Framework for Landslide Prevention and Management. *Remote Sensing*, **2** (9), 2259, **2010**.
- CHANG L., ZHANG R., WANG C. Evaluation and Prediction of Landslide Susceptibility in Yichang Section of Yangtze River Basin Based on Integrated Deep Learning Algorithm. *Remote Sensing*, **14** (11), **2022**.
- LIU C.Z., CHEN C.L. Achievements and countermeasures in risk reduction of geological disasters in China. *Journal of Engineering Geology*, **28** (2), 9, **2020**.
- YAN L., XU W., WANG H., WANG R., MENG Q., YU J., XIE W.-C. Drainage controls on the Donglingxing landslide (China) induced by rainfall and fluctuation in reservoir water levels. *Landslides*, **16** (8), 1583, **2019**.
- YIN Y.P. Preliminary study on the mitigation strategy of geological disasters in China. *The Chinese Journal of Geological Hazard and Control*, **015** (002), 1, **2004**.
- GORDO C., ZEZEZE J.L., MARQUES R. Landslide Susceptibility Assessment at the Basin Scale for Rainfall- and Earthquake-Triggered Shallow Slides. *Geosciences*, **9** (6), **2019**.
- JONES S., KASTHURBA A.K., BHAGYANATHAN A., BINOY B.V. Impact of anthropogenic activities on landslide occurrences in southwest India: An investigation using spatial models. *Journal of Earth System Science*, **130** (2), **2021**.
- DAI F.C., LEE C.F., NGAI Y.Y. Landslide risk assessment and management: an overview. *Engineering Geology*, **64** (1), 65, **2002**.
- DANDRIDGE C., STANLEY T., KIRSCHBAUM D., AMATYA P., LAKSHMI V. The influence of land use and land cover change on landslide susceptibility in the Lower Mekong River Basin. *Natural Hazards*, **115** (2), 1499, **2023**.
- JABOYEDOFF M., MICHOD C., DERRON M., VOUMARD J., LEIBUNDGUT G., SUDMEIER-RIEUX K., NADIM F., LEROI E.J.L. Engineered Slopes. Experience Theory And Practice, CRC Press, USA Human-induced landslides: toward the analysis of anthropogenic changes of the slope environment, 217, **2018**.
- ERCANOGLU M. An overview on the landslide susceptibility assessment techniques. *Environment Geoscience*, **2008**.
- AZARAFZA M., AZARAFZA M., AKGÜN H., ATKINSON P.M., DERAKHSHANI R.J.S.R. Deep learning-based landslide susceptibility mapping, **11** (1), 24112, **2021**.
- SHAHRI A.A., SPROSS J., JOHANSSON F., LARSSON S. Landslide susceptibility hazard map in southwest Sweden using artificial neural network. *Catena*, **183**, **2019**.
- ZHANG H., YIN C., WANG S., GUO B. Landslide susceptibility mapping based on landslide classification and improved convolutional neural networks. *Natural Hazards*, **116** (10), **2022**.
- GHOORBANZADEH O., SHAHABI H., CRIVELLARI A., HOMAYOUNI S., BLASCHKE T., GHAMISI P. Landslide detection using deep learning and object-based image analysis. *Landslides*, **19** (4), 929, **2022**.
- LIU M., LIU J., XU S., ZHOU T., MA Y., ZHANG F., KONECNY M. Landslide susceptibility mapping with the fusion of multi-feature SVM model based FCM sampling strategy: A case study from Shaanxi Province. *International Journal of Image and Data Fusion*, **12** (4), 349, **2021**.
- ZHOU C., YIN K., CAO Y., AHMED B., LI Y., CATANI F., POURGHASEMI H.R. Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. *Computers & Geosciences*, **112**, 23, **2018**.
- DUAN Y., TANG J., LIU Y., GAO X., DUAN Y. Spatial sensitivity evaluation of loess landslide in Liulin County, Shanxi based on sandom forest. *Scientia Geographica Sinica*, **42** (2), 343, **2022**.
- KAVZOGLU T., SAHIN E.K., COLKESEN I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, **11** (3), 425, **2014**.
- HUANG F., PAN L., YAO C. Landslide susceptibility prediction modelling based on semi-supervised machine learning. *Journal of Zhejiang University*, **55** (9), 1705, **2021**.
- MAVROULIS S., DIAKAKIS M., KRANIS H., VASSILAKIS E., KAPETANIDIS V., SPINGOS I., KAVIRIS G., SKOURTSOS E., VOULGARIS N., LEKKAS E. Inventory of historical and recent

- earthquake-triggered landslides and assessment of related susceptibility by GIS-based analytic hierarchy process: the case of Cephalonia (Ionian Islands, Western Greece). *Applied Sciences*, **12** (6), 2895, **2022**.
22. PETROVA E. Natural hazard impacts on transport infrastructure in Russia. *Natural Hazards and Earth System Sciences*, **20** (7), 1969, **2020**.
 23. VIET DU Q.V., NGUYEN H.D., PHAM V.T., NGUYEN C.H., NGUYEN Q.-H., BUI Q.-T., DOAN T.T., TRAN A.T., PETRISOR A.-I.J.G.I. Deep learning to assess the effects of land use/land cover and climate change on landslide susceptibility in the Tra Khuc river basin of Vietnam, 2172218, **2023**.
 24. LEE S., PRADHAN B. Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. *Journal of Earth System Science*, **115**, 661, **2006**.
 25. POURGHASEMI H.R., GAYEN A., PARK S., LEE C.-W., LEE S. Assessment of Landslide-Prone Areas and Their Zonation Using Logistic Regression, LogitBoost, and NaiveBayes Machine-Learning Algorithms. *Sustainability*, **10** (10), **2018**.
 26. ZHANG H., SONG Y., XU S., HE Y., LI Z., YU X., LIANG Y., WU W., WANG Y. Combining a class-weighted algorithm and machine learning models in landslide susceptibility mapping: A case study of Wanzhou section of the Three Gorges Reservoir, China. *Computers & Geosciences*, **158**, **2022**.
 27. KHABIRI S., CRAWFORD M.M., KOCH H.J., HANEBERG W.C., ZHU Y. An Assessment of Negative Samples and Model Structures in Landslide Susceptibility Characterization Based on Bayesian Network Models. *Remote Sensing*, **15** (12), **2023**.
 28. SHAHABI H., AHMADI R., ALIZADEH M., HASHIM M., AL-ANSARI N., SHIRZADI A., WOLF I.D., ARIFFIN E.H.J.R.S. Landslide Susceptibility Mapping in a Mountainous Area Using Machine Learning Algorithms, **15** (12), 3112, **2023**.
 29. TANG R., YAN E., CAI J. Back analysis of initial ground stress based on back-propagating neural network. *Electronic Journal of Geotechnical Engineering*, **18**, 5839, **2013**.
 30. YU Y., SI X., HU C., ZHANG J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, **31** (7), 1235, **2019**.
 31. XIE P., ZHOU A., CHAI B.J.I.A. The application of long short-term memory (LSTM) method on displacement prediction of multifactor-induced landslides, **7**, 54305, **2019**.
 32. GREFF K., SRIVASTAVA R.K., KOUTNIK J., STEUNEBRINK B.R., SCHMIDHUBER J. LSTM: A Search Space Odyssey. *Ieee Transactions on Neural Networks and Learning Systems*, **28** (10), 2222, **2017**.
 33. HUANG F., ZHANG J., ZHOU C., WANG Y., HUANG J., ZHU L. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides*, **17** (1), 217, **2020**.
 34. WANG Y., FANG Z., HONG H. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. *Science of the Total Environment*, **666**, 975, **2019**.
 35. LEE S., BAEK W.-K., JUNG H.-S., LEE S. Susceptibility Mapping on Urban Landslides Using Deep Learning Approaches in Mt. Umyeon. *Applied Sciences-Basel*, **10** (22), **2020**.
 36. QIN Z., ZHOU X., LI M., TONG Y., LUO H. Landslide Susceptibility Mapping Based on Resampling Method and FR-CNN: A Case Study of Changdu. *Land*, **12** (6), **2023**.
 37. GHORBANZADEH O., BLASCHKE T., GHOLAMNIA K., MEENA S.R., TIEDE D., ARYAL J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sensing*, **11** (2), **2019**.
 38. GIRSHICK R. Fast r-cnn, **2015**.
 39. LECUN Y., KAVUKCUOGLU K., FARABET C. Convolutional networks and applications in vision. *IEEE*, **2010**.
 40. SAHA S., ROY J., HEMBRAM T.K., PRADHAN B., DIKSHIT A., ABDUL MAULUD K.N., ALAMRI A.M. Comparison between Deep Learning and Tree-Based Machine Learning Approaches for Landslide Susceptibility Mapping. *Water*, **13** (19), **2021**.
 41. ZHANG G., WANG M., LIU K. Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. *International Journal of Disaster Risk Science*, **10** (3), 386, **2019**.
 42. NIKOObAKHT S., AZARAFZA M., AKGÜN H., DERAKHSHANI R.J.A.S. Landslide susceptibility assessment by using convolutional neural network, **12** (12), 5992, **2022**.
 43. BEHESHTIFAR S. Identification of landslide-prone zones using a GIS-based multi-criteria decision analysis and region-growing algorithm in uncertain conditions. *Natural Hazards*, **115** (2), 1475, **2023**.
 44. FEIZIZADEH B., ROODPOSHTI M.S., JANKOWSKI P., BLASCHKE T.J.C., GEOSCIENCES A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping, **73**, 208, **2014**.
 45. GORSEVSKI P.V., BROWN M.K., PANTER K., ONASCH C.M., SIMIC A., SNYDER J.J.L. Landslide detection and susceptibility mapping using LiDAR and an artificial neural network approach: a case study in the Cuyahoga Valley National Park, Ohio, **13**, 467, **2016**.
 46. CHEN W., ZHANG S.J.C. GIS-based comparative study of Bayes network, Hoeffding tree and logistic model tree for landslide susceptibility modeling, **203**, 105344, **2021**.
 47. IMTIAZ I., UMAR M., LATIF M., AHMED R., AZAM M.J.E.E.S. Landslide susceptibility mapping: improvements in variable weights estimation through machine learning algorithms – a case study of upper Indus River Basin, Pakistan, **81** (4), 112, **2022**.
 48. CHOI J., OH H.-J., WON J.-S., LEE S.J.E.E.S. Validation of an artificial neural network model for landslide susceptibility mapping, **60**, 473, **2010**.
 49. DI NAPOLI M., CAROTENUTO F., CEVASCO A., CONFUORTO P., DI MARTIRE D., FIRPO M., PEPE G., RASO E., CALCATERRA D. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides*, **17** (8), 1897, **2020**.
 50. FAWCETT T. An introduction to ROC analysis. *Pattern Recognition Letters*, **27** (8), 861, **2006**.
 51. MIAO Y., ZHU A., YANG L., BAI S., LIU J., DENG Y.J.M.R.D. Sensitivity of BCS for sampling landslide absence data in landslide susceptibility assessment, **34**, 432, **2016**.
 52. ZHOU X., HUANG F., WU W., ZHOU C., ZENG S., PAN L. Regional Landslide Susceptibility Prediction Based on Negative Sample Selected by Coupling Information Value Method. *Advanced Engineering Science/Gongcheng Kexue Yu Jishu*, **54** (3), **2022**.