https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

Research on landslide master control factor identification and susceptibility prediction modelling

Yin Xing[1], Yang Chen[2], Peng Wang[1], Saipeng Huang[3], Yunfei Xiang[4]

[1] School of Geography Science and Geomatics Engineering, Suzhou University of Science and Technology, Suzhou 215009, China; xyin0320@163.com; pwang@mail.usts.edu.cn
[2] School of Information Technology, Suzhou Institute of Trade & Commerce, Suzhou 215009, China; hntj_cy@163.com
[3] Key Laboratory of Continental Shale Hydrocarbon Accumulation and Efficient Development, Ministry of Education, Northeast Petroleum University, Daqing, 163318, China; huangspcugb@hotmail.com
[4] College of Civil Engineering, Nanjing Forestry University, Nanjing 210037, China; yfxiang@njfu.edu.cn

Correspondence to: Yin Xing (xyin0320@163.com)

Abstract. It is important to properly identify the primary control elements of landslide susceptibility because the modelling process and its uncertainties differ between machine learning predictions of susceptibility to landslides. In response to the aforementioned issues, the novel "weight mean method" is suggested to determine more precise landslide master factors. Support vector machine (SVM) and random forest (RF) are used as examples to discuss the prediction of landslide susceptibility and its uncertainty based on machine learning. For Ruijin City, Jiangxi Province, the landslide inventory and 12 different types of underlying environmental factors were acquired, and the factor frequency ratios were employed as input variables for SVM and RF. The landslides and randomly selected non-landslide samples were then divided into training and test sets, and the trained machine learning was used to predict and map landslide susceptibility. In order to assess modeling uncertainty and determine the landslide master control factor, subject work curves, means, and standard deviations were used. The results show that: (1) Machine learning can effectively predict regional landslide susceptibility, the accuracy of landslide susceptibility predicted by RF is higher than that of SVM, while its uncertainty is lower than that of SVM, but the overall susceptibility distribution patterns of both are similar. (2) The weight-mean approach determines that the slope, height, and lithology, in that order, are the primary controlling elements of the landslide in Ruijin City. In comparison to other machine learning models, the case studies and literature study demonstrate how dependable and susceptible the RF model is.

Keywords. landslide susceptibility prediction; uncertainty analysis; principal control factor identification; machine learning

1 Introduction

The intricate physical geography and geological formations of China produce the underlying environmental conditions for landslide development, and a variety of harmful engineering practices and wet weather conditions act as triggers (Chang et al., 2020; Huang et al., 2020). For the study of regional landslide spatial probabilities and their correlation with underlying environmental conditions, landslide susceptibility prediction modelling is crucial (Jiang et al., 2017; Huang et al., 2020; Steger et al., 2021; Yin et al., 2021). Landslide inventory, acquisition of underlying environmental factors, selection of susceptibility

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

prediction models, determination of model parameters, and analysis of susceptibility modelling results are just a few of the many steps involved in determining a slide's susceptibility (Lee et al., 2017; Huang et al., 2021). After gathering precise data, such as landslide inventories and their underlying environmental conditions, the study of the modelling processes shows that a major modelling uncertainty influences the choice of the susceptibility prediction model (Di Napoli et al., 2020).

The mapping link between landslides and the underlying environmental elements can be completely explored by a good susceptibility prediction model, and it may be used to build a non-linear function that connects the underlying environmental causes to the geographical probability of landslides. Some of the less trustworthy models either struggle to adequately expose the non-linear mapping relationships between them or do not fully exploit known landslide or factor information (Scherzer et al., 2019; Smith et al., 2021). Based on their creation and modelling methodologies, the generally used susceptibility models can be divided into heuristic models, conventional mathematical and statistical models, and machine learning models (Huang et al., 2020). Studies have demonstrated that heuristic models, like hierarchical analysis, require a high level of quantity and quality of historical landslide data, and that the uncertainty in determining how much weight to give to environmental elements is largely influenced by personal experience (Mallick et al., 2018). The size and grading of raster image values might affect traditional mathematical and statistical models like informativeness and multiple linear regression, making it challenging to depict the non-linear relationship between landslides and the underlying environmental causes (Sameen et al., 2020). In spite of the 'black box' modelling and analysis issue, the susceptibility modelling method is very easy and effective. Machine learning models are substantially more effective in predicting susceptibility than heuristic and conventional mathematical and statistical models (Merghadi et al., 2018).

The literature lists numerous machine learning models for predicting the likelihood of landslides, including BP neural networks, multilayer perceptrons, fuzzy mathematics, support vector machines (SVM), decision trees, random forests (RF), various types of integrated models (Chen et al., 2018; Buia et al., 2020; Hong et al., 2020; Talukdar et al., 2020; Youssef et al., 2020), and the recently developed deep learning models. To facilitate susceptibility prediction modelling in other domains, it is imperative to thoroughly separate out and expand the modelling process of machine learning models for forecasting landslide susceptibility. The characteristics of the susceptibility index distributions of various machine learning models can differ significantly even when there is a small difference in prediction accuracy, making it difficult to come to a consensus on which machine learning model performs better at modelling susceptibility (De Fauw et al., 2018; Min et al., 2019; Manibardo et al., 2022).

Therefore, in research regions with high landslide impact, it is crucial to conduct comparative investigations of machine learning models predicting landslide susceptibility. When comparing machine learning models for predicting landslide susceptibility, for instance, Zhang et al. (Zhang et al., 2021) and Huang et al. (Huang et al., 2017) found that SVM models offer the advantage of more consistent prediction outcomes and greater acceptability. Decision trees and RF models, according to Naemitabar et al. (Naemitabar et al., 2021) and Dou et al. (Dou et al., 2019), demonstrate notable advantages over other machine learning models for modelling landslide susceptibility. Even RF models demonstrate advantages over recently developed deep learning models in terms of high modelling efficiency and more reliable prediction accuracy. In order to

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

effectively validate the susceptibility prediction performance of classical machine learning models and to generalise these machine learning models, this study proposes to carry out a landslide susceptibility prediction study using SVM and RF as examples.

The selection of landslide samples and the underlying environmental parameters are intimately related to the modelling performance of machine learning models like SVM and RF. In order to determine the basic environmental factors necessary for landslide susceptibility modelling in the study area, the study was conducted by analyzing the development pattern of landslides in the study area, based on the ease of obtaining the basic environmental factors in the study area, and referring to the existing literature on the selection of factors for study areas with similar environmental backgrounds (Zhang et al., 2018). It is crucial to separate the primary controls for landslide development from the numerous underlying environmental factors after training and testing the SVM and RF models with the aforementioned landslide samples and underlying environmental factors. At this time, determining each basic environmental factor's contribution to the prediction of landslide susceptibility is the main method used to determine the primary controlling factors of landslides. Typically, the factors with the highest contribution rankings are selected as the primary controlling factors of landslides (Luo et al., 2018). The dominant factor of landslides in a given study area should be identified at a macro level according to theory, but in practice different machine learning models calculate the dominant factor of landslides differently, making it difficult to identify the dominant factor of landslides in a specific study area. As can be shown, the major control factor for landslides is calculated with a lot of uncertainty in the previous studies. This study suggests a novel "weighted average method" to address the issue by giving the landslide dominant factors calculated by various machine learning models weights before determining the average value to minimize the differences in the contributions of the factors calculated by various models. Due to the varied geography, loose soil slopes, and seasonal rains in Ruijin City, Jiangxi Province, numerous mounded landslides have formed, causing significant economic losses. Ruijin city is used in this work as an example to conduct research on landslide susceptibility modelling. In order to provide useful theoretical guidance for landslide susceptibility modelling in other regions, the uncertainty characteristics of various machine learning models for predicting landslide susceptibility are first analyzed using SVM and RF as examples, and the main control factors affecting landslide development in the underlying environmental factors are further calculated.

2 Research Methodology

2.1 Research ideas

In order to provide a more accurate mapping of landslide susceptibility in the research area, the prediction performance of the SVM and RF models was compared and examined in this work. (1) Determine each basic environmental factor's frequency ratio. Based on data from landslide inventories and the geological environment, 12 basic environmental variables were chosen; the frequency ratio approach was used to determine each basic environmental factor's frequency ratio; (2) Dividing the training and test sets: The 10-class factor frequency ratios, slippery slopes, and non-slippery slopes (marked as 1 and 0), were used as input variables for machine learning, and they were randomly split into training and test sets in the ratio of 70% and 30%;(3)

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

Creating and testing SVM and RF models utilizing training and test sets to establish the machine learning parameters; (4) Landslide susceptibility mapping: The research area's susceptibility to landslides was mapped in ArcGIS 10.3 and categorized into five susceptibility levels using the trained SVM and RF models; (5) Analyzing the outcomes of the forecast and determining the primary controlling variables: the subject operation characteristic curve (ROC) and additional The effectiveness of the susceptibility modelling was assessed using a range of uncertainty indicators, and the weighted mean technique was then utilized to determine the primary landslide control elements in the research area (Figure 1).

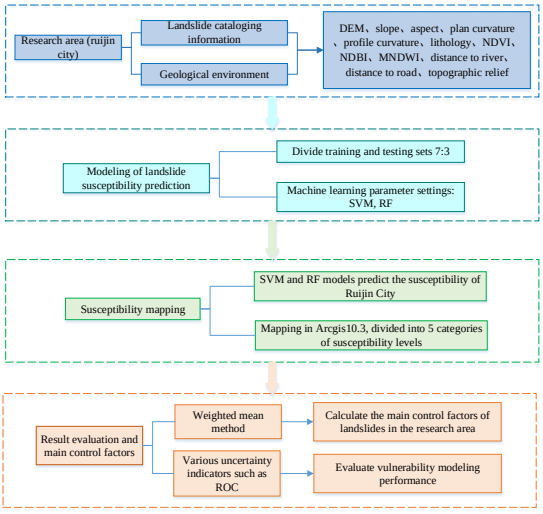


Figure 1. Selecting landslide impact factors, dividing the training and testing data into 7:3, and predicting landslide susceptibility maps using SVM and RF models.

2.2 Frequency ratio analysis of environmental factors

The frequency ratio technique captures the quantitative statistics of the linkage between the various attribute intervals of the factors on landslide susceptibility and enables for proper handling of the non-linear response relationship between landslides and their underlying environmental causes (Rodrigues et al., 2021). In this study, the number of rasters and landslide rasters contained within each environmental factor interval was found, and the total number of rasters and landslide rasters in the study area were calculated. The underlying environmental factors were divided into eight factor attribute intervals using the natural discontinuity method. Equation (1)'s non-linear relationship between environmental element qualities and landslide susceptibility can be more accurately represented by the frequency ratio technique.

$$FR = \frac{A / A'}{B / B'} \quad (1)$$

where: A is the number of landslide rasters occurring in the interval for each type of environmental factor; A' is the total number of landslide rasters in the area; B is the number of rasters in the interval in which the environmental factor is located; B' is the total number of rasters in the study area; and FR indicates the frequency ratio of that type of environmental factor.

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

2.3 Machine learning models

As illustrations of the modelling aspects of machine learning models to forecast landslide susceptibility, SVM and RF models are suggested.

2.3.1 Support vector machines

The SVM is built on a kernel function mapping to enhance the independent variables' dimensional properties and locate a hyperplane that maximizes the class spacing, resulting in linear differentiation of the output variables (Cervantes et al., 2020). First, assume that a set of data (x_i, y_i), $i = 1, 2, ..., n$, and $f(x) = \omega \bullet x + b$ are fitted by a linear regression function to determine ω and b. A slack variable ε is used to control for classification error, and the corresponding linear function is fitted as .

$$\begin{cases} y_i - f(x_i) \le \varepsilon + \xi_i \\ f(x_i) - y_i \le \varepsilon + \xi_i^*, i = 1, 2, ..., n \\ \xi_i, \xi_i^* \ge 0 \end{cases} \quad (2)$$

where: ξ_i, ξ_i^* are classification error factors respectively.

When ξ_i, ξ_i^* is greater than 0, it means that there is a classification error. At this point transformed to solve the minimization function problem, as shown in equation (3), where the constant C>0 is the degree of misclassification beyond the classification error ε. The linear fit function is then substituted into the Lagrange function as shown in equation (4).

$$R(\omega, \xi, \xi^*) = \frac{1}{2}\omega \cdot \omega + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (3)$$

$$f(x) = \omega \cdot x + b = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i \cdot x + b \quad (4)$$

where: α_i, α_i^* are the support vector coefficients.

2.3.2 Random forests

Using the vote results from each tree in numerous decision trees, random forest is mostly used to determine the best classification outcome (Li et al., 2018). Due to the method it is put back and the dataset created by randomly obtaining data features, each decision tree has more detailed information on the input variables. The integration of several decision trees is used to achieve resilience and prevent overfitting of the model. The capacity to provide the Gini index of the corresponding input variables, which is the ranking of the relevance of each input variable, is the key characteristic of a random forest. The RF tree is optimally segmented using an impurity measure, and the importance of the underlying environmental factor is calculated by the reduction in the Gini index of environmental factor k at the time of node segmentation D_{Gk}, which involves the calculation of the average Gini reduction as a percentage of the sum of the average Gini reductions of all underlying environmental factors, as in equation (5).

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

$$P_k = \frac{\sum_{h=1}^{n}\sum_{j=1}^{t} D_{Gkhj}}{\sum_{k=1}^{m}\sum_{h=1}^{n}\sum_{j=1}^{t} D_{Gkhj}} \quad (5)$$

where: m, n, t are the total number of underlying environmental factors, the number of classification trees and the number of individual tree nodes, respectively; D_{Gkhj} is the Gini index reduction of the k th factor at the j th node of the h th tree; P_k is the importance of the k th underlying environmental factor.

2.4 Research ideas

2.4.1 ROC accuracy evaluation

The sensitivity and specificity laws of the susceptibility prediction modelling are calculated using equations (6) and (7), and the ROC (Bui et al., 2019) curve is plotted with sensitivity (true positive rate, or TPR), as the vertical coordinate, and 1-specificity, or false positive rate, as the horizontal coordinate. The area under the ROC curve's (AUC) size indicates how accurately the model can predict outcomes.

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

Where (true positive) represents the total number of points correctly classified as landslides; (true negative) represents the total number of points correctly classified as non-landslides; (false negative) represents the number of points incorrectly classified as landslides; and (false positive) represents the number of points incorrectly classified as non-landslides.

2.4.2 Characteristics of the distribution of the susceptibility index

The standard deviation measures how widely distributed the landslide susceptibility index is, whereas the mean value indicates the average level of the regional landslide susceptibility index distribution (Tsangaratos et al., 2017). In order to reveal the susceptibility prediction performance under SVM and RF models and to obtain a prediction model with some objectivity through comparative analysis, the mean and standard deviation were used in this study to analyze the characteristics of the distribution of susceptibility index values as a whole. In general, it shows that the machine learning model has relatively less uncertainty in modelling the susceptibility prediction when the mean value is modest and the standard deviation is significant.

2.5 Weighted mean method of calculating landslide principal control factors

The importance of different types of environmental factors in landslide susceptibility modelling is determined by a number of factors, such as the characteristics of the factors themselves, the combination of factors, the machine learning model and the

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

scale of the study area, etc. These factors lead to large uncertainties in the identification of landslide master factors. In this study, the "weight mean method" is innovatively proposed to calculate the principal control factors under different machine learning model conditions: (1) Collecting management environment factors and constructing spatial datasets to predict landslide susceptibility using various types of machine learning models, respectively; (2) Compare the percentage importance ranking of each environmental factor in different machine learning models; (3) Assign weights according to the model AUC accuracy value, the higher the accuracy, the larger the corresponding weight value, multiply the weight value by the standardized importance of each factor and then add them together to obtain the final factor percentage importance, as in equation (8), where f_i is the percentage importance of an environmental factor, ω_i is the weight ratio of the accuracy of a model to predict landslide susceptibility, and W_i is the weighted percentage importance of a factor; (4) Conduct uncertainty analysis on the final factor importance ranking results under each working condition and obtain the landslide master control factor.

$$W_i = \sum_{i=1}^{n} \omega_i f_i \tag{8}$$

3 Information on the study area and its landslides

3.1 Introduction to Ruijin City

Ruijin is located in the southeastern part of Jiangxi Province, and the terrain in its territory slopes from northwest to southeast, with a high surrounding and a low middle. Its area is about 2441.2km^2 and its elevation is 139~1117m (Figure 2). The stratigraphic lithology of Ruijin is mainly composed of metamorphic rocks, carbonate rocks and clastic rocks. The southeastern part of the county is hilly and river basin, while the northeast, northwest and southwest are divided into mountainous areas. Induced by the complex natural environment, geological conditions, strong seasonal rainfall and slope excavation, more mounded landslides have developed in Ruijin. The large longitudinal slopes of the riverbeds and the deep and heavy river valleys in the area are important factors in the development of landslides. According to statistics, 370 landslides exist in Ruijin City, and the volume of landslides is mainly small to medium. The average area of the landslide and its affected area is about 13,000m2; the landslide body is mainly Quaternary powdered clay with debris, and the movement mode is mainly downward sliding of the slope as a whole; most of these landslides are located in the surrounding areas of dense residential areas, along roads or gullies, etc (Dongming et al., 2017).

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

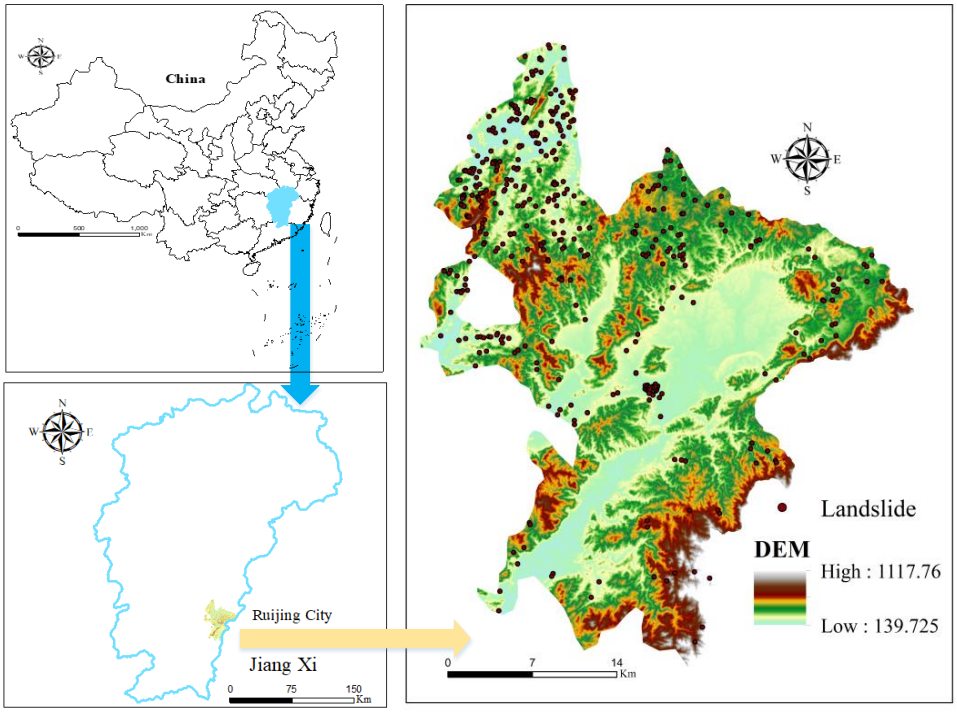


Figure 2. Ruijin City is located in the southeast of Jiangxi Province, in the eastern part of Ganzhou City. This article collected 370 landslides in Ruijin City and marked their locations on the map of Ruijin City.

3.2 System of environmental factors underlying landslide susceptibility

3.2.1 Data sources

The study area's basic data sources include: (1) historical landslide catalogues and related information from field research; (2) a digital elevation model (DEM), which is derived from the basic data of the geographic data cloud that is freely and publicly accessible and is primarily used for the acquisition of topographic and geomorphological basic environmental factors like elevation, slope, and slope direction; (3) A geological map of the study region at a scale of 1:100,000 (3) This geological map of the study area, used to collect stratigraphic lithology; (4) The normalized difference vegetation index (NDVI), normalized difference building index (NDBI), and normalized difference water body index (NDWBI) are all calculated using Landsat-8 remote sensing photos. Additionally, the prediction unit for landslide susceptibility in this investigation was a raster with a cell resolution of 30 m. In addition, the prediction cell for landslide susceptibility was selected as a raster with a cell resolution of 30 m. Using the "surface to raster" function in ArcGIS10.2 software, 370 landslide surfaces were transformed into 863383 raster cells, and the research area was divided into 4938 columns and 6600 rows, for a total of 2395696 raster cells.

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

3.2.2 Basic environmental factors and frequency ratio analysis

The reliability and accuracy of the evaluation results are influenced by the effective selection of the underlying environmental elements. The objective existence, importance, and inheritance of environmental influences are the major selection criteria (Ananthakrishnan et al., 2018). This study chose 12 categories of landslide base environmental parameters, including elevation, slope, slope direction, lithology, NDVI, NDBI, and MNDWI (Figure 3), based on these concepts. Using the natural interruption point method, the study's basic environmental factors were divided into six attribute intervals (stratigraphic lithology was divided into intervals based on stratigraphic combination), and the frequency ratio of each attribute interval was calculated (Table 1). Whether or not the interval is favourable for landslide development is determined by the frequency ratio (Achour et al., 2019).

(1) Topographic and geomorphological factors

Landslides are more likely to occur when the height is less than 335m, as indicated in Table 1 and Figure 3-a. Landslides are more likely to occur when the slope is between 8.8 °and 17.9 °or the slope direction is between 67.5 °and 245.5 °, which are both retrieved using DEM data. According to Table 1 and Figure 3-d, the definition of plane curvature is the slope direction, which reflects all ridgelines and valley lines on the surface in a horizontal direction. Profile curvature is characterized by the degree of change of the slope in the vertical direction and is defined as the slope of the slope (Table 1, Figure 3-e).

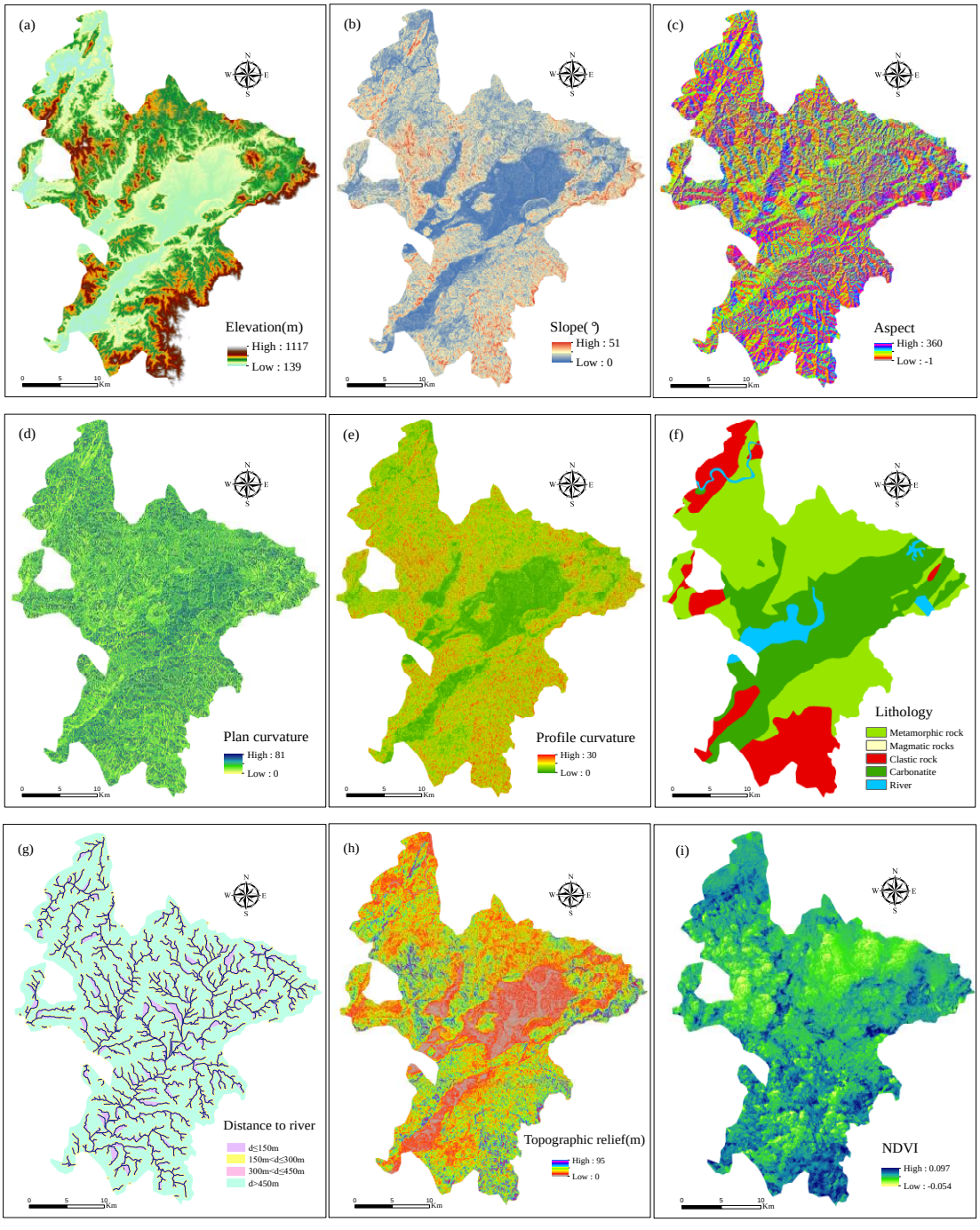
(2) Surface cover factor

While NDBI describes the density of surface buildings, NDVI measures the extent of local vegetation cover. A frequency ratio greater than 1 denotes an area that is more prone to landslides when NDVI is in the range of 0.018 to 0.025 and 0.033 to 0.098 or NDBI is in the range of -0.318 to -0.219 attributes (Table 1, Figure 3-i,g).

(3) Hydrological environmental factors

Water not only speeds up the erosion of geotechnical bodies but also increases the likelihood of landslides by weakening and misaligning the interbedded soils between sliding surfaces. Since surface water may be more clearly separated from water bodies by shading (Yang et al., 2018), the MNDWI is primarily used to elaborate the distribution of surface water. Table 1 and Figure 3-k show that the MNDWI index is vulnerable to landslides between 0.217 and 0.276.

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024



235

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

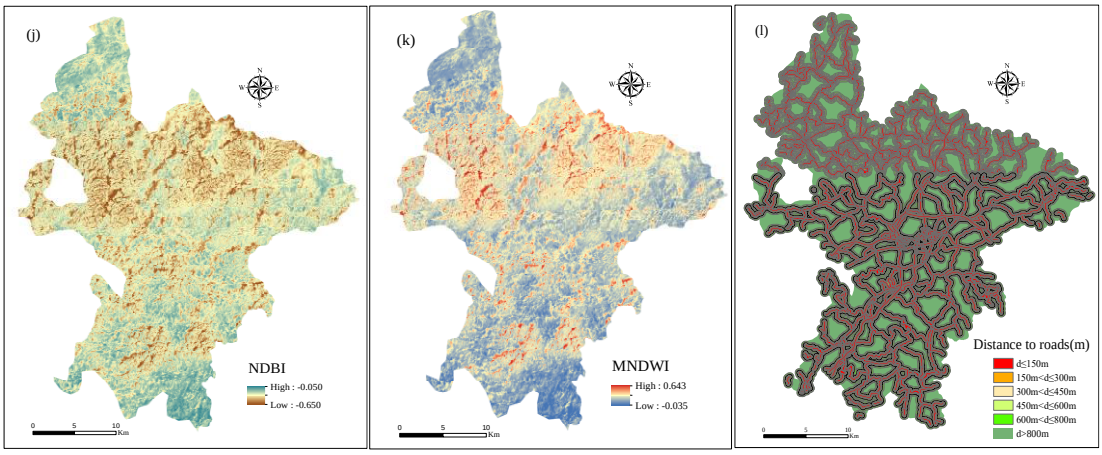


Figure 3. Basic environmental factors of landslide in Ruijin County: (a) Elevation: 139~1117m; (b) Slope:0~51; (c) Aspect: -1~360; (d) Plan curvature: 0~81; (e) Profile curvature: 0~30; (f) Lithology: Metamorphic rock; Magmatic rock; Clastic rock; Carbonatite; (g) Distance between landslide and river ;(h) Topographic relief of Ruijin: 0~90m; (i) NDVI: -0.054~0.097; (j) NDBI: -0.650~-0.050; (k) MNDWI: -0.035~0.643; (l) Distance between landslide and river.

Table 1 Attribute interval and frequency ratio of each evaluation factor.

Environmental factors	Values	Number of grids in the whole area	Grid scale/%	Landslide grid	Landslide grid scale /%	FR
Elevation(m)	139.7~250.9	836745	30.419	173	46.757	1.152
	250.9~335.3	796482	28.956	121	32.703	1.175
	335.3~423.5	578691	21.038	49	13.243	0.796
	423.5~538.6	332056	12.072	20	5.405	0.650
	538.6~695.9	147930	5.378	4	1.081	0.159
	695.9~1117.8	58787	2.137	3	0.811	0.376
Slope(°)	0~4.4	685218	24.911	41	11.081	0.260
	4.4~8.8	643535	23.395	125	33.784	0.986
	8.8~13.2	608755	22.131	113	30.541	1.276
	13.2~17.9	446520	16.233	56	15.135	1.945
	17.9~28.7	344703	12.532	34	9.189	0.632
	28.7~51.2	21960	0.798	1	0.270	0.398
Aspect	-1	155940	5.669	27	7.297	0
	0~22.5	297924	10.831	31	8.378	0.994
	22.5~67.5	354479	12.887	62	15.757	0.954

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

	67.5~112.5	359791	13.080	48	12.973	1.301
	112.5~157.5	332830	12.099	54	14.595	1.198
	157.5~202.5	332143	12.075	42	11.351	1.160
	202.5~247.5	378011	13.742	48	12.973	1.086
	247.5~292.5	370195	13.458	38	10.270	0.792
	292.5~337.5	169378	6.158	20	5.405	0.716
Profile curvature	0~2.029	884499	32.156	98	26.486	0.596
	2.029~4.057	773416	28.117	125	33.784	1.072
	4.057~6.324	561551	20.415	69	18.649	1.126
	6.324~8.949	324376	11.793	54	14.595	1.213
	8.949~14.529	187647	6.822	23	6.216	0.823
	14.529~30.428	19202	0.698	1	0.270	1.829
	0~13.422	651677	23.691	111	30.000	1.246
	13.422~24.927	625675	22.746	99	26.757	1.417
Plan curvature	24.927~37.710	471544	17.143	58	15.676	1.434
	37.710~52.091	354666	12.894	42	11.351	0.932
	52.091~67.749	301696	10.968	24	6.486	0.657
	67.749~81.491	345433	12.558	36	9.729	0.852
	0~6.022	651450	23.683	35	9.459	0.476
	6.022~12.420	721236	26.220	148	40.000	0.566
Topographic relief	12.420~18.819	641938	23.337	104	28.108	1.163
	18.819~22.969	293597	10.674	43	11.622	2.575
	22.969~35.379	385799	14.026	38	10.270	0.431
	35.379~95.975	72855	2.649	3	0.811	0.367
	Metamorphic rock	1218584	44.301	108	29.189	1.301
Lithology	Magmatic rock	503748	18.314	27	7.297	1.611
	Clastic rock	899363	32.696	19	5.135	0.209
	Carbonatite	128996	4.689	216	58.378	0.659
	-0.054~0.006	68098	2.476	5	1.351	0.192
	0.006~0.018	299115	10.874	34	9.189	0.803
NDVI	0.018~0.025	580373	21.099	56	15.135	1.009
	0.025~0.033	848420	30.843	146	39.459	0.955

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

NDBI	0.033~0.042	635132	23.089	87	23.514	1.075
	0.042~0.098	315488	11.469	42	11.351	1.141
	-0.650~-0.389	74963	2.725	13	3.513	1.101
	-0.389~-0.318	234632	8.529	28	7.568	0.928
	-0.318~-0.267	428674	15.584	58	15.676	1.418
	-0.267~-0.219	699581	25.433	92	24.865	1.233
	-0.219~-0.173	803445	29.209	110	29.729	0.901
	-0.173~-0.050	505331	18.371	69	18.649	0.729
MNDWI	-0.035~0.110	365882	13.301	48	12.973	1.374
	0.110~0.164	773621	28.125	118	31.892	1.221
	0.164~0.217	772212	28.073	94	25.405	0.952
	0.217~0.276	492158	17.892	67	18.108	1.174
	0.276~0.352	256718	9.333	29	7.838	1.082
	0.352~0.643	86035	3.128	13	3.514	0.708
Distance to river(m)	<150	155212	5.642	47	12.703	2.586
	150~300	55808	2.029	7	1.891	1.689
	300~450	279114	10.147	118	31.892	0.672
	>450	2274116	82.674	198	53.514	0.497
Distance to roads(m)	<150	265206	31.431	112	30.270	0.963
	150~300	366479	28.134	90	24.324	3.139
	300~450	337201	21.789	82	22.162	1.663
	450~600	599351	12.259	45	12.162	0.558
	600~800	773872	6.052	34	9.189	0.327
	>800	408582	0.335	7	1.891	0.127

(4) Underlying geological factors

Mudstone type and sandstone type rock formations are conducive to landslide development (Table 1, Figure 3-f).

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

4 Landslide susceptibility prediction results

4.1 Susceptibility modelling process for SVM and RF

The raster cell delineation of landslides and environmental factors is carried out at 30m resolution throughout Ruijin, then the specific implementation steps for landslide susceptibility prediction modelling from environmental factors and landslide cataloguing cells are as follows.

The raster cells of landslide and environmental factors were divided based on 30m resolution in the whole Ruijin city, then the specific implementation steps of landslide susceptibility prediction modelling from environmental factors and landslide cataloguing cells are as follows.

(1) ArcGIS 10.2 should be used to convert each environmental factor's raster cell to vector point format, then link the characteristics of the vector point objects of various environmental factors together and apply frequency ratio values. For the RF and SVM models, the frequency ratios of the environmental parameters in Section 1.2 served as independent variables (i.e., input variables).

(2) In ArcGIS software, the 370 landslide surfaces were transformed into 8633837 landslide raster cells, which were then transformed into landslide points and given a susceptibility value of 1. Likewise, 8633837 raster cells were chosen at random from the study region as non-landslide raster cells. These non-landslide raster cells were then transformed into non-landslide points and given a susceptibility value of 0. These raster cells have values of 1 and 0 assigned to them. The dependent variables (i.e., output variables) of the RF and SVM models are these raster cell attribute values that have been set to 1 and 0.

(3) The aforementioned vector point objects must be transferred from the ArcGIS 10.2 program to the appropriate machine learning modelling program, for example, RF models can be created with the RF package in R.

(4) In order to carry out the training and testing of the RF and SVM models, a total of 17,267,674 landslide and non-landslide points from step (2) are linked to their corresponding environmental factor frequency ratios. The landslide and non-landslide points are then randomly divided into two sets, the model training set and the test set, in a 7:3 ratio.

(5) The trained RF and SVM models were used to predict the landslide susceptibility of 2395696 point objects (including input variables, the output variables can be set arbitrarily) in the whole study area, and the susceptibility values of all points were further imported into ArcGIS10.2 software and converted into raster cells.

(6) Each raster cell's attribute value reflects the landslide susceptibility value, which may be separated into five susceptibility zones using the natural intermittent point method in ArcGIS10.2 software: very low, low, medium, high, and very high (Xiao et al., 2019).

The RF accuracy in this study was mainly adjusted using parameters such as the number of factor features m and the number of trees t. Based on the out-of-bag error screening method, the optimal parameters of m and t for the RF model to predict landslide susceptibility in Yanchang County were determined to be 6 and 477, respectively, while for the SVM model to predict landslide susceptibility, the regularization parameter C of the SVM model was determined to be 8, the regression accuracy e to be 0.1 and the RBF kernel function parameter to be 0.2 by the cross-validation method. the optimal parameters

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

obtained above were used to complete the training and testing of the RF and SVM models The best parameters obtained above were used to complete the training and testing of the RF and SVM models, and to predict the landslide susceptibility in Ruijin City. The landslide susceptibility index of Ruijin was further classified into five categories: very low, low, medium, high and very high susceptibility zones.

4.2 Analysis of landslide susceptibility results in Ruijin City

Figure 4 and Table 2 display the outcomes of the landslide susceptibility forecast using the RF and SVM models for Ruijin City. The area ratios of very low, low, medium, high, and very high landslide susceptibility areas predicted by the RF model are respectively 14.78%, 21.50%, 20.55%, 21.85%, and 21.35%, while the area ratios predicted by the SVM model are respectively 92.61%, 28.02%, 1.68%, 1.33%, and 1.58%. Table 2 demonstrates that the ratio of landslide frequency rises within each level as the level of landslide susceptibility rises. To determine the frequency ratio accuracy of the landslide susceptibility results, the frequency ratios of the very high and high susceptibility zones are divided by the sum of all frequency ratios. The results are 0.248, 0.249, 0.434, 0.755, and 3.076 for the very low to very high susceptibility zones, and 0.041, 0.37, and 0.37 for the very low to very high susceptibility zones for the SVM model. This comparison demonstrates that the RF model's frequency ratio is significantly more accurate than the SVM model's, indicating that the RF model's projected landslide susceptibility more accurately captures the spatial aggregation properties and distribution pattern of regional landslides. The landslide susceptibility maps of Ruijin City predicted by the SVM and RF models are, of course, comparable overall, with the area of the very low susceptibility zone predicted by the SVM model being larger than that of the RF model while the area of the very high susceptibility zone is smaller than that of the RF model. While the very high and high susceptibility areas are primarily centered in the middle portion of the city and stretch in a striped pattern to the southeast and southwest, the extremely low and low landslide susceptibility zones are typically found in the eastern section of Ruijin City and the surrounding area. The fundamental cause of this is because Ruijin's urban districts are situated along the central Yanhe River, where the river's long-term scouring effect on the banks and the frequent human engineering activities brought on by the growing urbanization are favourable to the formation of landslides. Additionally, the loose stratigraphy in hilly regions like the south and north, where gullies and ravines are interspersed, is brittle, and landslides are more prone to happen in times of severe rainfall.

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

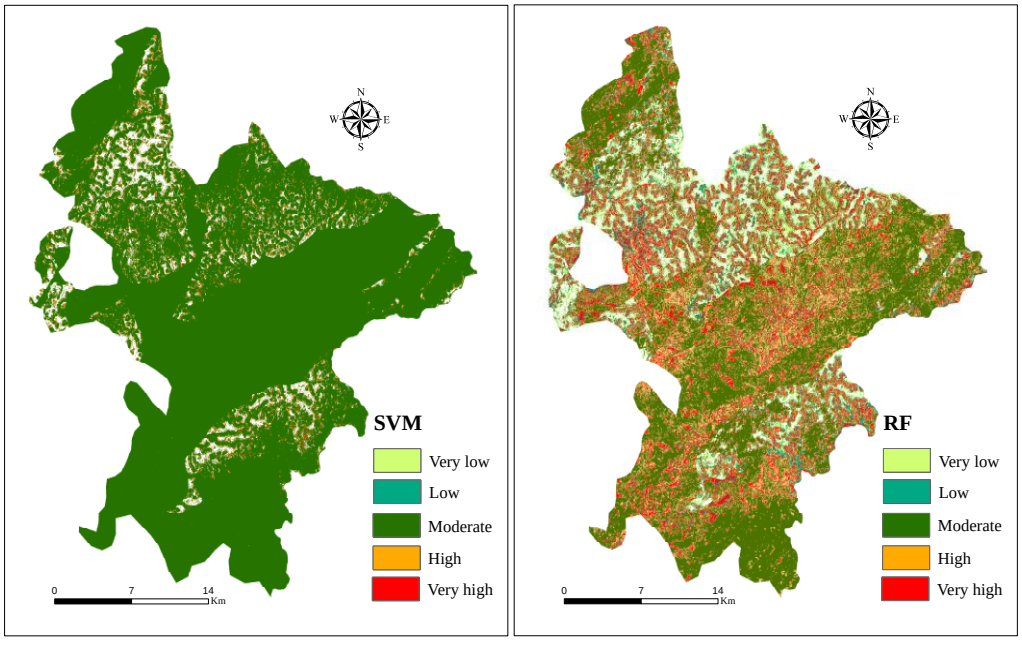


Figure 4. The graph on the left is a prediction of landslide susceptibility using SVM model. Green indicates a low probability of landslide occurrence, while red indicates a high probability of landslide occurrence. The darker the color, the greater the probability of landslide occurrence; The graph on the right is a prediction of landslide susceptibility using RF model. Green indicates a low probability of landslide occurrence, while red indicates a high probability of landslide occurrence. The darker the color, the greater the probability of landslide occurrence.

Table 2 Frequency ratio precision analysis of susceptibility graphs of RF and SVM models

Model	Susceptibility level	Area grid	Area grid scale	Landslide grid	Landslide grid scale	FR
RF	Very low	406548	14.78	13	1.08	0.248
	Low	591490	21.50	20	2.70	0.249
	Moderate	565391	20.55	33	7.03	0.434
	High	600890	21.85	61	13.24	0.755
	Very high	587270	21.35	243	75.95	3.076
SVM	Very low	2547431	92.61	14	0.54	0.041
	Low	380050	28.02	19	1.08	0.372
	Moderate	46152	1.68	22	1.08	3.544
	High	36621	1.33	29	2.97	5.887
	Very high	434135	1.58	285	94.32	4.881

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

5 Discussion

Discussions of modelling are now conducted from a variety of angles, such as the characteristics of uncertainty in the outcomes of RF and SVM models for predicting landslide susceptibility, the identification of master control factors for landslide development in Ruijin City, and the challenges and future directions in machine learning models for predicting landslide susceptibility.

5.1 Susceptibility modelling process for SVM and RF

5.1.1 Data sources

The ROC curve, which is primarily composed of the horizontal and vertical coordinates of 1-specificity and sensitivity, with 1-specificity indicating the proportion of landslides misclassified and sensitivity indicating the proportion of landslides correctly classified, is widely used in the evaluation of the overall accuracy of dichotomous classification in the field of landslides (Cantarino et al., 2018). Figure 5 illustrates the ROC curve's accuracy in assessing the RF and SVM models' prediction of landslide risk in extended counties. The area under the ROC curve (AUC value) values for the RF and SVM models are 0.8768 and 0.8621, respectively, as can be seen. This indicates that while both models have improved prediction performance overall, the RF model is more accurate at predicting susceptibility than the SVM model. The comparison of the SVM and RF models' ROC curve prediction accuracy yielded findings that were in line with each model's frequency ratio accuracy.

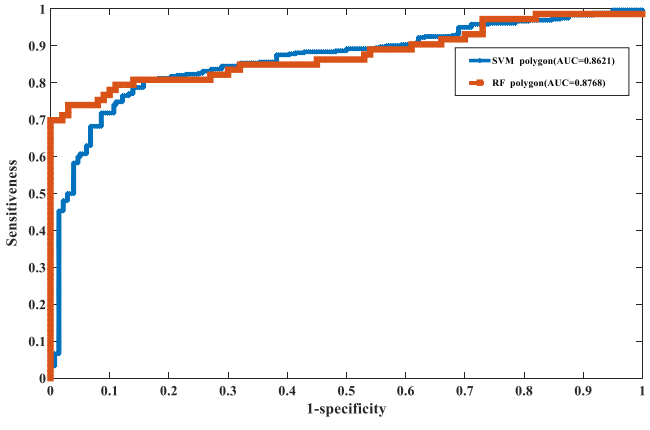


Figure 5. ROC curves for predicting landslide susceptibility using RF and SVM models. RF has a large AUC area and high prediction accuracy.

The results of the prediction of landslide susceptibility show that the SVM and RF models have concentrated distributions of landslide susceptibility indices in the very low and low susceptibility zones, but the distribution patterns of the remaining susceptibility zones are somewhat different; the susceptibility indices predicted by the SVM model show a small increase in the high and very high susceptibility zones, whereas the susceptibility indices predicted by the RF model show The SVM model projected a mean value of 0.363, which was higher than the RF model's expected value of 0.335 for the low susceptibility

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

range. Indirectly demonstrating that the RF model's susceptibility prediction uncertainty was lower than that of the SVM model, the RF model had a lower mean value of susceptibility index than the SVM model with higher prediction accuracy.

In addition, the standard deviation of the landslide susceptibility index predicted by the SVM model is 0.286, which is higher than that of the RF model at 0.254. The dispersion of the susceptibility index of both models is exactly opposite to its mean (SD) SD(RF) < SD(SVM), indicating that the SVM model is better than the RF model in differentiating the susceptibility index of landslides in Ruijin City. Overall, both the SVM and RF models can better reflect the differences in susceptibility indices within different raster cells, and effectively reflect as much known landslide inventory information as possible with a lower susceptibility index, indirectly indicating that advanced machine learning models can effectively predict landslide susceptibility. Combining the above prediction accuracy and susceptibility index characteristics, the uncertainty in the susceptibility modelling of the RF model is lower than that of the SVM model.

5.2 Identification of the main control factors for landslides in Ruijin City

The frequency ratio of each factor in Table 1 for the pattern of each base environmental factor attribute interval affecting landslide development shows that elevation is 139.7–335.3m, slope is 8.8–17.9 ° or slope direction is 67.5–245.5 °, profile curvature is 2.029–8.949, plane curvature is 0–37.71, lithology type is sandstone and mudstone type, and NDVI is in the range of 0–37. Landslides tend to occur in areas with NDBI values between -0.318 and -0.219, a slope between 0 and 67.5 degrees, an elevation between 335.3 and 1117.8 meters, a plan curvature between 37.71 and 81.491 degrees, and a profile curvature between 0.029 and 8.949 degrees.

NDVI values between -0.054-0.018 and 0.025-0.161, NDBI values between -0.389-0.318 and -0.219-0.050, and profile curvature between 0 and 8.949 and 8.949 to 30.428 are all incompatible with the development of landslides. Studies on landslide susceptibility might use the significance of the underlying environmental causes for landslides as a theoretical reference. In this work, Origin 9.0 software was used to analyze and process 12 environmental elements from the SVM and RF models to determine the relevance ranking of each factor (Figure 6). The importance of environmental factors is determined by the mean Gini index reduction, which is employed in the RF model as a percentage of the total mean Gini reduction values for all environmental factors.

Figure 6a and 6b show that the importance and magnitude of environmental factors influencing landslide occurrence differ between SVM and RF modelling, and a thorough comparison of the two types of models reveals both similarities and differences in how different models handle environmental factors. The relevance of the environmental elements was objectively assessed using the weighted average approach. The accuracy of the SVM and RF models in predicting landslide susceptibility was used to first determine the weights of the corresponding factor contributions, and then a weighted average was used to determine the average weight values of the various underlying environmental factors influencing landslide susceptibility (Figure 7-a). The average importance of slope, elevation and lithology are 0.23, 0.20 and 0.174 respectively, while the average importance of NDVI and MNDWI are 0.055 and 0.038 respectively (Figure 7-b).

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

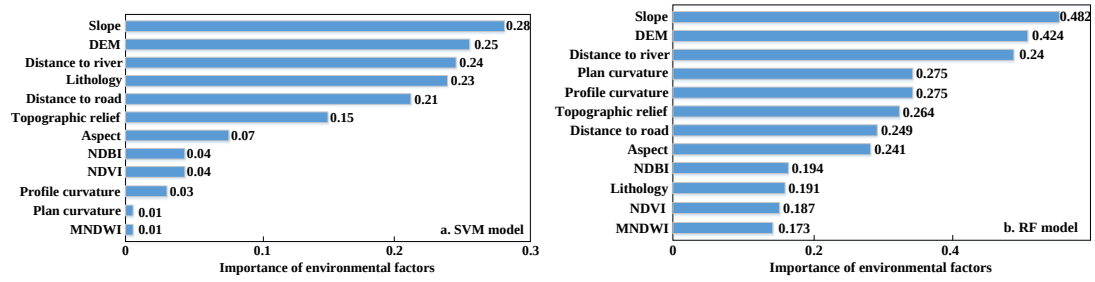


Figure 6. Figures 6a and 6b show that SVM and RF modelling have different importance and magnitude in dealing with environmental factors that affect landslide occurrence.

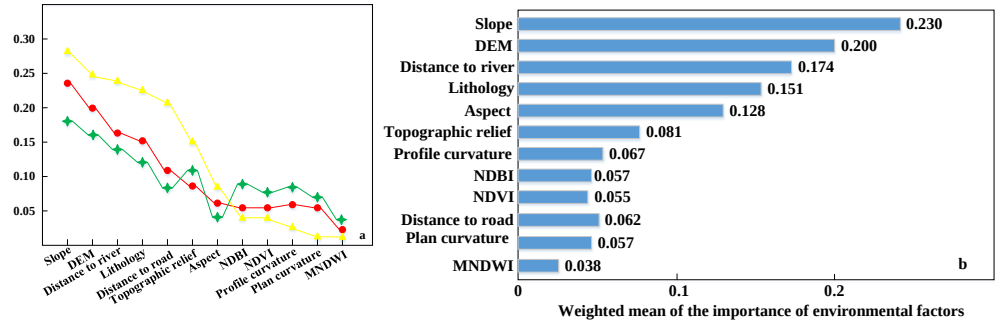


Figure 7. The accuracy of SVM and RF models in predicting landslide susceptibility is first used to determine the weights of corresponding factor contributions, and then the weighted average is used to determine the average weight values of various potential environmental factors that affect landslide susceptibility (Figure 7-a). The average importance of slope, elevation, and lithology are 0.23, 0.20, and 0.174, respectively, while the average importance of NDVI and MNDWI are 0.055 and 0.038, respectively.

5.3 Problems and frontiers of machine learning models for predicting susceptibility

A contentious and challenging issue has been how to choose and construct an appropriate landslide susceptibility model. In the past, researchers have utilized a variety of models to forecast landslide susceptibility in various locations and have come to certain conclusions (Sheikh et al., 2019; Xing et al., 2020; Xing et al., 2021). Information quantity models, in contrast to logistic regression, can only indirectly reflect the contribution of factors to susceptibility and are also impacted by factor classification and raster image components. For instance, physical models like SINMAP primarily focus on local slope stability but are not very specific for large areas and are prone to spatial errors. The SVM model can tackle non-linear or high-dimensional identification issues with few samples despite being challenging to comprehend and use.

In summary, there are advantages and disadvantages and applicability of each of the above models, which need to be explored and analysed in the literature to determine the better model. Compared with the earlier heuristic models and conventional mathematical and statistical models, machine learning models including artificial neural networks, SVM, decision trees, RF and deep learning have been widely used for landslide susceptibility prediction modelling in the last 15 years, and their prediction effectiveness has been generally recognised by many scholars (Al-Najjar et al., 2020). In particular, the RF model has distinctive features compared with other machine learning models: its prediction accuracy and training test efficiency are

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

high, it can be applied on a large scale, it can handle high-dimensional data, it can resist noise interference and has strong modelling adaptability, etc. (Tomasevic et al., 2020).

The RF model has also been used to model the susceptibility of certain landslide prone areas, for example, Huang et al (Huang et al, 2018) found that the RF model achieved very reasonable susceptibility grading results in predicting the landslide susceptibility of the Hubei section of the Three Gorges reservoir area. The RF model achieved very reasonable susceptibility grading results; Ahmed et al (Youssef et al., 2016) established RF, SVM and logistic regression models to predict landslide susceptibility in the Three Gorges reservoir area and found that the prediction accuracy of RF was higher than other models; Akinci et al (Akinci et al, 2021) and Ali et al (Ali et al., 2020) made a comparative study of the RF model and other machine learning models for landslide susceptibility prediction and found that the RF model had stronger modelling adaptability.

The findings of this study support those of the literature review discussed above, namely that RF is a trustworthy and accurate model for predicting the vulnerability of a given area to landslides. Machine learning models actually have a promising future in the field of susceptibility modelling due to their higher prediction accuracy as the research on susceptibility prediction models continues to advance and mature. Optimised random forest models, multi-class machine learning integrated models, convolutional neural networks, and deep learning techniques like fully connected sparse self-coding are some of the machine learning models for susceptibility modelling that are currently being developed. These models are based on coupled machine learning models that take into account the weight of the evidence or the amount of information (Bui et al., 2019; Islam et al., 2020). It should be highlighted that despite maintaining high forecast accuracy and reliability, the aforementioned models shouldn't be unduly complex, as this prevents them from being verified in engineering practice and for general applicability.

6 Discussion

(1) Both the SVM and RF models have good landslide susceptibility prediction performance, and their predicted distribution patterns of landslide susceptibility in Ruijin City are generally similar, with the very low and low susceptibility areas being primarily distributed in the eastern part of Ruijin City and its surroundings, and the very high and high susceptibility areas being primarily concentrated in the central part of the county and extending in a strip in the southeast and southwest

(2) The weight mean method's calculation results demonstrate that slope, elevation, and lithology are the key determining variables of landslide development in Ruijin, while NDVI and MNDWI, among other parameters, have less of an impact on this process.

(3) The literature review demonstrates that the RF model is a highly accurate and reliable susceptibility prediction model compared to many other machine learning models. The accuracy of the RF model in predicting landslide susceptibility in Ruijin is higher than that of the SVM model, while its uncertainty is lower than that of the SVM model.

(4) While increasing the accuracy of susceptibility prediction, attention should be paid to reducing the complexity of the models in order to facilitate their widespread use. Machine learning primarily predicts landslide susceptibility in the direction of various coupled models, integrated models, and deep learning models.

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

Acknowledgments The authors would highly thank the National Service Center for China University of Geosciences for providing relevant data.

Author Contributions Conceptualization, Yin Xing; methodology, Yang Chen and Saipeng Huang; formal analysis, Yin Xing and Yang Chen; investigation, Saipeng Huang and Peng Wang; writing—original draft preparation, Yin Xing; writing—review and editing, Yin Xing, Yang Chen, Saipeng Huang, Peng Wang and Yunfei Xiang; supervision, Yang Chen; funding acquisition, Yunfei Xiang. Authorship must be limited to those who have contributed substantially to the work reported.

Funding Natural Science Fund for Colleges and Universities of Jiangsu Province (22KJB420002); Suzhou University of Science and Technology Talent Introduction Initiation Project (332214808).

Data Availability Statement Not applicable.

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Achour, Y., Pourghasemi, H. R., 2019. How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? Geoscience Frontiers.11(3), 871-873.

Akinci, H., Zeybek, M., 2021. Comparing classical statistic and machine learning models in landslide susceptibility mapping in Ardanuc (Artvin), Turkey. Natural Hazards. 108(2): 1515-1543.

Ali, S.A., Parvin, F., Vojtekov á, J., Costache, R., Linh, N.T.T., Pham, O.B., Vojtek, M., Gigović, L., Ahmad, A., 2020. GIS-Based Landslide Susceptibility Modeling: A Comparison between Fuzzy Multi-Criteria and Machine Learning Algorithms. Geoscience Frontiers .12(2), 857-876.

Al-Najjar, H.A.H., Pradhan, B., 2020. Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. Geoscience Frontiers. 12(2), 625-637.

Ananthakrishnan, A. N., Bernstein, C. N., Iliopoulos, D., Macpherson, A., Neurath, M.F., Ali, R.A.R., Vavricka, S.R., Fiocchi, C., 2018. Environmental triggers in IBD: a review of progress and evidence. Nat Rev Gastroenterol Hepatol. 15(1): 39-49.

Bui, D. T., Hoang, N.D., Mart ínez-Álvarez, F., Ngo, P.T.T., Hoa, P.V., Pham, T.D., Samui, P., 2019. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. Science of the Total Environment. 701, 134413.

Bui, D. T., Tsangaratos, P.,Ngo, P.T.T., Pham, T.D., Pham, B.T., 2019. Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. Science of The Total Environment. 668, 1038-1054.

Buia, D. T., Tsangaratos, P., Nguyen, V.T., Liem, N.V., Trinh, P.T., 2020. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. Catena. 188, 104426.

Cantarino, I., Carrion, M.A., Goerlich, F., Iba ñez, V.M., 2019. A ROC analysis-based classification method for landslide susceptibility maps. Landslides. 16, 265-282.

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

Cervantes, J., Garcia-Lamont, F., Rodr ́guez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 408: 189-215.

Chang, Z., Du, Z., Zhang, F., Huang, F., Chen, J., Li, W., Guo, Z., 2020. Landslide Susceptibility Prediction Based on Remote Sensing Images and GIS: Comparisons of Supervised and Unsupervised Machine Learning Models. Remote Sensing. 12 (3): 502.

Chen, W., Zhang, S., Li, R., Shahabi, H., 2018. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and na ïve Bayes tree for landslide susceptibility modeling. Science of the Total Environment. 644: 1006-1018.

De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., et al. 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine. 24(9): 1342-1350.

Di Napoli, M., Carotenuto, F., Cevasco, A., Confuorto, P., Di Martire, D., et al. 2020. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. Landslides. 17(8): 1897-1914.

Gu. D., Huang, D., Yang, W., Zhu, J., Fu, G., 2017. Understanding the triggering mechanism and possible kinematic evolution of a reactivated landslide in the Three Gorges Reservoir. Landslides. 14(6): 2073-2087.

Dou, J., Yunus, A.P., Bui, D.T., Merghadi, A., Sahana, M., et al. 2019. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Science of the Total Environment. 662: 332-346.

Hong, H., Tsangaratos, P., Ilia, I., Loupasakis, C., Wang, Y., 2020. Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping. Science of the Total Environment. 742: 140549.

Huang, F., Huang, J., Jiang, S., Zhou, C., 2017. Landslide displacement prediction based on multivariate chaotic model and Extreme Learning Machine. Engineering Geology. 218, 173-186.

Huang, F., Zhang, J., Zhou, C., Wang, Y., Huang, J., Zhu, L., 2020. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. Landslides. 17(1): 217-229.

Huang, F., Zhou, Y., Jiang, S., Huang, J., Chang, Z., Chen, J., 2021. Uncertainty study of landslide susceptibility prediction considering the different attribute interval numbers of environmental factors and different data-based models. Catena. 202, 105250.

Huang, F., Cao, Z., Guo, J., Jiang, S., Li, S., Guo, Z., 2020. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. Catena. 191, 104580.

Huang, F., Cao, Z., Jiang, S., Zhou, C., Huang, J., Guo, Z., 2020. Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. Landslides. 17(12), 2919-2930.

Huang, Y., Zhao, L., 2018. Review on landslide susceptibility mapping using support vector machines. Catena. 165, 520-529.

Islam, A.R.M.T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K.U., et al. 2020. Flood susceptibility modelling using advanced ensemble machine learning models. Geoscience Frontiers. 12(3), 101075.

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

Jiang, S., Huang, J., Huang, F., Yang, J., Yao, C., Zhou, C., 2017. Modelling of spatial variability of soil undrained shear strength by conditional random fields for slope reliability analysis. Applied Mathematical Modelling. 63, 374-389.

Lee, J.H., Sameen, M.I., Pradhan, B., Park, H.J., 2017. Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. Geomorphology. 303, 284-298.

Li, Y., Yan, C., Liu, W., Li, M., 2018. A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification. Applied Soft Computing. 70, 1000-1009.

Luo, W., Liu, C., 2018. Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods. Landslides. 15(3), 465-474.

Mallick, J., Singh, R.K., AlAwadh, M.A., Islam, S., Khan, R.A., Qureshi, M.N., 2018. GIS-based landslide susceptibility evaluation using fuzzy—AHP multi-criteria decision-making techniques in the Abha Watershed, Saudi Arabia. Environmental Earth Sciences. 77(7), 276.

Manibardo, E. L., Lana, I., Ser, J.D., 2022. Deep Learning for Road Traffic Forecasting: Does it Make a Difference? IEEE Transactions on Intelligent Transportation Systems. 23(7), 6164-6188.

Merghadi, A., Abderrahmane, B., Bui, D.T., 2018. Landslide Susceptibility Assessment at Mila Basin (Algeria): A Comparative Assessment of Prediction Capability of Advanced Machine Learning Methods. ISPRS Int. J. Geo-Inf. 7(7), 268.

Min, M., Bai, C., Guo, J., Sun, F., Liu, C., et al. 2019. Estimating Summertime Precipitation from Himawari-8 and Global Forecast System Based on Machine Learning. IEEE Transactions on Geoscience and Remote Sensing 57(5): 2557-2570.

Naemitabar, M., Asadi, M.Z., 2021. Landslide zonation and assessment of Farizi watershed in northeastern Iran using data mining techniques. Natural Hazards, 108(3), 2423-2453.

Rodrigues, S.G., Silva, M.M., Alencar, M.H., 2021. A proposal for an approach to mapping susceptibility to landslides using natural language processing and machine learning. Landslides, 18(7), 2515-2529.

Sameen, M.I., Pradhan, B., Lee, S., 2020. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. Catena. 186, 104249.

Scherzer, S., Lujala, P., Rød, J.K., 2019. A community resilience index for Norway: An adaptation of the Baseline Resilience Indicators for Communities (BRIC). International Journal of Disaster Risk Reduction, 36, 101107.

Sheikh, V., Kornejady, A., Ownegh, M., 2019. Application of the coupled TOPSIS—Mahalanobis distance for multi—hazard—based management of the target districts of the Golestan Province, Iran. Natural Hazards. 96, 1335-1365.

Smith, H.G., Spiekermann, R., Betts, H., Neverman, A.J., 2021. Comparing methods of landslide data acquisition and susceptibility modelling: Examples from New Zealand. Geomorphology, 381, 107660.

Steger, S., Mair, V., Kofler, C., Pittore, M., Zebisch, M., Schneiderbauer, S., 2021. Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling—Benefits of exploring landslide data collection effects. Science of The Total Environment. 776, 145935.

https://doi.org/10.5194/egusphere-2023-3134
Preprint. Discussion started: 6 February 2024

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.A., Rahman, A., 2020. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. Remote Sensing. 12(7), 1135.

Tomasevic, N., Gvozdenovic, N., 2020. An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & Education. 143, 103676.

Tsangaratos, P., Ilia, I., Hong, H., Chen, W., Xu, C., 2017. Applying Information Theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China. Landslides, 14(3), 1091-1111.

Xiao, W., Lv, X., Zhao, Y., Sun, H., Li, J., 2019. Ecological resilience assessment of an arid coal mining area using index of entropy and linear weighted analysis: A case study of Shendong Coalfield, China. Ecological Indicators. 109, 105843.

Xing, Y. Yue, J., Chen, C., Qin, Y., Hu, J., 2020. A hybrid prediction model of landslide displacement with risk-averse adaptation. Computers & Geosciences, 141, 104527.

Xing, Y., Yue, J., Chen, C., Cai, D., Hu, J., Xiang, Y., 2021. Prediction interval estimation of landslide displacement using adaptive chicken swarm optimization-tuned support vector machines. Applied intelligence, 51(11), 8466-8483.

Yang, X., Qin, Q., Grussenmeyer P., Koehl, M., 2018. Urban surface water body detection with suppressed built-up noise based on water indices from Sentinel-2 MSI imagery. Remote Sensing of Environment. 219, 259-270.

Yin, G., Luo, J., Niu, F., Lin, Z., Liu, M., 2021. Machine learning-based thermokarst landslide susceptibility modeling across the permafrost region on the Qinghai-Tibet Plateau. Landslides.18, 2639-2649.

Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Alkatheeri, M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides, 13(5), 839-856.

Youssef, A. M., Pourghasemi, H.R., 2020. Landslide susceptibility mapping using machine learning algorithms and comparison. Geoscience Frontiers. 12(2), 639-655.

Zhang, L., Shi, B., Zhu, H., Yu, X., Han, H., Fan, X., 2021. PSO-SVM-based deep displacement prediction of Majiagou landslide considering the deformation hysteresis effect. Landslides. 18(1), 179-193.

Zhang, P., Qin, C., Hong, X., Kang, G., Qin, M., et al. 2018. Risk assessment and source analysis of soil heavy metal pollution from lower reaches of Yellow River irrigation in China. Science of the Total Environment. 633, 1136-1147.