



OPEN ACCESS

EDITED BY

Zhang Cong,
Central South University Forestry and
Technology, China

REVIEWED BY

Yi Chang,
East China University of Technology, China
Pengjiao Jia,
Soochow University, China
Fei Huang,
Flinders University, Australia

*CORRESPONDENCE

Ying Yang,
✉ 493662400@qq.com

RECEIVED 21 February 2025

ACCEPTED 01 April 2025

PUBLISHED 06 May 2025

CITATION

Wang J, Yang Y, Yang X, Lu Y, Liu Y, Hu D and
Hu Y (2025) The prediction of karst-collapse
susceptibility levels based on the ISSA-ELM
integrated model.
Front. Earth Sci. 13:1581090.
doi: 10.3389/feart.2025.1581090

COPYRIGHT

© 2025 Wang, Yang, Yang, Lu, Liu, Hu and Hu.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

The prediction of karst-collapse susceptibility levels based on the ISSA-ELM integrated model

Jiaxin Wang¹, Ying Yang^{2*}, Xian Yang³, Yulong Lu⁴, Yang Liu⁴,
Da Hu³ and Yongjia Hu⁵

¹Department of Natural Resources, Hunan Vocational College of Engineering, Changsha, China, ²Department of Discipline Inspection and Supervision, Hunan Vocational College of Engineering, Changsha, China, ³Hunan Engineering Research Center of Structural Safety and Disaster Prevention for Urban Underground Infrastructure, Hunan City University, Yiyang, China, ⁴School of Earth Sciences and Spatial Information, Hunan University of Science and Technology, Xiangtan, China, ⁵School of Resources and Safety Engineering, Central South University, Changsha, China

Karst collapse, a sudden geological hazard with complex mechanisms and low predictability, presents significant threats to urban safety and sustainable development by jeopardizing human lives and infrastructure. To address the limitations of conventional prediction methods, in this study, we introduce an enhanced predictive model, the improved sparrow search algorithm-optimized extreme learning machine (ISSA-ELM), for accurate karst-collapse susceptibility assessment. The methodology incorporates two key innovations: first, it applies a Singer chaotic mapping technique to enhance the sparrow search algorithm (SSA), effectively mitigating local optima entrapment by increasing population diversity and enhancing global search capabilities. Second, the optimized ISSA automatically adjusts the initial weights and thresholds of the ELM, whereas a five-fold cross-validation is used to determine the optimal hidden layer configuration, forming an adaptive and intelligent prediction framework. When validated against 20 datasets from a representative karst region, the proposed model achieved exceptional performance, with a mean absolute error (MAE) of 0.0544 and a coefficient of determination (R^2) of 0.9914, significantly surpassing the prediction accuracy of conventional ELM and SSA-ELM models. The results underscore the ISSA-ELM's superior nonlinear fitting capability, enhanced generalization performance, and outstanding stability in practical engineering applications. This research offers a solid scientific foundation for risk classification and hazard mitigation strategies while introducing a novel methodological framework through the integration of innovative algorithms. The proposed technical pathway provides significant theoretical advancements and practical engineering values for geological disaster prediction systems.

KEYWORDS

karst collapse, susceptibility prediction, improved sparrow search algorithm, extreme learning machine, ISSA-ELM integrated model

1 Introduction

Karst collapse, a prominent geological hazard, arises from the interaction between karst formation processes and groundwater dynamics. It often results in sudden surface subsidence or structural failure which occurs due to the

progressive development of subsurface voids and fractures (Bouzerda et al., 2024; Wang et al., 2024; Xiaozhen et al., 2024). With its concealed nature, intricate triggering mechanisms, and significant destructive potential, karst collapse poses substantial challenges to urban development and infrastructure security in vulnerable regions. Conducting an accurate susceptibility assessment is crucial for enhancing disaster prevention frameworks and guiding sustainable land-use planning.

The development of karst collapse prediction methodologies has progressed through distinct phases, evolving from qualitative evaluations (Gao et al., 2009; Luo and Shen, 2018) to more sophisticated quantitative analyses. Early research primarily established qualitative models based on genetic mechanism analysis. Subsequent interdisciplinary advancements introduced semi-quantitative approaches, including Fisher discriminant analysis (Huang et al., 2011; Jiang and Jiang, 2012) and the Analytic Hierarchy Process (AHP) (Jiyuan et al., 2021; YanHua et al., 2022). Additionally, quantitative frameworks such as fuzzy logic systems (Zhang et al., 2021; Zhuang, 2022) and grey system theory (Qiu, 2004; Meng et al., 2009; Gao et al., 2018; Ding et al., 2019) were developed. More recently, machine learning techniques have significantly enhanced prediction accuracy by modeling complex nonlinear relationships. Techniques such as backpropagation neural networks (BPNNs) (Bao and Hu, 2002; Chen et al., 2005; Yang et al., 2023; Journal of Chemistry, 2023), entropy-weighted cloud models (Chen et al., 2019; Wei et al., 2021), support vector machines (SVMs) (Lai and Qiao, 2008; En et al., 2011; Zhou et al., 2020), random forests (RFs) (Wang et al., 2022; Ren et al., 2023), and extreme learning machines (ELMs) (Xie et al., 2021) have been widely applied. Despite these advancements, current methodologies face several limitations. First, conventional statistical methods (e.g., AHP and fuzzy analysis) rely heavily on expert judgment, leading to subjective biases and challenges in determining objective weights. Second, data-driven models (e.g., BPNN and RF) often experience optimization challenges, including convergence to local minima and sensitivity to parameter selection. Additionally, traditional ELM implementations suffer from instability due to random weight initialization, compromising their reliability in engineering applications.

To address these constraints, in this study, we propose an optimized prediction framework (ISSA-ELM) that integrates an improved sparrow search algorithm (ISSA) with the extreme learning machine architecture. The method uses a three-fold optimization strategy: First, principal component analysis (PCA) reduces the dimensionality of collapse-influencing factors, minimizing multicollinearity while retaining essential diagnostic information. Second, Singer chaotic mapping enhances the standard sparrow search algorithm (SSA) by improving population diversity, expanding the search space, and preventing premature convergence. Third, an adaptive optimization framework combines the refined ISSA with K-fold cross-validation to optimize ELM hyperparameters, ensuring robust model generalization. Empirical validation through case studies in representative karst regions demonstrates that the ISSA-ELM outperforms conventional methods in predictive accuracy, particularly for susceptibility zonation. This methodological advancement offers a novel paradigm for regional-scale geological risk management, providing both

theoretical insights and practical values for geological disaster prevention and land-use planning.

2 Theory of the ISSA-ELM integrated model

2.1 Principal component analysis

PCA, initially proposed by Pearson (1901), is a widely adopted dimensionality reduction technique. It converts multiple indicators into a smaller set of comprehensive variables (Zhou et al., 2020; Cao et al., 2015). This approach not only reduces data complexity but also preserves key relationships among variables. The main steps of PCA are as follows.

2.1.1 Data standardization

Given n samples and p influencing factors, the original data matrix is represented as shown in Equation 1:

$$Z = (z_{ij})_{n \times p}, (i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, p). \quad (1)$$

Standardization is applied using the formula as shown in Equation 2:

$$Z^* = (z_{ij}^*)_{n \times p}, (i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, p), \quad (2)$$

where

$$z_{ij}^* = \frac{z_{ij} - \frac{1}{n} \sum_{i=1}^n z_{ij}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2}}, (i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, p). \quad (3)$$

In Equation 3, z_{ij} is the observed value of the j th variable of the i th sample, and \bar{z}_j is the sample mean of the j th variable.

2.1.2 Correlation coefficient matrix calculation

The correlation coefficient matrix R can be calculated using Equation 4:

$$R = \frac{1}{n-1} Z^{*T} Z^*, \quad (4)$$

where Z^* represents the standardized data matrix.

2.1.3 Eigenvalue and eigenvector computation

The eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p$ of the correlation coefficient matrix R are calculated, and their corresponding orthogonal eigenvectors are as follows:

$$a_i = (a_{i1}, a_{i2}, a_{i3}, \dots, a_{ip})^T, (i = 1, 2, 3, \dots, p).$$

2.1.4 Variance contribution rate and cumulative Variance contribution rate

The variance contribution rate β_i and the cumulative variance contribution rate $\beta(i)$ can be calculated using Equations 5, 6:

$$\beta_i = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i}, (i = 1, 2, 3, \dots, p), \quad (5)$$

$$\beta(i) = \sum_{i=1}^p \beta_i, (i = 1, 2, 3, \dots, p). \quad (6)$$

2.1.5 Principal component selection

Principal components are selected when the cumulative variance contribution rate reaches 85% or higher, ensuring most of the original data information is retained (Zhou et al., 2020).

2.2 Sparrow search algorithm

The SSA is a swarm intelligence optimization algorithm introduced by Xue and Shen (2020). It simulates the foraging and anti-predation behaviors of sparrow flocks to find optimal solutions. The sparrow population consists of discoverers, followers, and sentinels, represented by an $n \times d$ matrix (as shown in Equation 7), where n is the number of sparrows, d denotes the dimension, and x_{nd} represents the j th dimension of the i th sparrow. The fitness value for each sparrow is represented as F_x (as shown in Equation 8).

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}, \quad (7)$$

$$F_x = \begin{bmatrix} f([x_{11} & x_{12} & \cdots & x_{1d}]) \\ f([x_{21} & x_{22} & \cdots & x_{2d}]) \\ \vdots \\ f([x_{n1} & x_{n2} & \cdots & x_{nd}]) \end{bmatrix}. \quad (8)$$

2.2.1 Discoverer update

Discoverers locate new food sources using Equation 9:

$$X_{ij}^{t+1} = \begin{cases} X_{ij}^t \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{max}}\right), & R_2 < ST \\ X_{ij}^t + Q \cdot L, & R_2 \geq ST \end{cases}, \quad (9)$$

where $iter_{max}$ denotes the maximum number of iterations; X_{ij}^t is the position of the i th sparrow in the j th dimension at the t -th iteration; $\alpha \in (0, 1]$, $Q \in [0, 1]$, and $R_2 \in [0, 1]$ are random numbers, representing the warning values, respectively; L is a matrix of ones; and ST takes values between $[0.5, 1]$, representing the safety value.

When $R_2 \geq ST$, it indicates that the foraging area is not safe, and all sparrows need to fly to other safe foraging locations.

2.2.1.1 Follower update

Followers adjust their positions by tracking discoverers to accelerate convergence using Equation 10:

$$X_{ij}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^t - X_{ij}^t}{i^2}\right), & i > \frac{n}{2} \\ X_p^{t+1} + |X_{ij}^t - X_p^{t+1}| A^+ \cdot L, & i \leq \frac{n}{2} \end{cases}, \quad (10)$$

where X_p^{t+1} is the best position where the discoverers are located, X_{worst}^t is the current globally worst position, and $A^+ = A^T \cdot (A \cdot A^T)^{-1}$ is a matrix, with each element being 1 or -1.

When $i > \frac{n}{2}$, it indicates that the i th follower has not obtained food and needs to move to other places to forage.

2.2.1.2 Sentinel update

Sentinels monitor threats and adjust their positions using Equation 11:

$$X_{ij}^{t+1} = \begin{cases} X_{best}^t + \theta |X_{ij}^t - X_{best}^t|, & f_i > f_g \\ X_{ij}^t + K \left| \frac{X_{ij}^t - X_{worst}^t}{(f_i - f_w) + \varepsilon} \right|, & f_i = f_g \end{cases}, \quad (11)$$

where X_{best}^t is the global optimal position in the sparrow population at the t -th generation, ε is a constant, θ is the step-size control coefficient, $K \in [-1, 1]$ is a random number, f_i is the fitness value of the i th sparrow currently, f_g is the current global best fitness value, and f_w is the current global worst fitness value.

When $f_i > f_g$, it indicates that the sparrow is currently at the edge of the population and is vulnerable to predator attacks; when $f_i = f_g$, it indicates that the sparrows in the middle of the population are aware of the danger of being preyed upon and need to move closer to other sparrows.

2.3 Improved sparrow search algorithm

The conventional SSA often experiences issues related to local convergence and reduced population diversity in later iterations (Lv et al., 2021; Liu et al., 2022). To mitigate these challenges, the Singer chaotic mapping is introduced to enhance population initialization. The formula for Singer chaotic mapping is shown in Equation 12:

$$X_{k+1} = u(7.86x_k - 23.32x_k^2 + 28.75x_k^3 - 13.3.2875x_k^4), \quad (12)$$

where $u \in (0.9, 1.08)$ is a control parameter and $x_k \in [0, 1]$ is a random number.

As shown in Figure 1, the value range of the Singer mapping lies within the interval $[0, 1]$. By replacing the random initialization method with its chaotic characteristics, the search space can be made more uniform, enhancing the uniformity and diversity of the sparrow population, and thereby improving the global search ability.

The steps of the improved SSA are outlined as follows:

- (1) Use the chaotic mapping to initialize the population, the number of iterations, and the proportions of discoverers and followers.
- (2) Compute the fitness values, and arrange them in order.
- (3) Update the positions of discoverers.
- (4) Update the positions of followers.
- (5) Update the positions of sentinels.
- (6) Recalculate the fitness values, and update the positions of the sparrows.
- (7) Repeat until the maximum number of iterations is reached.

The flowchart of the improved SSA is presented in Figure 2.

2.4 Extreme learning machine

The ELM is a computationally efficient algorithm designed for training single-hidden-layer feedforward neural networks (SLFNs).

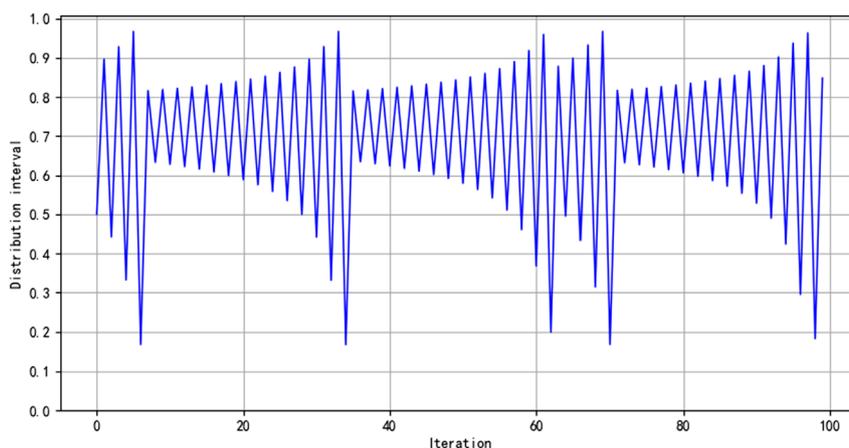


FIGURE 1 Distribution of the Singer map after 100 iterations.

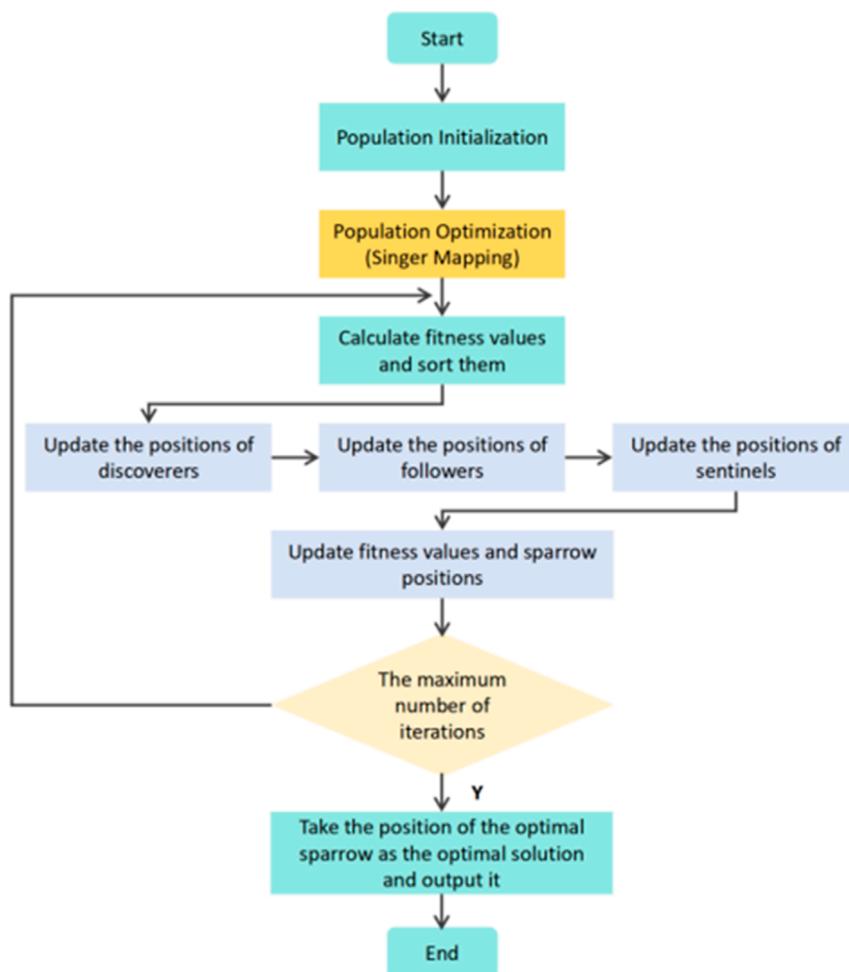
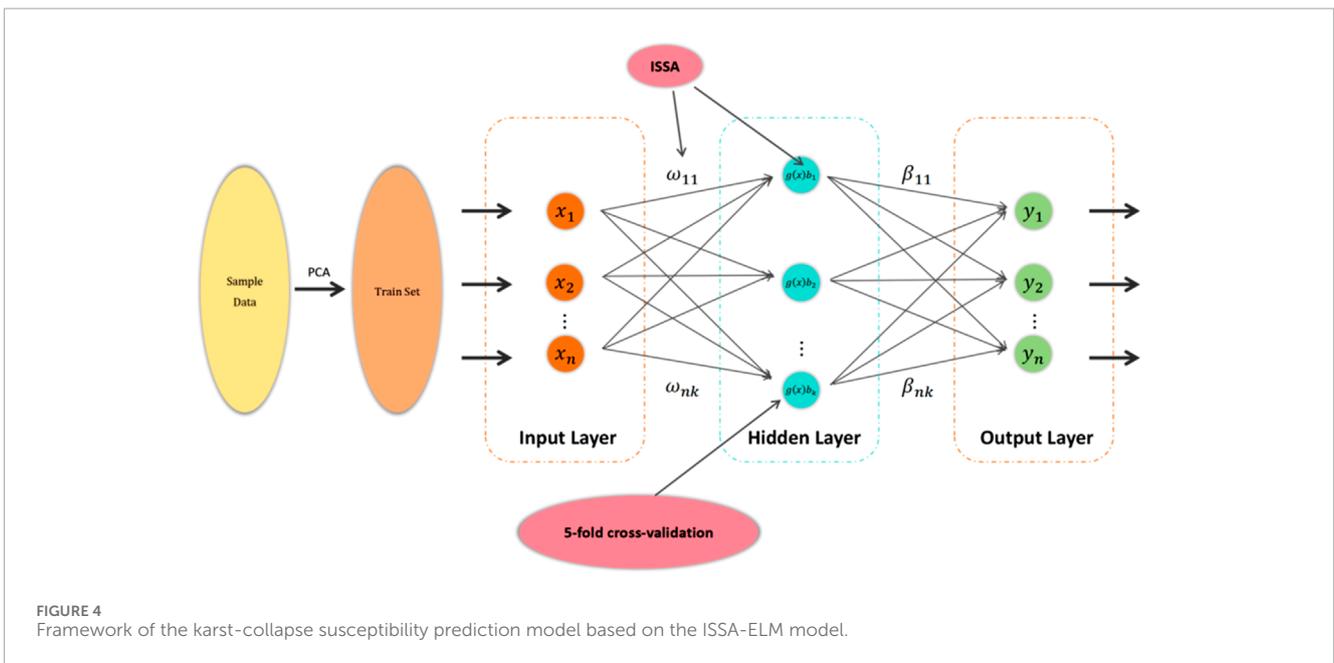
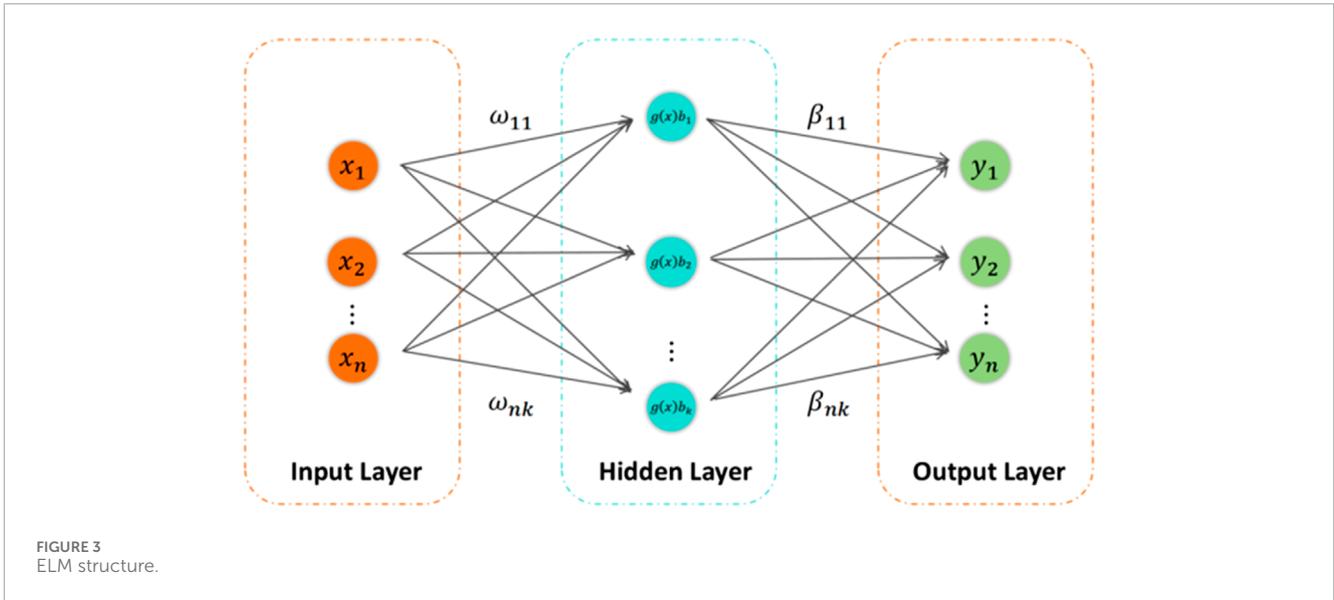


FIGURE 2 Flowchart of the ISSA.



First introduced by Huang et al. (2006), ELM offers significant improvements in training speed and generalization capability. Its applications are widespread across various domains, including classification, regression, and prediction tasks.

The ELM algorithm follows a simple structure, as illustrated in Figure 3, comprising an input layer, a hidden layer, and an output layer. Unlike traditional neural networks, ELM assigns random values to the input-layer weights and hidden-layer threshold. The output-layer weights are then calculated using a generalized matrix, enabling efficient and accurate chaotic recognition outcomes.

The mathematical representation of the output model for a single-hidden-layer neural network with L hidden-layer nodes and an activation function $g(x)$ is given in Equation 13:

$$y_j = \sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i), (j = 1, 2, 3, \dots, n), \tag{13}$$

where x_j denotes the input value, y_j represents the actual output value, ω_i is the input weight matrix, b_i denotes the hidden-layer threshold, β_i is the output weight matrix, and $g(x)$ is typically a sigmoid function (Jiang and Shi, 2023)

The primary objective of the ELM algorithm is to minimize the output error, which can be expressed in Equation 14:

$$\|H\beta - T\| = \min, \tag{14}$$

where there exist β_i , ω_i , and b_i such that

$$H\beta = T. \tag{15}$$

This relationship shown in Equation 15 consists of N equations that can be represented in a matrix form as shown in Equation 16:

$$H = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & \cdots & g(\omega_L \cdot x_1 + b_L) \\ g(\omega_1 \cdot x_2 + b_1) & \cdots & g(\omega_L \cdot x_2 + b_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1 \cdot x_N + b_1) & \cdots & g(\omega_L \cdot x_N + b_L) \end{bmatrix}, \quad (16)$$

where H is the hidden-layer output matrix, β is the weight matrix connecting the hidden and output layers, and T represents the desired output.

The least-squares solution for the output-layer weights can be efficiently obtained using the Moore–Penrose generalized inverse matrix H^+ , as shown in Equation 17:

$$\beta = H^+ T. \quad (17)$$

A notable limitation of ELM is its sensitivity to the random initialization of input-layer weights and hidden-layer threshold, leading to potential instability. To mitigate this issue, the ISSA is applied for parameter optimization. By enhancing the search process and avoiding local optima, the ISSA improves both the accuracy and robustness of the ELM model.

2.5 Hyperparameter optimization of the prediction model

The ELM algorithm randomly assigns input-layer weights and hidden-layer threshold. Without appropriate adjustment, this randomness may reduce the functionality of certain hidden-layer nodes, leading to algorithmic instability. Furthermore, although a larger number of hidden-layer nodes can enhance performance, an excessive number of nodes may result in overfitting, whereas too few nodes can cause underfitting, thereby reducing the prediction accuracy (Xie et al., 2021).

To mitigate these issues, in this study, we employ a combined approach using the ISSA and K-fold cross-validation to optimize the hyperparameters of the ELM prediction model. The optimized parameters include input-layer weights, hidden-layer thresholds, and the number of hidden nodes (N). The hyperparameter optimization process consists of the following steps:

- (1) Initialization: set the sparrow population size, maximum number of iterations, proportions of initial discoverers and followers, and the number of hidden nodes.
- (2) Fitness function calculation: define the mean squared error (MSE) between the predicted and actual values as the model's fitness function, represented in Equation 18:

$$\text{Fitness} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2, \quad (18)$$

where y_i represents the actual grade of karst-collapse susceptibility of the i th sample, \bar{y}_i denotes the predicted value for the i -th sample, and n is the number of samples.

- (3) Ranking and selection: compute the fitness values and rank individuals. Those with higher fitness values are designated as discoverers in the sparrow population.
- (4) Position update: update the positions of discoverers, followers, and vigilant individuals using Formulas 9–11, and then, recalculate fitness values.
- (5) Iteration check: assess whether the maximum number of iterations is reached. If so, terminate training and output the global optimal parameters (optimal weights, hidden-layer biases, and N). Otherwise, repeat steps (2–4).
- (6) K-fold cross-validation: perform five-fold cross-validation to determine the optimal number of hidden nodes. The data are split into five subsets, with each subset used once as a validation set, whereas the remaining subsets form the training set. The optimal number of hidden nodes is determined by minimizing the average error across the five models, as shown in Equation 19:

$$\text{MAE} = \frac{1}{5} \sum_{i=1}^n (y_i - \bar{y}_i). \quad (19)$$

- (7) Output: provide the final optimal parameter combination, including input-layer weights, hidden-layer threshold, and the number of hidden nodes.

2.6 Prediction process of karst collapse susceptibility using the ISSA-ELM model

The ISSA-ELM-based karst-collapse susceptibility prediction model is illustrated in Figure 4, with its overall flowchart depicted in Figure 5. The prediction process consists of the following steps:

- (1) Data collection and preprocessing: Real-world data are collected from key karst collapse exploration areas. Data preprocessing is performed by removing outliers and imputing missing values using the mode. PCA is then applied for dimensionality reduction.
- (2) Data division: The dataset is split into a training set and a test set. The training set is used to train the ISSA-ELM model, whereas the test set is used for prediction and model evaluation.
- (3) Hyperparameter optimization: The ELM model's hyperparameters, including the sparrow population size, maximum iterations, proportions of discoverers and followers, and the number of hidden nodes, are initialized. The initial model is trained and evaluated using MSE and MAE. The optimal hyperparameters are determined when the error reaches its minimum.
- (4) Prediction: The ISSA-ELM model, with the optimized hyperparameters, is used to predict karst-collapse susceptibility. The final model parameters are applied for prediction tasks.

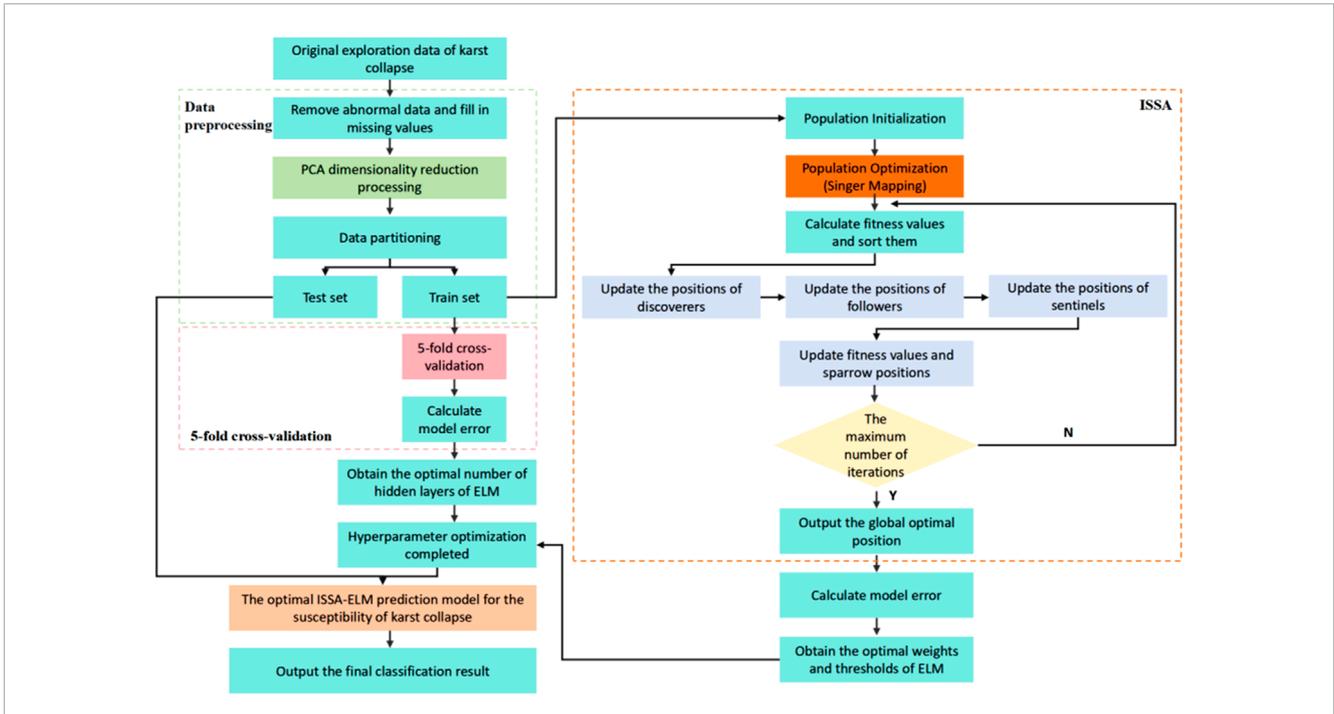


FIGURE 5 Overall flowchart of the karst-collapse susceptibility prediction model based on the ISSA-ELM model.

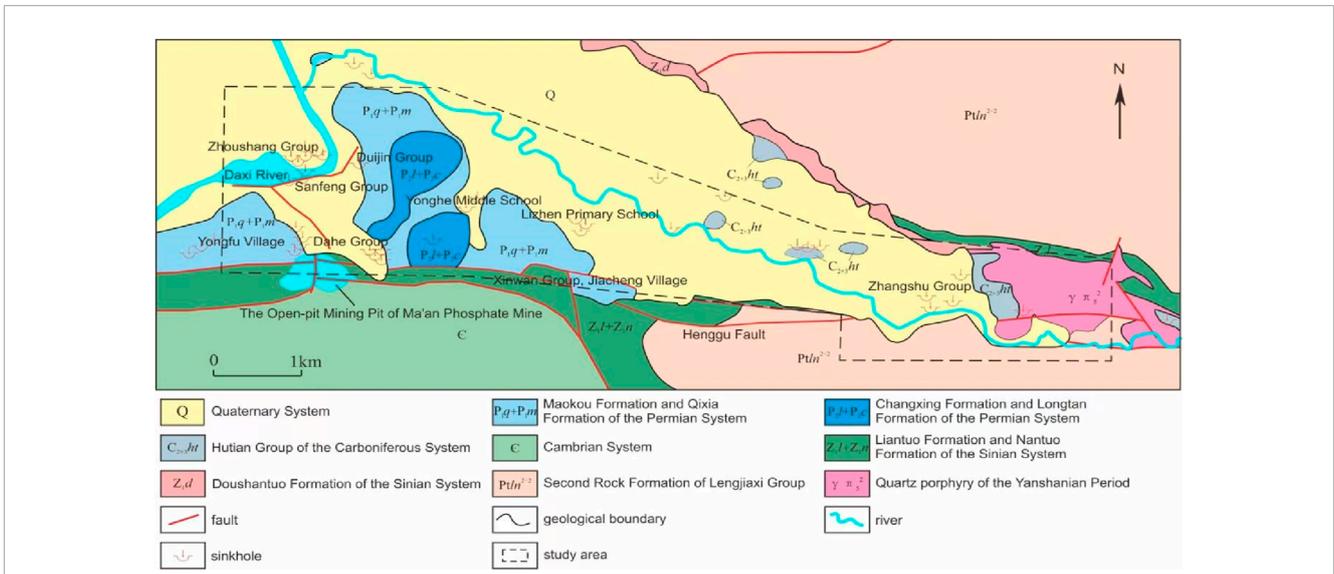


FIGURE 6 Relationship diagram among karst development, strata, and faults in the study area.

2.7 Evaluation methods

To comprehensively evaluate the predictive performance of various karst-collapse susceptibility models, the following indicators are applied.

2.7.1 Coefficient of determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{20}$$

In Equation 20, y_i is the actual value of the i th sample, \hat{y}_i is the predicted value of the i th sample, \bar{y} is the mean of the actual values, and n is the total number of samples. A coefficient value closer to 1 indicates a better model fit.

2.7.2 Mean squared error

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (21)$$

2.7.3 Mean absolute error

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (22)$$

Smaller MSE (Equation 21) and MAE (Equation 22) values indicate a more accurate model, with the predicted values closely matching the actual values.

3 General situation of the project

3.1 Overview of the geological environment

The study area is located in Yonghe Town, in the northeastern part of Liuyang City, Hunan Province. The terrain generally slopes from higher elevations in the south and northeast to lower elevations in the central and western regions. The primary geological and environmental conditions are as follows:

- (1) Physical geography: the region is characterized by denudation–erosion hilly landforms and dissolution–erosion valley landforms. The hills are primarily located in the southern part and the Qibaoshan area to the east. These positive landforms are composed mainly of sandstones, slates, and dolomites from the Upper Sinian System (Zb), Cambrian System (ϵ), Longtan Formation (P₂l), and Changxing Formation (P₂c) of the Upper Permian System. In contrast, negative landforms composed of dolomites, marls, and limestones interbedded with shales are prevalent in the northern and western parts. These belong to the Hutian Group (C₂₊₃ht) of the Middle and Upper Carboniferous Systems and the Qixia Formation (P₁q) of the Lower Permian System.

Dissolution–erosion valley landforms are dominant in the northern and western regions, characterized by broad valleys and alluvial plains formed through erosional and depositional processes. These areas are covered by upper silty clay and sandy gravel layers from the Holocene Series (Q_h^{al}) of the Quaternary System. The underlying bedrock mainly comprises marls, limestones interbedded with shales, and dolomites. The area experiences an average annual rainfall of 1574.7 mm, with maximum and minimum recorded rainfalls of 2096.1 mm and 1079.1 mm, respectively. Rainfall is most concentrated in June, contributing over 16% of the annual total. The Daxi River, the second largest tributary of the Liuyang River, flows from north to west across the study area (Figure 7).

- (2) Stratigraphic lithology: the stratigraphic sequence in the study area, from the oldest to the youngest, includes the second segment of the second lithologic unit of the Lengjiaxi Group (Ptln²⁻²), the Liantuo Formation (Z₁l) and Nantuo Formation (Z₁n) of the Lower Sinian System, the Doushantuo Formation (Z₂d) of the Upper Sinian System, the Cambrian System (ϵ), the Hutian Group (C₂₊₃ht) of the Middle and Upper Carboniferous Systems, the Qixia Formation (P₁q) and Maokou Formation (P₁m) of the Lower Permian System, the Longtan Formation (P₂l) and Changxing Formation (P₂c) of the Upper Permian System, and the Quaternary System (Q) (Figure 6).

Soluble rock strata are predominantly found in the Changxing Formation, the Qixia Formation, and the Hutian Group. The Hutian Group and the first (P₁q¹) and second (P₁q²) segments of the Qixia Formation consist of interbedded limestones, shales, argillaceous limestones, and dolomites, displaying well-developed karst features.

- (3) Geological structure: the Yonghe Syncline is a minor fold structure oriented in a north–northwest direction. The syncline core consists of the Changxing Formation, with the wings comprising the Longtan Formation, the Maokou Formation, and the Qixia Formation. The strata dip gently, with angles ranging from 10° to 20°.

Several faults are present, including two strike-slip faults, two reverse faults, one transverse thrust fault, and two unidentified faults. The most prominent is the transverse thrust fault, which traverses the northern edge of the Ma'anling Phosphate Mine (Figure 6).

- (4) Hydrogeological conditions: the groundwater in the area consists of karst water, bedrock weathering fissure water, and pore water within loose rock formations.

3.2 Overview of karst collapses

A total of 38 karst collapses were documented within the 14-km² exploration area (Figure 7), resulting in a collapse density of 2.71 occurrences per km². These collapses are primarily concentrated in regions with well-developed soluble rocks, significant fault activity, and strong surface water–groundwater interactions. The most affected areas include the Yueshan–Ou's–Dahe area of Yongfu Village; the Juxiang Community, Yonghe Old Street, Huayuan Village of Yanxi Town; and the Nanshan–Jiacheng Village, Li Zhen Primary School (Old), Xinwan–Xinping regions.

Most collapses are small-scale, with long-axis lengths ranging from 1 to 7 m, accounting for 73% of incidents (Figure 8).

Xinwu Group, Jiacheng Village collapse (Figure 9a): this circular collapse pit has a long-axis length of approximately 3.5 m, a short-axis length of 2.5 m, and a depth of 1.2 m. The pit contains water, with the water level approximately 0.2 m below the surface. Bedrock is not visible, and the surrounding area consists of alluvial gravel-bearing silty clay, negatively impacting agricultural activities.

Xinwan Group, Jiacheng Village collapse (Figure 9b): this collapse is nearly circular, with a long-axis length of 2.3 m, a short-axis length of 2.2 m, and a depth of 10 m. The water level inside the pit is approximately 5 m. The primary triggering factors were

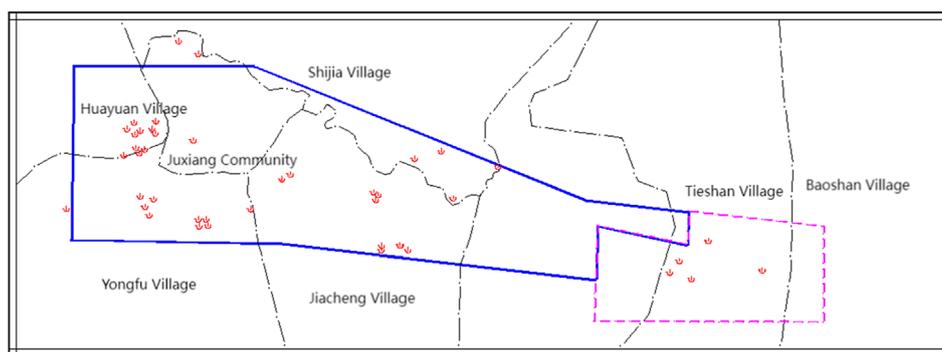


FIGURE 7 Distribution map of karst-collapse points in the exploration area (the blue solid line represents the scope of the exploration area, and the dotted line represents the scope of the prediction area).

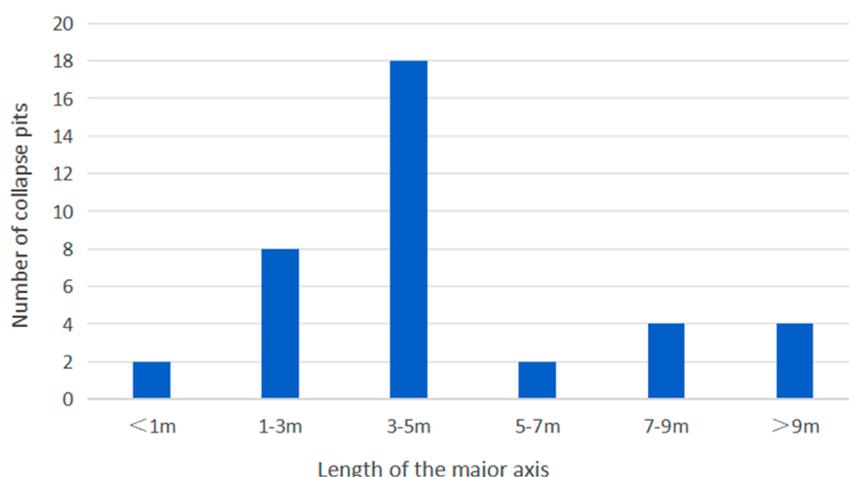


FIGURE 8 Scale situation of the long axis of the collapse pits.

groundwater extraction and heavy rainfall. Situated on a river terrace, the surrounding soil consists of silty clay with pebbles, forming a single-layer structure approximately 1 m thick. The underlying bedrock is limestone. Due to its instability, this collapse presents a continuous risk to local production and daily life.

3.3 Formation mechanism of karst collapses

The occurrence of karst collapses results from the complex interplay and cumulative effects of multiple formation mechanisms (Hu et al., 2001; Chen and Cao, 2023; Feng, 2025). Based on comprehensive investigations, in-depth research, and detailed exploration within the study area, the primary formation conditions include karst development characteristics, overburden properties, and hydrodynamic features.

(1) Impact of the karst development degree on karst collapses: karst development serves as the essential foundation and

prerequisite for the occurrence of karst collapses. Statistical analysis indicates that most collapse pits in the study area are located within or near intense karst development zones. Notably, collapses are most prevalent in strata with well-developed karst, particularly within the second segment of the Qixia Formation (P_1q^2) of the Lower Permian System and the Hutian Group ($C_{2+3}ht$) of the Middle and Upper Carboniferous Systems. This strong correlation underscores the critical influence of karst development on the formation of collapses.

(2) Influence of overburden characteristics on karst collapses: the overburden comprises Quaternary loose soil layers that cover the bedrock surface. Collapse pits are primarily concentrated in covered karst regions, where the thickness of the Quaternary loose layer generally does not exceed 20 m. Areas with alluvial and diluvial deposits that lack a clay layer at the base are particularly vulnerable to collapse formation. From a geotechnical perspective, these soil layers exhibit low shear strength and limited resistance to seepage-induced



FIGURE 9

Examples of planar morphological characteristics of karst-collapse pits. (a) Karst collapse in the Xinwu Group of Jiacheng Village (T23). (b) Karst collapse in the Xinwan Group of Jiacheng Village (T27). (c) Karst collapse in the Xinping Group of Jiacheng Village (T30). (d) Karst collapse in the Shixing Group of Shijia Village (T32).

deformation, making them susceptible to piping processes and subsequent collapses.

For instance, in the collapse-affected section spanning the Zhoushang Group of Huayuan Village, Juxiang Community, Yonghe Old Street, Mingshan Group of Yongfu Village, Yueshan, Ou's Group, and Dahe Group, the soil structure consists predominantly of gravel-bearing silty clay formed through alluvial and diluvial processes, along with residual slope-derived crushed stone soil. The total thickness of the soil layer ranges from 5.5 to 13.5 m. In these areas, the upper silty clay layer is relatively thin, and the absence of a water-retaining clay layer at the base facilitates the infiltration of atmospheric precipitation, surface water seepage, and the upward movement of water and sand during the increase in the karst water level. Consequently, the thickness and physical properties of the overburden significantly regulate the collapse process.

- (3) Effect of hydrodynamic characteristics on karst collapses: the study area contains two distinct groundwater layers: the upper-layer Quaternary pore water and the lower-layer karst water. The upper-layer pore water is primarily stored in silty clay, pebble-bearing soil, and residual crushed stone soil from Quaternary alluvial and diluvial facies. Recharge sources include atmospheric precipitation and seepage from nearby surface water bodies such as the Daxi River. During recharge, seepage forces can trigger piping within the soil layer, leading to cavity formation, which may expand and eventually result in collapse.

The lower-layer karst water mainly originates from the limestone aquifer of P_1q and the limestone karst-cave aquifer of $C_{23}ht$, both exhibiting significant karst development and high water-bearing capacity. Groundwater extraction often lowers the karst water level, prompting the continuous recharge of upper-layer pore water into the karst aquifer. Fine-grained soil particles are transported by the flowing water, further accelerating the collapse process. In conclusion, the coexistence of the two-layer groundwater system, the presence of a significant hydraulic gradient, and the infiltration of surface water bodies jointly alter the hydrodynamic characteristics of the region, playing a pivotal role in the formation of karst collapses.

3.4 Determination of influencing factors for the susceptibility of karst collapses

The development of karst collapses is influenced by the combined effects of multiple factors. Nonlinear relationships exist not only among these factors but also between each factor and the occurrence of karst collapses (Chen and Cao, 2023). Through extensive on-site investigations and comprehensive analyses of the formation mechanisms of numerous karst collapses, this study identified the key factors for predicting karst-collapse susceptibility.

3.4.1 Karst development degree

Soluble rocks provide the material basis for karstification, with their solubility directly impacting the extent of karst development. In the region extending from the Zhoushang Group of Huayuan Village to Juxiang Community, Yonghe Old Street, Yueshan Group, Ou's Group, and Dahe Group, 29 karst collapses have occurred. Borehole data from 15 boreholes revealed karst caves in five of them, yielding a karst-cave encounter rate of 33.3%. Among the six identified caves, five developed within the second member of the Lower Permian Qixia Formation (P_1q^2), characterized by thick-bedded dark gray limestone with a faint flesh-red tinge, interbedded with thin layers of marl and dolomite. Geophysical interpretation indicated that 76.98% of the 505 detected anomalies were karst related, predominantly in the Qixia Formation.

3.4.2 Distance from intra-area faults

Fault structures provide conduits for groundwater flow and play a decisive role in karst development. On-site investigations revealed that karst collapses are concentrated along concealed and unidentified faults on both sides of the NW-trending secondary faults within the northern plate of the Henggu Fault. Notably, four collapses in Yongfu Village, Yonghe Town, occurred approximately 200 m west of a concealed fault. This positive correlation highlights the influence of fault zones on collapse occurrence.

3.4.3 Thickness of the overlying soil layer

The development and distribution of karst collapses are closely linked to the thickness of the Quaternary layer. Survey results

indicated that the Quaternary layer in the Dahe Group, Yueshan Group, Ou's Group, and Mingshan Group of Yongfu Village, as well as in the Zhoushang Group, Juxiang Community, and Yonghe Old Street in Huayuan Village, ranged from 5.50 to 13.5 m. Additionally, the thickness of the upper silty clay layer was generally ≤ 3 m. The presence of a thin silty clay layer facilitates water infiltration and groundwater-level fluctuations, increasing the likelihood of collapses.

3.4.4 Groundwater fluctuation range

Karst water fluctuates between confined and unconfined states near the bedrock surface, contributing to collapse occurrence. Survey data from the Yueshan Group and Ou's Group regions indicated that groundwater levels ranged from 9.27 to 25.90 m during the wet season, with the bedrock surface buried at depths of 15.00–30.70 m. During the dry season, groundwater levels decrease by 5–10 m, maintaining their position near the bedrock surface. This fluctuation promotes the “bellows effect,” characterized by piping and suction erosion of the overlying soil, which increases the risk of collapse.

3.4.5 Distance from the center of the artificial pumping funnel

Before October 2016, mining at the Ma'an Phosphate Mine reached an elevation of approximately ± 0 m, affecting an area within a 2-km radius, primarily around Huayuan Village, Juxiang Community, Yongfu Village, and Yonghe Middle School. Investigations documented 89 instances of house and ground cracking, accounting for 44.9% of all cases, alongside 22 collapse pits (57.9%). Since the cessation of mining activities, a significant decrease in collapses and cracking incidents has been observed, underscoring the impact of groundwater extraction on collapse susceptibility.

3.4.6 Development density of existing ground collapses

A total of 38 ground collapses were identified within the study area, distributed as follows: 16 collapses in the Mingshan Group, Yueshan Group, Ou's Group, and Dahe Group of Yongfu Village; 12 in the Zhoushang Group, Juxiang Community, and Yonghe Old Street in Huayuan Village; three in the Yonghe Middle School and Nanshan Group area of Jiacheng Village; seven in the Li Zhen Primary School (Old), Xinwan Group, and Xiping Group area; and four in the Tieshan Village and Qibaoshan Village area. The density of existing collapses was considered a key predictive indicator.

Following the relevant guidelines (Ministry of Natural Resources of the People's Republic of China, 2023) and building on previous studies (Huang et al., 2011; YanHua et al., 2022), in this study, we integrated field survey results, geophysical exploration data, and borehole data to identify six key factors influencing karst-collapse susceptibility (Table 1). Through a comprehensive analysis of the field survey data, the susceptibility levels of karst collapse were assessed. To effectively represent the likelihood of collapse occurrences, the susceptibility levels were classified into three categories: low susceptibility, medium susceptibility, and high susceptibility, corresponding to category labels 1, 2, and 3, respectively (Table 2).

4 Case analysis

4.1 Data collection and preprocessing

The karst-collapse-affected area in Yonghe Town, Liuyang City, Hunan Province, was selected as the research subject. A total of 20 sets of sample data were collected within the exploration zone, comprising 16 samples from collapsed pits and four samples from non-collapsed pits representing locations with no recent collapses within 1 km of the sampling points and also considered similar geological settings (like lithology and stratigraphy) and hydrogeological conditions (groundwater levels, flow, and quality) (Table 3). This dataset was used to evaluate the feasibility of the ISSA-ELM model for predicting the karst-collapse susceptibility.

Given the complex interrelationships among various influencing factors of karst collapses, using all factors as inputs could introduce information redundancy. To mitigate this, PCA was applied to reduce the dimensionality and remove correlations among six key factors: the degree of karst development, the distance from faults, the thickness of the overlying soil layer, the groundwater fluctuation range, the distance from the center of the artificial pumping funnel, and the development density of preexisting ground collapses. This approach retained the essential information while minimizing redundancy.

The original data were standardized to eliminate the impact of differing units and magnitudes. A standardized covariance matrix was then calculated (Table 4), followed by an in-depth correlation analysis to explore the relationships among the influencing factors. Eigenvalues and contribution rates were derived from the covariance matrix, and principal components were selected based on a cumulative contribution rate exceeding the widely accepted threshold of 85% (Table 5).

As shown in Table 5, the cumulative contribution rate of the first five principal components was 96.87%, surpassing the 85% threshold. Therefore, these five principal components were chosen to represent the original six variables.

Based on the eigenvector matrix (Table 6), the principal components were expressed as follows:

The karst-collapse-affected area in Yonghe Town, Liuyang City, Hunan Province, was selected as the research subject. A total of 20 sets of sample data were collected within the exploration zone, comprising 16 samples from collapsed pits and four samples from non-collapsed pits representing locations with no recent collapses within 1 km of the sampling points (Table 3). This dataset was used to evaluate the feasibility of the ISSA-ELM (improved sparrow search algorithm-extreme learning machine) model for predicting karst-collapse susceptibility.

Given the complex interrelationships among various influencing factors of karst collapses, using all factors as inputs could introduce information redundancy. To mitigate this, PCA was applied to reduce dimensionality and remove correlations among six key factors: the degree of karst development, the distance from faults, the thickness of the overlying soil layer, the groundwater fluctuation range, the distance from the center of the artificial pumping funnel, and the development density of preexisting ground collapses. This approach retained essential information while minimizing redundancy.

TABLE 1 Classification of influencing factors for karst-collapse susceptibility.

Factor	Karst-collapse susceptibility levels		
	Low-susceptibility criteria	Medium-susceptibility criteria	High-susceptibility criteria
Drilling saturation/%	<3	3~10	>10
Distance to faults/m	>100	50~100	0~50
Overburden thickness/m	Total >20 m; upper clay >3 m	10< Total ≤20 m; upper clay ≤3 m	Total ≤10 m; upper clay ≤3 m
Groundwater fluctuation/m·a ⁻¹	<3	3~8	>8
Distance to pumping center/m	>500	200~500	0~200
Collapse density (per 10 km ²)	<2	2~10	>10

TABLE 2 Class labels for the grading of expected output parameters.

Susceptibility level	Low	Medium	High
Label	1	2	3

The original data were standardized to eliminate the impact of differing units and magnitudes. A standardized covariance matrix was then calculated (Table 4), followed by an in-depth correlation analysis to explore the relationships among the influencing factors. Eigenvalues and contribution rates were derived from the covariance matrix, and principal components were selected based on a cumulative contribution rate exceeding the widely accepted threshold of 85% (Table 5).

As shown in Table 5, the cumulative contribution rate of the first five principal components was 96.87%, surpassing the 85% threshold. Therefore, these five principal components were chosen to represent the original six variables.

Based on the eigenvector matrix, the principal components were expressed as follows:

$$X_1 = -0.0407Z_1^* + 0.6521Z_2^* - 0.0783Z_3^* + 0.3718Z_4^* + 0.6459Z_5^* + 0.1069Z_6^*$$

$$X_2 = -0.5044Z_1^* - 0.0975Z_2^* - 0.6139Z_3^* + 0.4390Z_4^* - 0.2019Z_5^* - 0.3543Z_6^*$$

$$X_3 = 0.3891Z_1^* - 0.1321 + 0.2367Z_3^* + 0.1785Z_4^* + 0.2218Z_5^* - 0.8330Z_6^*$$

$$X_4 = -0.0242Z_1^* - 0.2971Z_2^* + 0.4789Z_3^* + 0.7582Z_4^* - 0.1295Z_5^* + 0.2999Z_6^*$$

$$X_5 = 0.1712Z_1^* + 0.6665Z_2^* + 0.1441Z_3^* + 0.1148Z_4^* - 0.6869Z_5^* - 0.1431Z_6^*$$

where Z_1^* , Z_2^* , Z_3^* , Z_4^* , Z_5^* , and Z_6^* represent the standardized values of the degree of karst development, the distance from faults, the thickness of the overlying soil layer, the groundwater fluctuation range, the distance from the pumping funnel center,

and the development density of existing ground collapses, respectively.

The principal components were subsequently calculated using these expressions (Table 7) and used for training the ISSA-ELM model.

4.2 Optimization of model parameters

From the PCA-processed dataset, 14 samples were designated as the training set, whereas the remaining six samples were used as the test set. This 7:3 split ensured a balanced approach, maintaining sufficient data for training while reserving a representative subset for evaluating the model's generalization performance.

The ISSA-ELM model parameters were configured as follows: population size $N = 20$, maximum iterations = 100, warning threshold $ST = 0.6$, proportion of discoverers $PD = 0.7$, and proportion of scouts $SD = 0.2$.

The number of hidden-layer nodes significantly affects model performance. In the ELM model, the number of input-layer nodes was set to 5, corresponding to the number of principal components. The sigmoid function was applied as the activation function. The number of hidden-layer nodes was selected within the range $[2n+1, N]$ (where n is the number of nodes in the input layer $n = 5$, and N denotes the total number of input samples $N = 20$), resulting in a range of $[11, 20]$ (Jiang and Shi, 2023). A five-fold cross-validation approach was used to determine the optimal number of hidden-layer nodes, balancing computational efficiency and robust evaluation. The MAE index results (Figure 10) indicated that the optimal number of hidden layers was 12 as it produced a concentrated MAE distribution without significant outliers.

The PCA-processed data were input into the ISSA-ELM and SSA-ELM models for iterative comparisons. The resulting optimization curves (Figure 11) showed that the ISSA-ELM model achieved rapid convergence. After 63 iterations, its fitness value stabilized at approximately 0.0047. In contrast, the SSA-ELM model exhibited a slower convergence rate, requiring 70 iterations to reach a fitness value of approximately 0.009. These findings demonstrate that the ISSA-ELM model is closer to the global optimal solution

TABLE 3 Sample dataset.

Sample	Degree of karst development/%	Distance to faults/m	Overburden thickness/m	Groundwater fluctuation/m·a ⁻¹	Distance to pumping center/m	Collapse density (per 10 km ²)	Label
1	12.3	32	18.3	7.64	12	7	3
2	9.8	55	15	8.6	35	20	3
3	15.6	33	18	8.87	56	19	3
4	8.5	42	8	13.7	78	12	3
5	11.2	28	11	17	90	11	3
6	13.7	47	10	2.8	110	8	3
7	10.1	120	65	10.2	520	2	1
8	7.9	150	37	11.76	650	14	1
9	14.4	48	20	15	132	9	3
10	12.9	33	15	11	155	13	3
11	16.2	200	28.3	12.1	710	11	1
12	9.2	31	12.5	5.37	178	13	3
13	11.8	45	6.7	8.3	19	16	3
14	10.7	37	14	11.13	45	5	3
15	13.1	350	37	15.4	830	12	1
16	8.8	39	14	11.67	67	12	3
17	15.3	35	17	10.06	89	18	3
18	12.6	48	14	3.12	123	14	3
19	9.5	44	18	5.7	145	7	3
20	14.7	37	19	2.57	107	15	3

TABLE 4 Covariance matrix of six variables.

Variable	Z_1^*	Z_2^*	Z_3^*	Z_4^*	Z_5^*	Z_6^*
Z_1^*	1.0526	-0.1031	0.1103	-0.1029	0.1080	0.0525
Z_2^*	-0.1031	1.0526	-0.0776	0.2224	0.7858	0.1873
Z_3^*	0.1103	-0.0776	1.0526	-0.2154	0.0873	0.0397
Z_4^*	-0.1029	0.2224	-0.2154	1.0526	0.3014	-0.0652
Z_5^*	0.1080	0.7858	0.0873	0.3014	1.0526	0.0141
Z_6^*	0.0525	0.1873	0.0397	-0.0652	0.0141	1.0526

TABLE 5 Eigenvalues of the covariance matrix for each variable.

Principal component variable	X_1	X_2	X_3	X_4	X_5	X_6
Eigenvalue	2.00	1.34	1.06	0.97	0.76	0.20
Variance contribution rate, %	31.74	21.16	16.73	15.28	11.96	3.13
Cumulative variance contribution rate, %	31.74	52.90	69.63	84.91	96.87	100.00

TABLE 6 Eigenvectors corresponding to the eigenvalues of the covariance matrix of six variables.

Variable	X_1	X_2	X_3	X_4	X_5
Z_1^*	-0.0407	-0.5044	0.3891	-0.0242	0.1712
Z_2^*	0.6521	-0.0975	-0.1321	-0.2971	0.6665
Z_3^*	-0.0783	-0.6139	0.2367	0.4789	0.1441
Z_4^*	0.3718	0.4390	0.1785	0.7582	0.1148
Z_5^*	0.6459	-0.2019	0.2218	-0.1295	-0.6869
Z_6^*	0.1069	-0.3543	-0.8330	0.2999	-0.1431

and outperforms the SSA-ELM model in terms of both optimization capabilities and search speed.

4.3 Prediction of karst-collapse susceptibility

The optimal parameter combination was identified using the ISSA in conjunction with a five-fold cross-validation technique. These optimized parameters were then applied to the ELM model, resulting in the ISSA-ELM model. To comprehensively evaluate the effectiveness of PCA for data preprocessing and to assess the predictive capability of the ISSA-ELM model in karst-collapse susceptibility prediction, validation experiments were conducted using two independent datasets. One dataset consisted of 20 sets of raw field-measured data from the exploration area, whereas the

other contained PCA-processed data to reduce dimensionality and mitigate redundancy.

The training datasets were used to develop the ISSA-ELM model, which was subsequently applied to the test datasets to generate predictions. The predicted values were then compared to the actual values to assess the model's predictive accuracy (refer to Figures 12–14). A comparative analysis was also performed to evaluate the relative performance of the ISSA-ELM model, the SSA-ELM model, and the conventional ELM model.

The fitting results demonstrated that the predicted and actual values for all three models exhibited a generally linear relationship, indicating a reasonable degree of accuracy. However, the coefficient of determination ($R^2 = 0.9914$) for the PCA-processed ISSA-ELM model was significantly higher than that for both the non-PCA-processed ISSA-ELM model ($R^2 = 0.8945$) and the SSA-ELM model ($R^2 = 0.9837$). This finding suggests that applying PCA preprocessing effectively enhances the predictive accuracy by reducing noise and eliminating redundant information.

Furthermore, the ISSA-ELM model's predicted values displayed a superior alignment with actual values, reflecting the model's ability to capture complex nonlinear relationships among the influencing factors of karst collapse. This improvement in predictive capability highlights the robustness and reliability of the ISSA-ELM model as an effective tool for karst-collapse susceptibility assessment. Its strong performance offers valuable insights for practical applications in geological hazard management and regional risk assessment.

4.4 Analysis of error metrics

To further evaluate the performance of the ISSA-ELM model, three quantitative metrics were applied: R^2 , MSE, and MAE, as

TABLE 7 Sample principal components of 20 sets of original data.

Sample	X_1	X_2	X_3	X_4	X_5	Label
1	0.1951	0.8412	0.2243	0.8250	0.4164	3
2	-1.1088	-0.8936	-0.6128	-1.1902	-0.1288	3
3	-1.1156	1.1636	-1.8450	-0.4318	0.2906	3
4	1.0692	-1.4289	-0.3521	0.2497	-0.4030	3
5	0.1672	-1.0043	-0.9085	1.4766	0.2303	3
6	0.8426	1.6515	1.0372	-0.4781	0.3287	3
7	-0.3019	-1.1554	-0.5059	-1.5938	-0.8299	1
8	0.3988	-2.0851	1.5171	-0.6149	-0.3268	1
9	-0.1935	0.3527	-0.7559	1.8877	0.5824	3
10	0.3567	0.2279	-0.5183	0.2046	-0.1037	3
11	-0.5726	0.2954	1.2158	0.6263	1.3463	1
12	-0.8876	-0.1501	1.1124	-0.2751	-0.8738	3
13	-0.6103	-0.4283	0.0356	-0.9649	0.9998	3
14	0.7837	0.2155	0.0255	1.6396	-0.1804	3
15	-0.0526	-1.2964	1.6167	0.4785	0.8863	1
16	0.0057	-1.3587	0.4027	0.1493	0.1907	3
17	-0.9001	0.8008	-1.4816	-0.2934	0.4536	3
18	1.9534	1.9457	-1.1564	-1.2251	-1.8506	3
19	0.6418	0.7898	0.5852	0.7657	-1.5597	3
20	-0.6714	1.5165	0.3638	-1.2357	0.5313	3

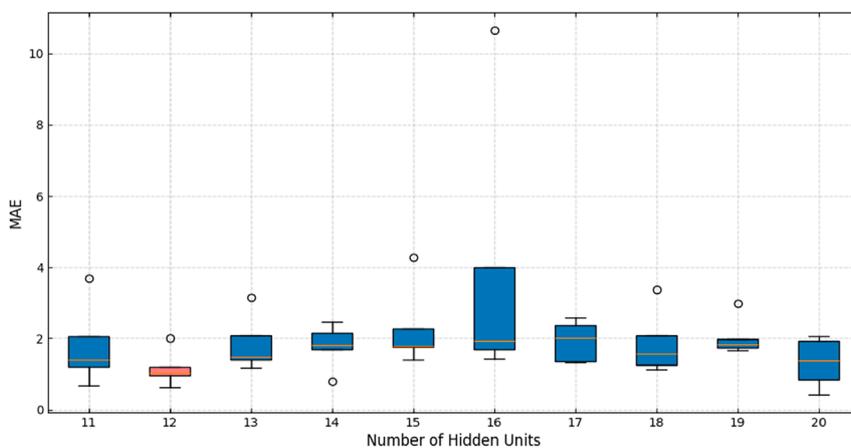


FIGURE 10 Error graph of five-fold cross-validation.

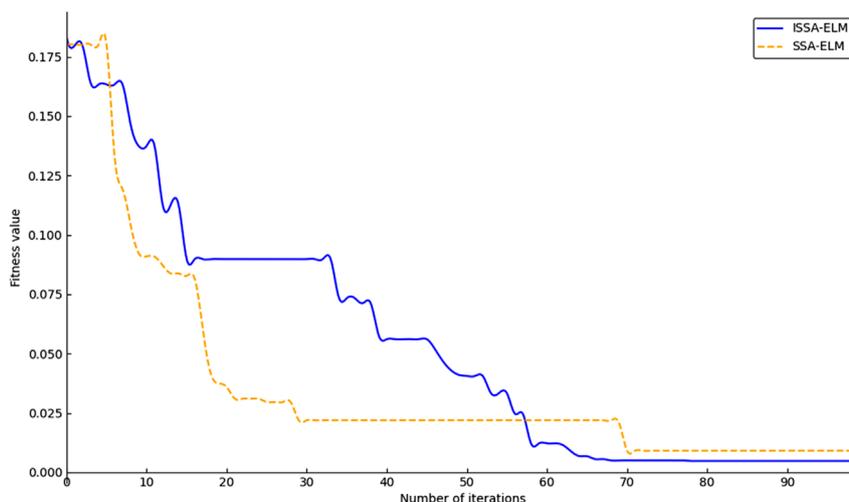


FIGURE 11 Adaptation graph.

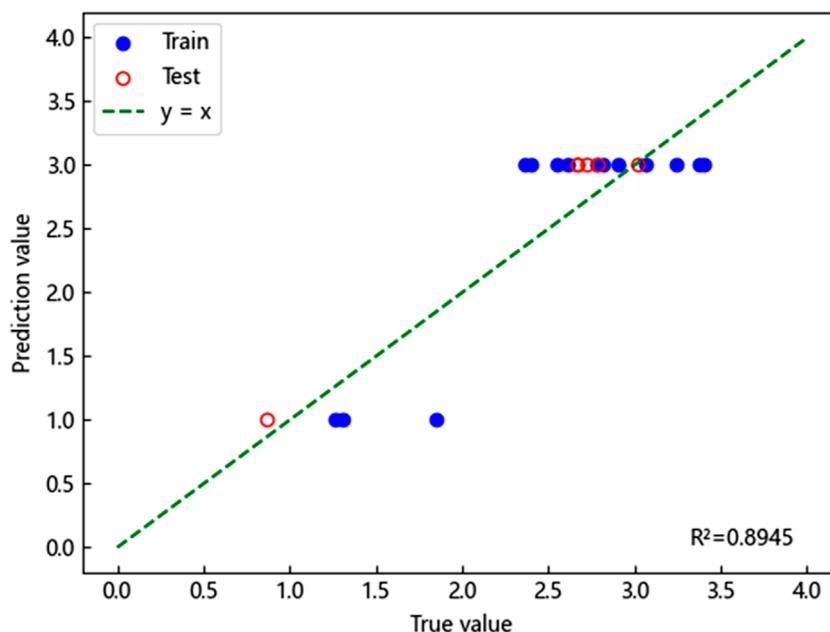


FIGURE 12 Prediction results of the ISSA-ELM model (without PCA processing).

described in Section 2.7. A comprehensive comparative analysis was conducted to assess the predictive accuracy and generalization capability of the ISSA-ELM model against the SSA-ELM model and the conventional ELM model (Figure 15)

The results presented in Figure 15 indicate that the ISSA-ELM model achieved the lowest MSE (0.0047) and MAE (0.0544) among all models. In comparison, the SSA-ELM model reported an MSE of 0.0091 and an MAE of 0.0766, whereas the ELM model exhibited a significantly higher MSE of 0.9988 and an MAE of 0.6903. These results clearly demonstrate the superior predictive accuracy and

robustness of the ISSA-ELM model. The enhanced convergence of the ISSA, facilitated by the integration of the Singer chaotic map, effectively mitigated the limitations of traditional sparrow search algorithms, preventing premature convergence and ensuring optimal parameter selection.

Furthermore, the goodness-of-fit coefficient (R^2) of the ISSA-ELM model was markedly higher than those of the SSA-ELM and ELM models. This result reflects the ISSA-ELM model's excellent generalization performance, indicating its ability to make accurate predictions across diverse datasets. The reduced error values and

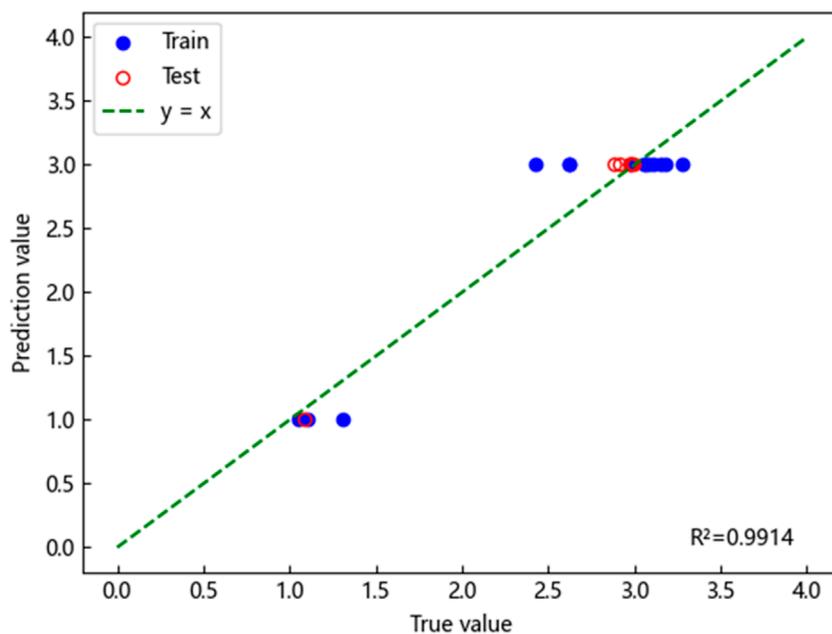


FIGURE 13 Prediction results of the ISSA-ELM model (with PCA processing).

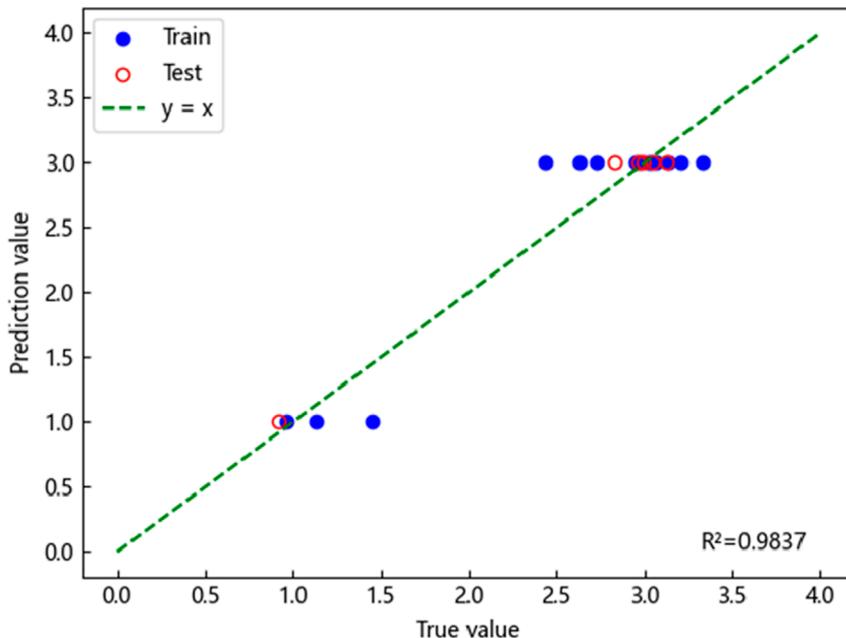


FIGURE 14 Prediction results of the SSA-ELM model.

consistent prediction accuracy further validate the model’s reliability and stability.

In summary, the ISSA-ELM model exhibits substantial advantages in karst-collapse susceptibility prediction. Its remarkable accuracy, minimized prediction errors, and robust generalization

capacity make it an effective and reliable tool for geohazard risk assessment. Consequently, the model can serve as a valuable decision-support system for policymakers, urban planners, and geotechnical engineers engaged in geological disaster prevention and management.

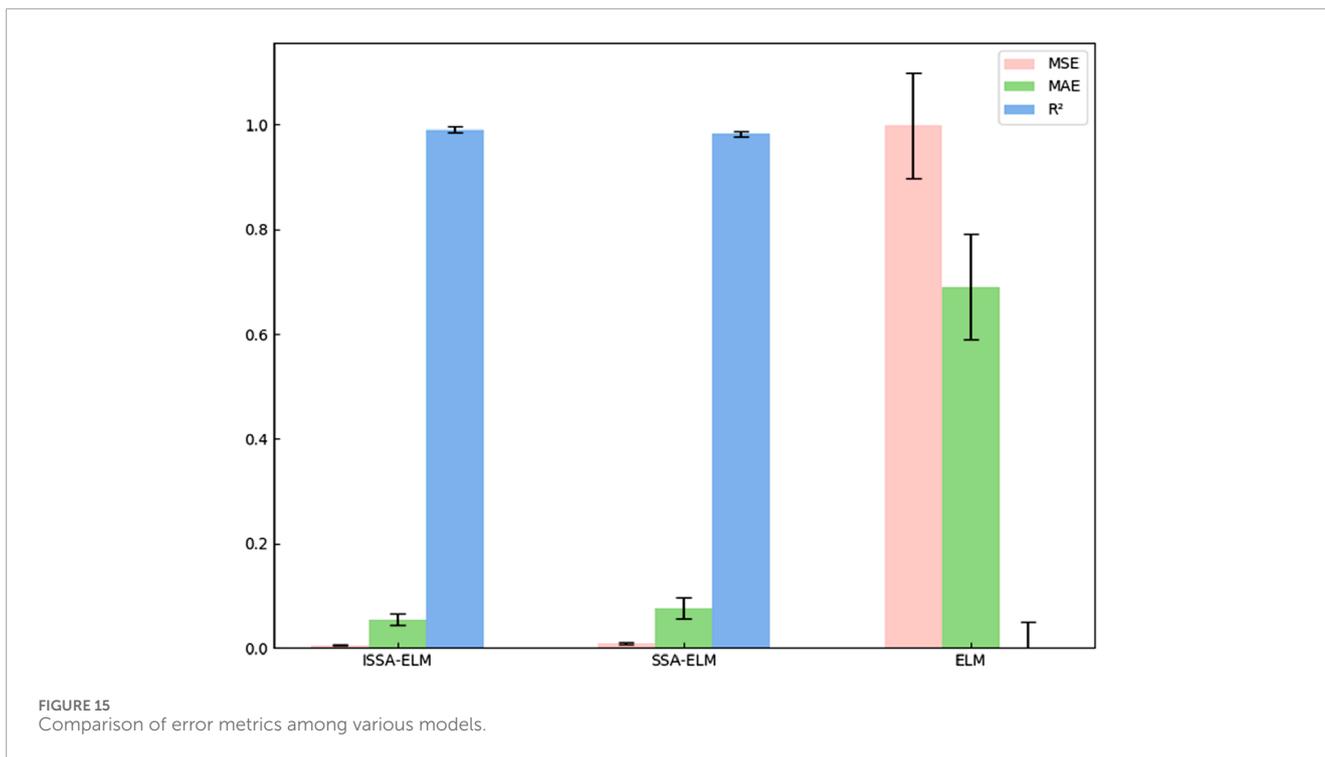


FIGURE 15 Comparison of error metrics among various models.

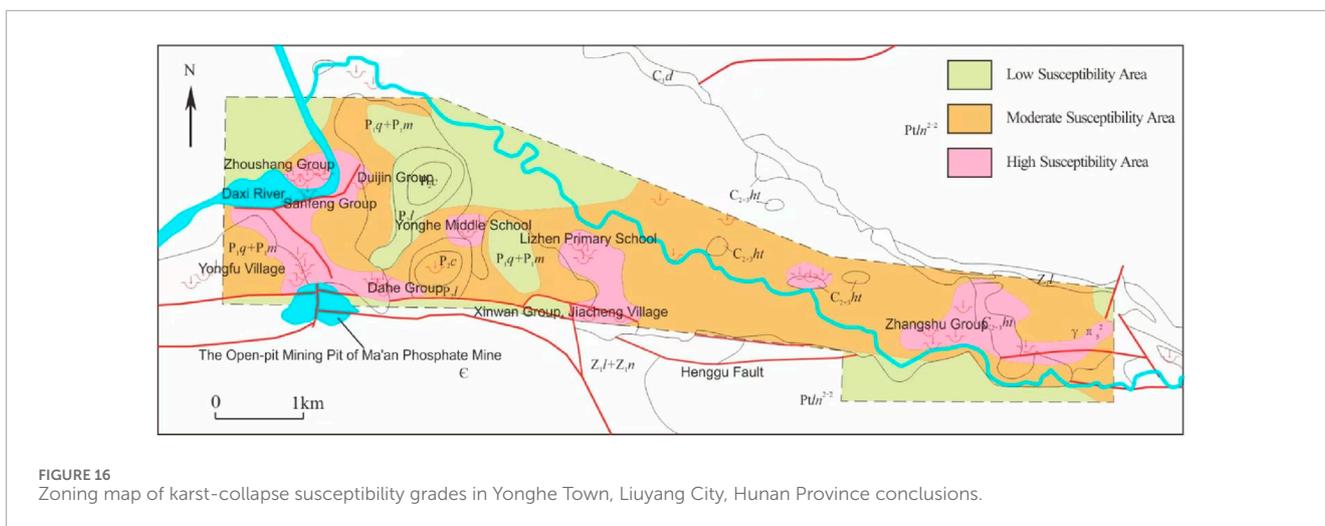


FIGURE 16 Zoning map of karst-collapse susceptibility grades in Yonghe Town, Liuyang City, Hunan Province conclusions.

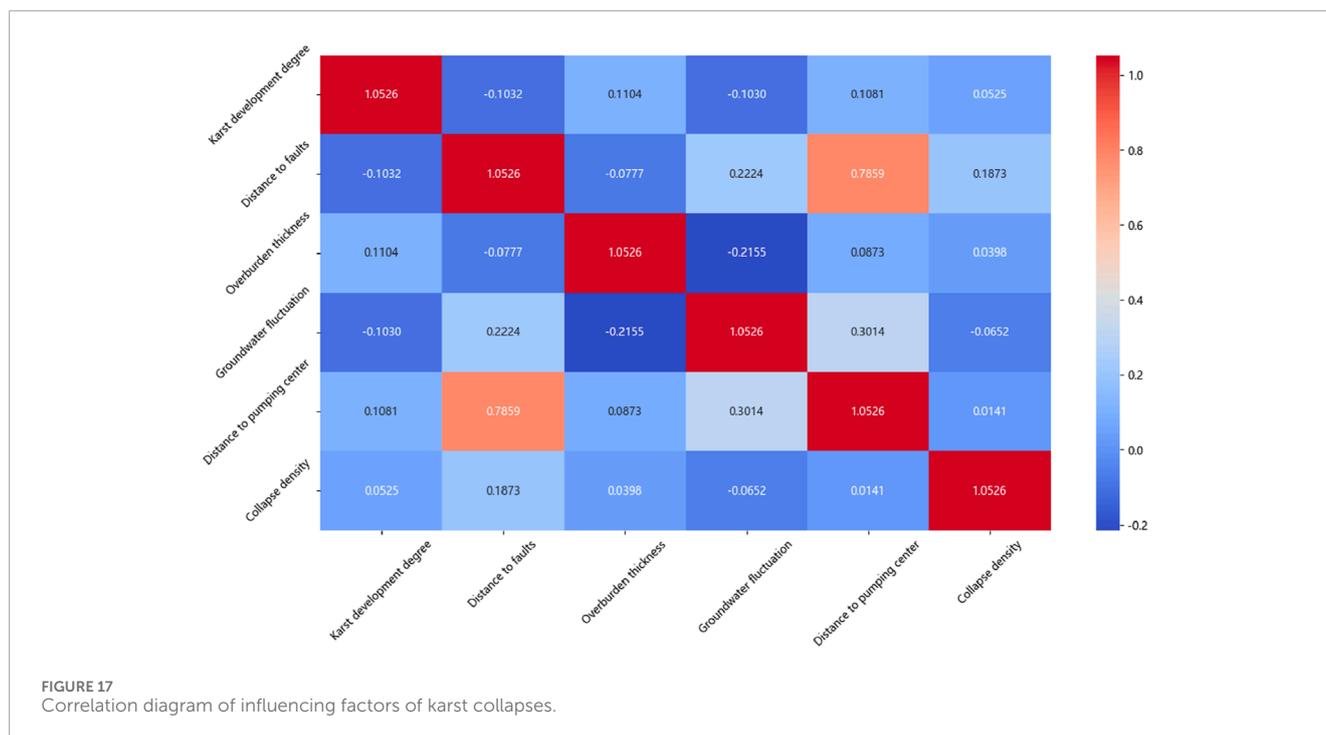
4.5 Model validation

To comprehensively evaluate the effectiveness and accuracy of the ISSA-ELM model in predicting the karst-collapse susceptibility, the evaluation area shown in Figure 7 was designated as the primary study region. Subsequently, the prediction scope was expanded to cover the entire Yonghe Town. A karst-collapse susceptibility zoning map was generated using MAPGIS software, as illustrated in Figure 16.

The results reveal that Yonghe Middle School, Lizhen Primary School—Xinwan Group of Jiacheng Village, Sanfeng Group—Dahe Group, and Zhangshu Group are classified as high-susceptibility

zones. Conversely, Duijin Group and Zhoushang Group are identified as low-susceptibility areas, with other regions falling under medium-susceptibility zones.

Notably, numerous collapse pits were identified in Jingquan Village, Tieshan Village, and Qibaoshan Village, all located in proximity to the Qibaoshan mining area. Although most pits have been backfilled, a substantial secondary collapse occurred in the Zhangshu Group of Tieshan Village. This incident further substantiates the predictive reliability of the ISSA-ELM model, which accurately identified high-susceptibility regions. The model's capacity to predict secondary collapses post-backfilling highlights its practical applicability in real-world scenarios.



The correlation analysis of influencing factors, presented in Section 4.1 (Figure 17), further corroborates these findings. Proximity to intra-area faults and the distance from the pumping funnel center were identified as primary determinants of karst-collapse susceptibility. These insights provide a strong basis for the implementation of targeted disaster prevention and mitigation measures.

To mitigate the risk of karst collapses effectively, comprehensive monitoring and preventive measures are essential in high-susceptibility regions. Advanced remote sensing, geophysical exploration techniques, and real-time monitoring systems can provide early warnings and facilitate rapid response efforts. Additionally, evidence-based management strategies, including regulated groundwater extraction and optimized land-use planning, should be prioritized to minimize collapse incidents.

Furthermore, this research provides valuable theoretical support for urban planning, particularly in the determination of disaster-avoidance distances. Policymakers and urban planners can utilize the susceptibility zoning map to make informed decisions, ensuring that critical infrastructure and densely populated areas are located outside high-susceptibility zones.

5 Conclusion

- (1) In this study, we used PCA for dimensionality reduction in karst-collapse susceptibility factors. Comparative analysis confirmed that the application of PCA significantly enhanced the predictive accuracy of the ISSA-ELM model, resulting in a

10.83% increase in R^2 . This demonstrates the effectiveness of PCA in reducing data redundancy and extracting critical features.

- (2) To overcome the limitations of the traditional SSA, including local optima convergence and slow convergence speed, the Singer chaotic map was introduced for population initialization. This enhancement resulted in the development of the ISSA, which was integrated with the ELM model. The ISSA-ELM model exhibited superior predictive performance across 20 measured datasets, achieving an MAE of 0.0544, an MSE of 0.0047, and an R^2 of 0.9914. These results highlight the model's exceptional accuracy and robust generalization capability compared to conventional ELM and SSA-ELM models.
- (3) The optimal number of hidden-layer nodes in the ELM model was determined through a rigorous five-fold cross-validation approach. This method effectively mitigated the risks of underfitting and overfitting while maximizing the use of limited data. The selected optimal number of nodes ($N = 12$) further demonstrated the effectiveness of this approach in fine-tuning model hyperparameters.
- (4) The ISSA-ELM model was applied for the first time to predict karst-collapse susceptibility. The generated susceptibility zoning map showed a significant spatial correlation between high-susceptibility zones and densely populated areas, including schools and villages. Through correlation analysis, the primary influencing factors were identified, offering a solid theoretical foundation for determining disaster-avoidance distances in urban planning. The findings provide essential guidance for policymakers in

developing evidence-based disaster prevention and mitigation strategies.

- (5) Despite its promising results, this study has certain limitations:
- ① Limited dataset: the dataset used for model training and validation was relatively small, which may affect the model's generalizability.
 - ② Activation function selection: only the sigmoid activation function was applied, with no comparative analysis involving other functions. Future studies could explore Tanh or ReLU activation functions to assess their impact on model performance.
 - ③ Data imbalance: the dataset contained limited non-disaster-point data, potentially resulting in an incomplete representation of non-disaster areas. Techniques such as the synthetic minority over-sampling technique (SMOTE) could be applied to address this imbalance.
 - ④ Normalization methods: the study did not investigate the effects of different normalization methods, leaving the potential impact of alternative approaches unexplored.
- (6) To further refine and validate the ISSA-ELM model, future research should focus on the following aspects:
- ① Dataset expansion: incorporating additional data from various karst regions will improve model training and validation.
 - ② Activation function evaluation: comparative studies of different activation functions, including Tanh and ReLU, can optimize model performance.
 - ③ Algorithmic enhancements: developing hybrid optimization strategies could further mitigate the risk of local optima in the sparrow algorithm.
 - ④ Data balancing: applying algorithms like SMOTE will balance the sample distribution and enhance model robustness.
 - ⑤ Normalization analysis: evaluating the impact of different normalization methods will provide insights into their influence on prediction accuracy.
 - ⑥ Generalizability assessment: applying the ISSA-ELM model to other karst regions with diverse geological and hydrological conditions will assess its broader applicability.

By addressing these aspects, future research can further enhance the predictive accuracy, stability, and practical applicability of the ISSA-ELM model. Ultimately, this will contribute to more effective karst-collapse prevention and improved risk management strategies.

Data availability statement

The original contributions presented in the study are included in the article/[supplementary material](#); further inquiries can be directed to the corresponding author.

Author contributions

JW: conceptualization, data curation, and writing – original draft. YY: methodology, supervision, and writing – review and editing. XY: data curation, supervision, validation, and writing – original draft. YuL: data curation, investigation, and writing – review and editing. YaL: data curation, investigation, and writing – review and editing. DH: software, validation, and writing – review and editing. YH: software and writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

The authors are grateful to the Hunan Institute of Territorial Spatial Survey and Monitoring, providing access to field data and technical support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feart.2025.1581090/full#supplementary-material>.

References

- Bao, H. M., and Hu, C. S. (2002). Neural network prediction of karst ground collapse. *J. Eng. Geol.* 10 (3), 299–304. doi:10.3969/j.issn.1004-9665.2002.03.014
- Bouzerda, M., Mehdi, K., Boualla, O., Fadili, A., Najib, S., and Saied, M. (2024). Inventory and geomorphological analysis of karstic collapse dolines in Sahel-Doukkala (Morocco). *Carbonates Evaporites* 39 (4), 93. doi:10.1007/s13146-024-01007-7
- Cao, B., Bai, G., and Li, H. (2015). Prediction and analysis of gas content based on PCA-GA-BP neural network. *J. Saf. Sci. Technol.* (5), 84–90. doi:10.11731/j.issn.1673-193x.2015.05.013
- Chen, D., and Cao, Y. (2023). Numerical simulation of karst collapse in qingling Town, wuhan. *J. Phys. Conf. Ser.* 2565 (1), 012034. doi:10.1088/1742-6596/2565/1/012034
- Chen, X. J., Chen, L. J., Song, Y., and Pengyan, B. (2019). Prediction and analysis of karst collapse with entropy - normal cloud model. *J. Eng. Geol.* 27 (6), 1389–1394. doi:10.13544/j.cnki.jeg.2018-347
- Chen, H. F., Zhu, M. Q., Xia, R. Y., Tang, J. S., and Liang, B. (2005). Analysis on epikarst spring with BP ANN in Luota, Hunan Province. *Carsologica Sin.* (4), 300–304. doi:10.3969/j.issn.1001-4810.2005.04.008
- Ding, H., Wu, Q., Zhao, D., Mu, W., and Yu, S. (2019). Risk assessment of karst collapse using an integrated fuzzy analytic hierarchy process and grey relational analysis model. *Geomechanics Eng.* 18 (5), 515–525. doi:10.12989/gae.2019.18.5.515
- En, W., Long-Cang, S., Li-Hong, L., and Bi-Juan, H. (2011). Prediction model for water level of sinkholes in karst region based on improved SVM. *J. Hohai Univ. Nat. Sci.* 39 (1), 20–23. doi:10.1007/s12182-011-0118-0
- Feng, T. (2025). *Mechanism and stability analysis of karst collapse in jingquan water source area of tengzhou city*. Singapore: Springer. doi:10.1007/978-981-97-7251-3_20
- Gao, C., Li, S., Wang, J., Li, L., and Lin, P. (2018). The risk assessment of tunnels based on grey correlation and entropy weight method. *Geotech. Geol. Eng.* 36, 1621–1631. doi:10.1007/s10706-017-0415-5
- Gao, Z. J., Ma, H. H., Wang, M., and Cheng, S. C. (2009). Preliminary exploration of the prediction model for karst ground collapse. *Chin. J. Geol. Hazard Control.* 20 (4), 66–71. doi:10.3969/j.issn.1003-8035.2009.04.014
- Hu, R., Yeung, M., Lee, C. S., W., and J., X. (2001). Regional risk assessment of karst collapse in Tangshan, China. *Environ. Geol.* 40 (11), 1377–1389. doi:10.1007/s002540100319
- Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing* 70 (1/3), 489–501. doi:10.1016/j.neucom.2005.12.126
- Huang, R. D., Han, M., Zhang, X. J., Zhang, H., Jin, H., and Hua, Z. (2011). Classification and prediction of karst collapse tendency grade based on Fisher discriminant method. *China Saf. Sci. J.* 21 (9), 70–76. doi:10.3969/j.issn.1003-3033.2011.09.012
- Jiang, C. L., and Jiang, Z. Q. (2012). Prediction of karst collapse based on Fisher discriminant analysis method. *J. Earth Sci. Environ.* 34 (1), 5. doi:10.3969/j.issn.1672-6561.2012.01.012
- Jiang, S. Q., and Shi, B. (2023). Prediction of failure pressure of pipelines with corrosion defects based on the ISSA-ELM model. *Hot Work. Technol.* 52 (12), 70–75 + 80. doi:10.14158/j.cnki.1001-3814.20221045
- Jiyuan, H., Mahdi, M., Jiayao, W., Qin, F., Zhang, J., Wu, W., et al. (2021). Karst collapse risk zonation and evaluation in wuhan, China based on analytic hierarchy process, logistic regression, and InSAR angular distortion approaches. *Remote Sens.* 13 (24), 5063. doi:10.3390/rs13245063
- Journal of Chemistry (2023). Retracted: study on early warning of karst collapse based on the BP neural network. 1. doi:10.1155/2023/9818621
- Lai, Y. B., and Qiao, C. S. (2008). An intelligent prediction model for karst collapse based on support vector machine. *J. Beijing Jiaot. Univ.* 32 (1), 5. doi:10.3969/j.issn.1673-0291.2008.01.009
- Liu, C. A., Feng, X. L., Sun, C. H., and Zhao, L. J. (2022). Maximum 2 - D entropy image segmentation method based on improved sparrow algorithm. *Laser Technol.* 46 (2), 274–282. doi:10.7510/jgjs.issn.1001-3806.2022.02.020
- Luo, X. J., and Shen, J. (2018). Research progress and prospect of karst ground collapse in China. *Carsologica Sin.* 37 (1): 101–111. doi:10.11932/karst20180106
- Lv, X., Mu, X. D., Zhang, J., and Wang, Z. (2021). Chaotic sparrow search optimization algorithm. *J. Beijing Univ. Aeronautics Astronautics* 47 (8), 1712–1720. doi:10.13700/j.bh.1001-5965.2020.0298
- Meng, Y., Huang, J. M., Lei, M. T., Li, Y., and Dai, J. L. (2009). Quantitative forecasting method of karst collapse based on grey Verhulst model. *Carsologica Sin.* (1), 17–22. doi:10.3969/j.issn.1001-4810.2009.01.004
- Ministry of Natural Resources of the People's Republic of China (2023). DZ/T 0447-2023 specification for karst collapse survey. 1. Beijing: China Standard Press. 10–31.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series* 6. 2 (11), 559–572. doi:10.1080/14786440109462720
- Qiu, X. R. (2004). Grey - fuzzy comprehensive evaluation on the stability of karst collapse. *Hydrogeology and Eng. Geol.* 31 (4), 4. doi:10.3969/j.issn.1000-3665.2004.04.010
- Ren, T., Tian, G. L., Ning, Z. J., Zhou, A. H., Li, K., and Chen, S. (2023). Evaluation of karst collapse susceptibility based on geographical detector and random forest. *J. Catastrophology* 38 (3), 227–234. doi:10.3969/j.issn.1000-811X.2023.03.035
- Wang, G., Hao, J., Wen, H., and Cao, C. (2022). A random forest model of karst ground collapse susceptibility based on factor and parameter coupling optimization. *Geocarto Int.* 37 (27), 15548–15567. doi:10.1080/10106049.2022.2102216
- Wang, Z. Z., Zhuang, Z. H., Hu, F. Y., and Huang, W. L. (2024). Formation conditions and susceptibility assessment of karst collapses in the northern hilly area of Guangzhou City. *J. China Geol. Hazard Control* 35 (4), 163–172. doi:10.16031/j.cnki.issn.1003-8035.202311008
- Wei, A., Li, D., Zhou, Y., Deng, Q., and Yan, L. (2021). A novel combination approach for karst collapse susceptibility assessment using the analytic hierarchy process, catastrophe, and entropy model. *Nat. Hazards* 105, 405–430. doi:10.1007/s11069-020-04317-w
- Xiaozhen, J., Jianling, D., Zhiwen, Z., Li, X. J., Ma, X., Zhou, W., et al. (2024). An overview on karst collapse mechanism in China. *Carbonates and Evaporites* 39 (3), 71. doi:10.1007/s13146-024-00986-x
- Xie, J. F., Tan, F., Jiao, Y. Y., Zou, J., and Mao, Z. (2021). Prediction of karst ground collapse based on factor analysis-GA-ELM model. *J. Eng. Geol.* 29 (2), 236–244. doi:10.13544/j.cnki.jeg.2020-219
- Xue, J. K., and Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. *Systems Science and Control Engineering* 8 (1), 22–34. doi:10.1080/21642583.2019.1708830
- Yang, Z., Li, B., Wu, H., Li, M., Fan, J., Chen, M., et al. (2023). Water consumption prediction and influencing factor analysis based on PCA-BP neural network in karst regions: a case study of Guizhou Province. *Environmental Science and Pollution Research* 30 (12), 33504–33515. doi:10.1007/s11356-022-24604-2
- YanHua, X., BingHui, Z., YuXin, L., Liu, B. C., Zhang, C. F., and Lin, Y. S. (2022). Evaluation of the karst collapse susceptibility of subgrade based on the AHP method of ArcGIS and prevention measures: a case study of the quannan expressway, section K1379+300-K1471+920. *Water* 14 (9), 1432. doi:10.3390/w14091432
- Zhang, J., Bi, P., Wei, A. H., Tao, Z. B., and Zhu, H. C. (2021). Assessment of susceptibility to karst collapse in the Qixia Zhongqiao district of Yantai based on fuzzy comprehensive method. *Carsologica Sinica* (2), 215–220. doi:10.11932/karst2021y07
- Zhou, A. H., Niu, P. F., Yuan, Y., and Huang, H. C. (2020). Prediction of karst surface collapse risk in Fankou lead - zinc mine area based on PCA-PSO-SVM. *Carsologica Sinica* (4), 622–628. doi:10.11932/karst2020y30
- Zhuang, S. Y. (2022). Assessment of karst collapse susceptibility in yongding district, Longyan city using the fuzzy analytic hierarchy process. *Geology of Fujian* 41 (4), 323–331. doi:10.3969/j.issn.1001-3970.2022.04.008