A Graph–Transformer Method for Landslide Susceptibility Mapping

Qing Zhang, Yi He, *Member, IEEE*, Yalei Zhang, Jiangang Lu, Lifeng Zhang, Tianbao Huo, Jiapeng Tang, Yumin Fang, and Yunhao Zhang

***Abstract*—Landslide susceptibility mapping (LSM) is of great significance for regional land resource planning and disaster prevention and reduction. The machine learning (ML) method has been widely used in the field of LSM. However, the existing LSM model fails to consider the correlation between landslide and disaster-prone environment (DPE) and lacks global information, resulting in a high false alarm rate of LSM. Therefore, we propose an LSM method with graph–transformer that considers the DPE characteristics and global information. First, correlation analysis and importance analysis are employed on nine landslide contributing factors, and the landslide dataset is generated by combining remote sensing image interpretation and field verification. Second, a graph constrained by environment similarity relationship is constructed to realize the correlation between landslide and DPE. Then, the transformer module is introduced to construct a graph–transformer model that considers the global information. Finally, the LSM is generated and analyzed, and the accuracy of the proposed model is compared and evaluated. The experimental results show that the environment similarity relationship graph effectively improves the accuracy of the models and weakens the influence of environmental differences on the models. Compared with graph convolutional network, graph sample and aggregate, and graph attention network models, the area under the curve (AUC) value of the proposed model is more than 2.05% higher under the environment similarity relationship. In addition, the AUC value of the proposed model is more than 8.8% higher than that of traditional ML models. In conclusion, our proposed model framework can get better evaluation results than most existing methods, and its results can provide effective ways and key technical support for landslide disaster investigation and control.**

***Index Terms*—Environment similarity relationship, graph, landslide susceptibility mapping (LSM), transformer.**

Manuscript received 23 April 2024; revised 24 July 2024; accepted 30 July 2024. Date of publication 2 August 2024; date of current version 26 August 2024. This work is supported in part by the National Natural Science Foundation of China under Grant 42201459, in part by the Gausu Science and Technology Program under Grant 23JRRA881, in part by the Key and Development Project of Lanzhou Jiaotong University under Grant LZJTU-ZDYF2301, and in part by the Science and Technology Project of Gansu Province under Grant 22JR5RA067. *(Corresponding author: Yi He.)*

Qing Zhang, Yi He, Jiangang Lu, Lifeng Zhang, Tianbao Huo, Jiapeng Tang, Yumin Fang, and Yunhao Zhang are with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China, and with the National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China, and also with the Gansu Provincial Key Laboratory of Science and Technology in Surveying & Mapping, Lanzhou 730070, China (e-mail: 1439112766@qq.com; heyi@mail.lzjtu.cn; rs_lujg@qq.com; 119273207@qq.com; htb666@126.com; 3153292919@qq.com; fgymnhlcq@163.com; 596532969@qq.com).

Yalei Zhang is with the Zhangye Mineral Exploration Institute of Gansu Nonferrous Metals Geological Exploration Bureau, Zhangye 734000, China.

Digital Object Identifier 10.1109/JSTARS.2024.3437751

I. Introduction

LANDSLIDE disasters often appear in people's production and life, their strong destructive power will not only change the topography but also cause people's lives and property losses [1], [2]. Therefore, it is very important to evaluate the susceptibility of landslides scientifically and prevent landslide disasters in time. Landslide susceptibility mapping (LSM) refers to predicting the spatial distribution of the probability of landslide occurrence according to the characteristics of topography and geomorphology in the study area [3], [4], which is of great significance to the prevention of landslide disasters and the planning and siting of cities.

With the rapid development of artificial intelligence and Big Data, scholars [5], [6], [7] in LSM fields have begun to use machine learning (ML) methods to evaluate landslide susceptibility, such as multilayer perceptron (MLP) [8], convolutional neural network (CNN) [9], and gate recurrent unit (GRU) [10]. Compared with traditional landslide susceptibility assessment methods, ML has a more powerful feature analysis ability, which can further improve the reliability of LSM [11], [12], [13]. In recent years, as an important branch of ML, deep learning (DL) has inherited and developed the advantages of ML algorithms [14]. For landslide disaster, its formation conditions are relatively complex, and the amount of data involved is also relatively large. Theoretically, DL can better extract the formation characteristics of landslide disasters, efficiently and quickly process various landslide-related data, and more accurately assess the probability of landslides. For example, Zhao et al. [15] proposed a local and global feature extraction network, which combines the advantages of CNN and transformer models. Chen et al. [16] proposed a new hybrid framework combining landslide conditioning factors with swin transformer for landslide analysis. However, although these DL methods can obtain reliable LSM in simple small scenes, for complex environment scenes, there are still limitations that cannot consider the correlation between landslide and disaster-prone environment (DPE), and lack of feature extraction ability, resulting in a high false alarm rate of LSM.

The graph is generally represented as $G = (V, E)$, where V is the set of nodes in the graph, and E is the set of edges between nodes [17], [18]. Each edge represents the relationship between nodes [19]. Considering the complexity between landslide and DPE, compared with a single feature sequence or image, the graph structure has a high degree of data abstraction and

expression ability, and can more effectively express the complex nonlinear relationship between landslide and DPE. Graph neural network (GNN) is a kind of models that specialize in processing graphs and is widely used in social networks, recommendation systems, and knowledge graphs [20], [21], [22]. Nowadays, the GNN model has been applied in the field of LSM. For example, Wang et al. [23] proposed an LSM method based on graph convolutional networks (GCNs) and slope units. Zeng et al. [24] proposed a method based on graph sample and aggregate (GraphSAGE) and triangulated irregular network units to evaluate the susceptibility of landslides in Fengjie and Fulin counties, Chongqing. Yang et al. [25] combined graph attention network (GAT) and extreme gradient boosting models to propose a hybrid model for LSM. However, the graphs used in these models are based on spatial topological relationships. Although the graph based on spatial topological relationships can correlate the information of nodes around the sample, at a large scale, due to the heterogeneity of the geographical environment, the features of many neighboring nodes are irrelevant, resulting in the connection of many invalid edges. In addition, although these GNN models used above have achieved good results in LSM, they only consider the information of the neighbor nodes when the central node aggregates the neighborhood features, but do not consider the global information. As a result, the expression ability of the model is limited, and the accurate generalization of the whole graph information cannot be realized.

The prior knowledge of landslides shows similar geographical and geological environments will produce landslides with similar characteristics regardless of the distance [24], [26]. The characteristics of landslides can influence each other among similar subenvironment areas. In this interrelated and influencing environment, the clustering and correlation of similar landslide characteristics are formed, which can better help to summarize landslide characteristics. Therefore, the correlation between subenvironmental regions with similar geological conditions should be considered in LSM. Moreover, the transformer is a neural network model based on the self-attention mechanism that is capable of globally modeling each element in a sequence and making connections between the elements [15], [27], [28]. The self-attention mechanism allows the transformer to process each input sequence by focusing its attention on relevant parts at different locations in the sequence, effectively capturing global information [16].

In this article, we propose a graph–transformer model specifically designed for LSM. The graph–transformer can make use of the advantages of the graph in DPE correlation and transformer in extracting global information, which enables it to extract landslide characteristics more comprehensively and improve the reliability of LSM. First, landslide contributing factors (LCFs) and historical landslide database are constructed. Second, based on the environment similarity relationship, the directed graph of environment consistency constraint is constructed. Then, the transformer module is introduced to build the graph–transformer model. Finally, the performance and results of the model are evaluated and analyzed.

Two major research contributions come from our research. 1) We construct a directed graph that considers the correlation between landslide and DPE. 2) We construct a graph–transformer model that considers global information.

II. Materials

A. Experimental Scene

The Bailong River Basin (BRB) is located between 103°–105.5°E and 32.6°–34.4°N (see Fig. 1), which is one of the four regions with high incidence of geological disasters in China. The BRB is located in the northeastern part of the Qinghai-Tibet Plateau, which is the edge of the eastern collision deformation zone between the Indian plate and the Eurasian plate [29], [30]. Coupled with the influence of the north-south tectonic belt, the geological structure of the area is very complicated, and there are many high mountains and valleys. BRB is also located in the middle and north section of the North-South seismic belt in China, and many seismic belts pass through the area, with complex fault structures, active neotectonic movement, and frequent earthquakes [31]. In the basin, soft and hard rock layers alternate with each other, and weak interlayers develop. Under the action of internal and external forces such as structural failure and weathering, the rocks are extremely broken and their stability is extremely poor, which provides rich material basis and conditions for the development of landslides. The annual average rainfall in BRB is 587.2 mm, and the annual rainfall distribution is uneven. The total rainfall from May to September accounts for 75%–85% of the annual rainfall, mostly in the form of continuous rainfall or heavy rain, which provides the induced conditions for the development of geological disasters. In addition, due to the high population density and the shortage of residential land, human activities such as slope excavation, slope leveling, and landfilling are frequent, which are not conducive to slope stability. And because most of the traffic lines are built along the river valley, human engineering activities are particularly intense. In recent years, due to the combination of extreme weather (8·12 flash flood disaster in Longnan), earthquakes (5·12 Wenchuan earthquake, 8·8 Jiuzhaigou earthquake) [32], and human engineering activities, the frequency of landslides in this area has increased significantly, which has seriously threatened and restricted the safety of people's lives and property and the economic development of cities and towns.

B. Landslide Inventory

Using multiperiod high-resolution remote sensing images, we established a landslide database in the study area through manual visual interpretation and detailed field investigation, combined with unmanned aerial vehicle (UAV) images (see Fig. 2). We identified 625 landslides, and the largest area of which was 4.4 km^2. The landslide distribution is shown in Fig. 1.

C. LCFs of BRB

The development of landslides is affected by many environmental factors such as topography, geology, hydrology, and human activities [33], [34]. Based on the geographical condition

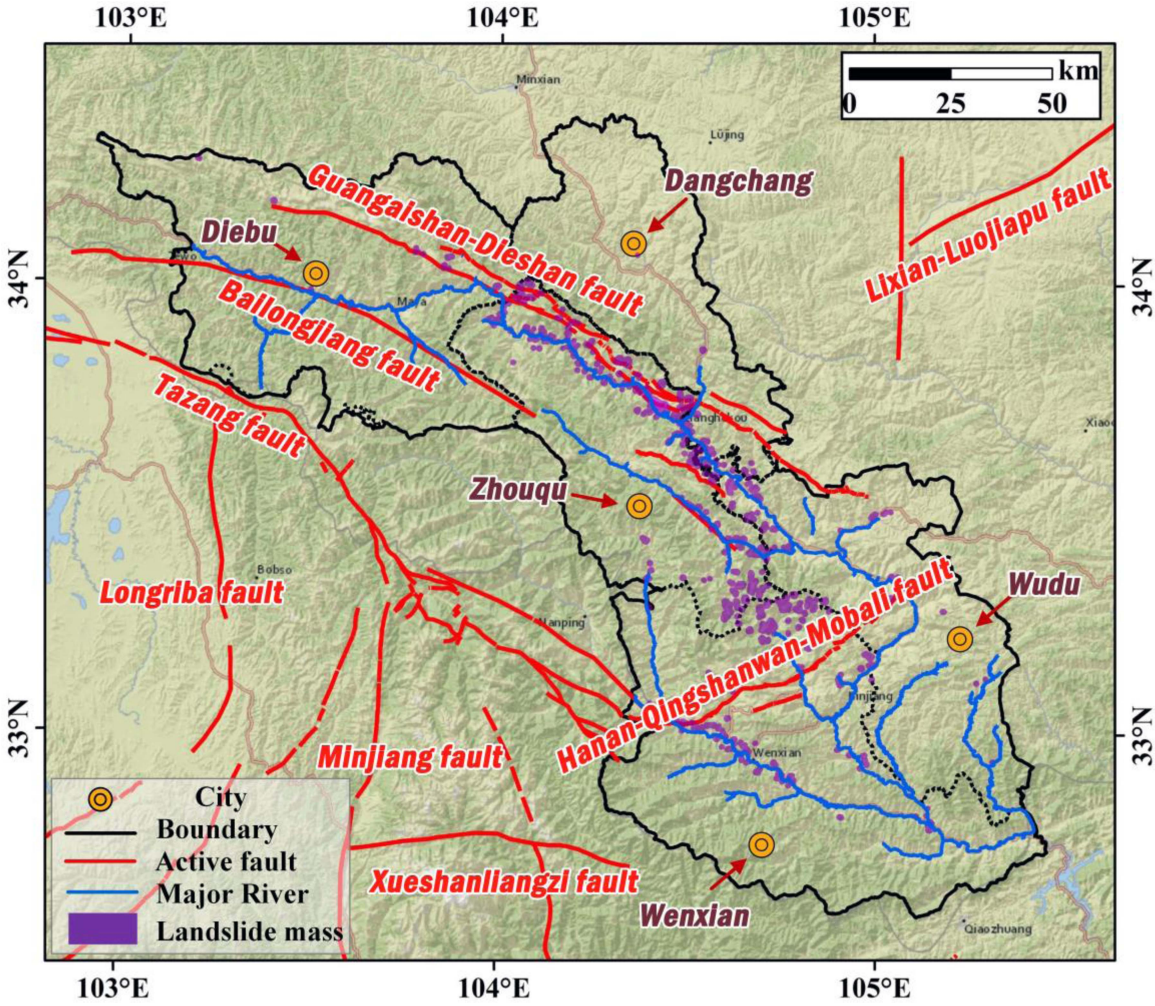


Fig. 1. Location of BRB.

of BRB and the collected database, we selected nine LCFs: altitude, slope, aspect, lithology, distance from faults, distance from rivers, distance from roads, normalized difference vegetation index (NDVI), and average annual rainfall (see Fig. 3).

The altitude, slope, and aspect are generated by using the digital elevation model (DEM) of BRB. The DEM is from the Geospatial Data Cloud Platform with a resolution of 30 m.[1] Lithology and fault data in the study area were extracted from 1:100 000 scale geological maps of the China Geological Survey.[2] The river and road network data were provided by OpenStreetMap,[3] and the Euclidean distance function in ArcGIS software was used to generate the river network distance distribution map and road distance distribution map of BRB. Using the Moderate-Resolution Imaging Spectroradiometer satellite data,[4] the NDVI map is generated on the environment for visualizing images remote sensing image processing platform. The rainfall data were obtained from the National Earth System Science Data Center,[5] and the rainfall factor was obtained by averaging the annual cumulative rainfall from 2019 to 2023.

The spatial resolution of the extracted data is different. On the premise of ensuring that the data resolution meets the accuracy of the result and is convenient for subsequent data processing, all the LCFs are uniformly resampling to a resolution of 30 m $\times$ 30 m.

III. Methods

First, we select nine geographical environment factors as LCFs and carry out correlation analysis and importance analysis of LCFs. Next, we combined remote sensing image interpretation and field verification to generate landslide datasets. Second, the graph dataset is constructed based on spatial topological relationships and environment similarity relationships. Then, a graph–transformer model considering the characteristics of the DPE and global information is constructed and compared with GCN, GraphSAGE, GAT, MLP, CNN, and GRU models. Finally, the generated LSM is analyzed statistically, and the accuracy of

[1][Online]. Available at: https://www.gscloud.cn/.
[2][Online]. Available at: https://www.cgs.gov.cn/.
[3][Online]. Available at: https://www.openstreetmap.org/.
[4][Online]. Available at: https://ladsweb.modaps.eosdis.nasa.gov/.
[5][Online]. Available at: https://www.geodata.cn/.

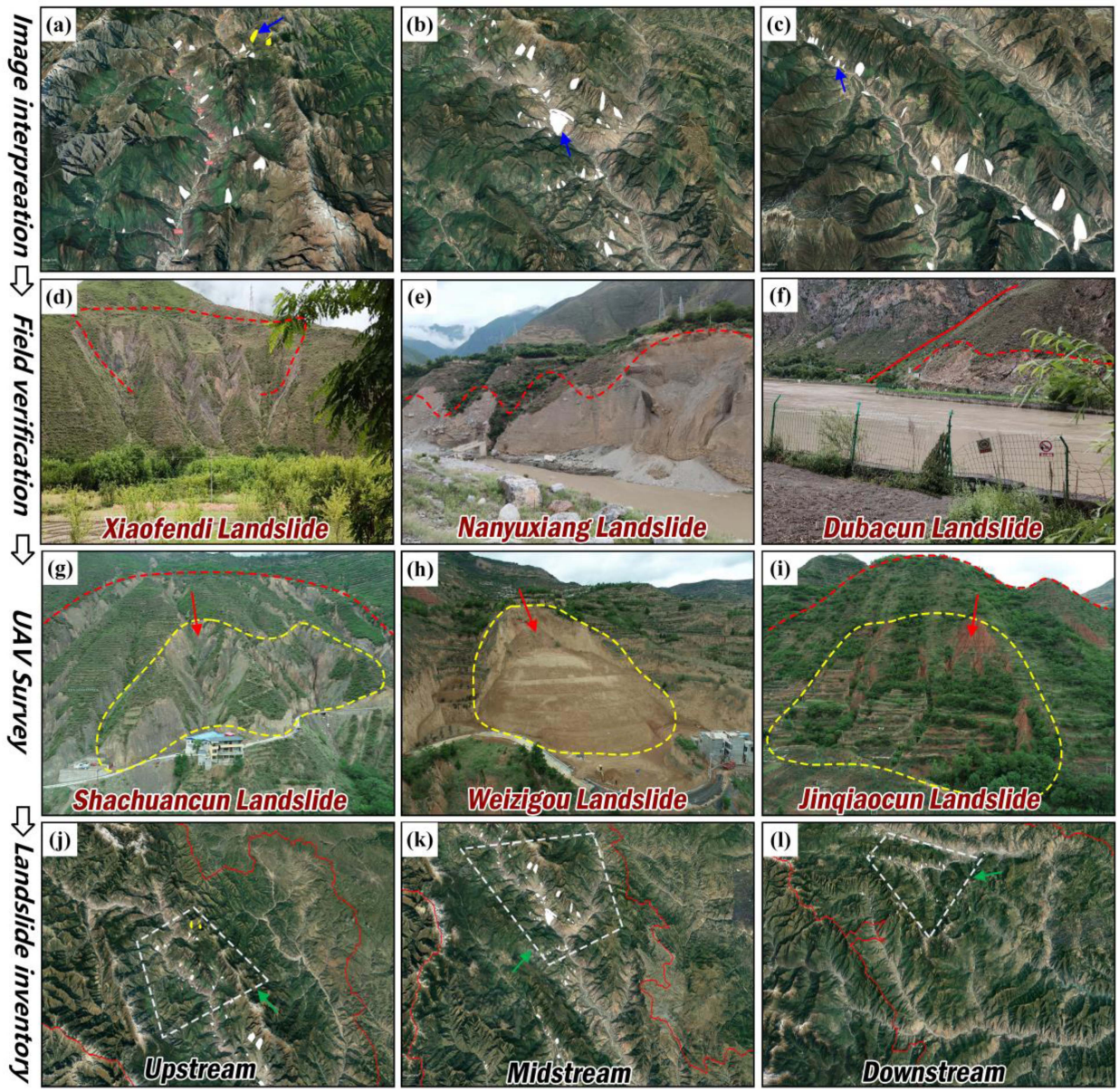


Fig. 2. Interpretation and inventory of landslide along the BRB: (a)–(c) Landslide identification via remote sensing images. (d)–(f) Field verification. (g)–(i) UAV images. (j)–(l) Compilation of landslide inventory.

the model is compared and evaluated. The detailed flow chart is shown in Fig. 4.

A. Evaluation of Contributing Factors

1) Colleration Analysis: To avoid the problem of model training error and low accuracy of LSM due to collinearity between LCFs, correlation analysis of each factor is needed to verify its independence. In this article, the Pearson correlation coefficient is used to measure the degree of correlation between LCFs. The formula is as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (1)$$

x_i and y_i are the ith observation of the two variables, $\overline{x}$ and $\overline{y}$ are the mean values of the two variables, and n is the number of samples.

2) Relief-F: To prevent the model from learning redundant feature information, it is necessary to further evaluate the importance of LCFs. The relief-F method is a feature selection algorithm, which can assign different weights to features according to the relevance of each feature and category [35]. Features with weights smaller than a certain threshold value will be removed. The formula is as follows:

$$W_i = W_i - \sum_{j=1}^{k} \frac{\mathrm{diff}(A_i, R, H_i)}{mk}$$

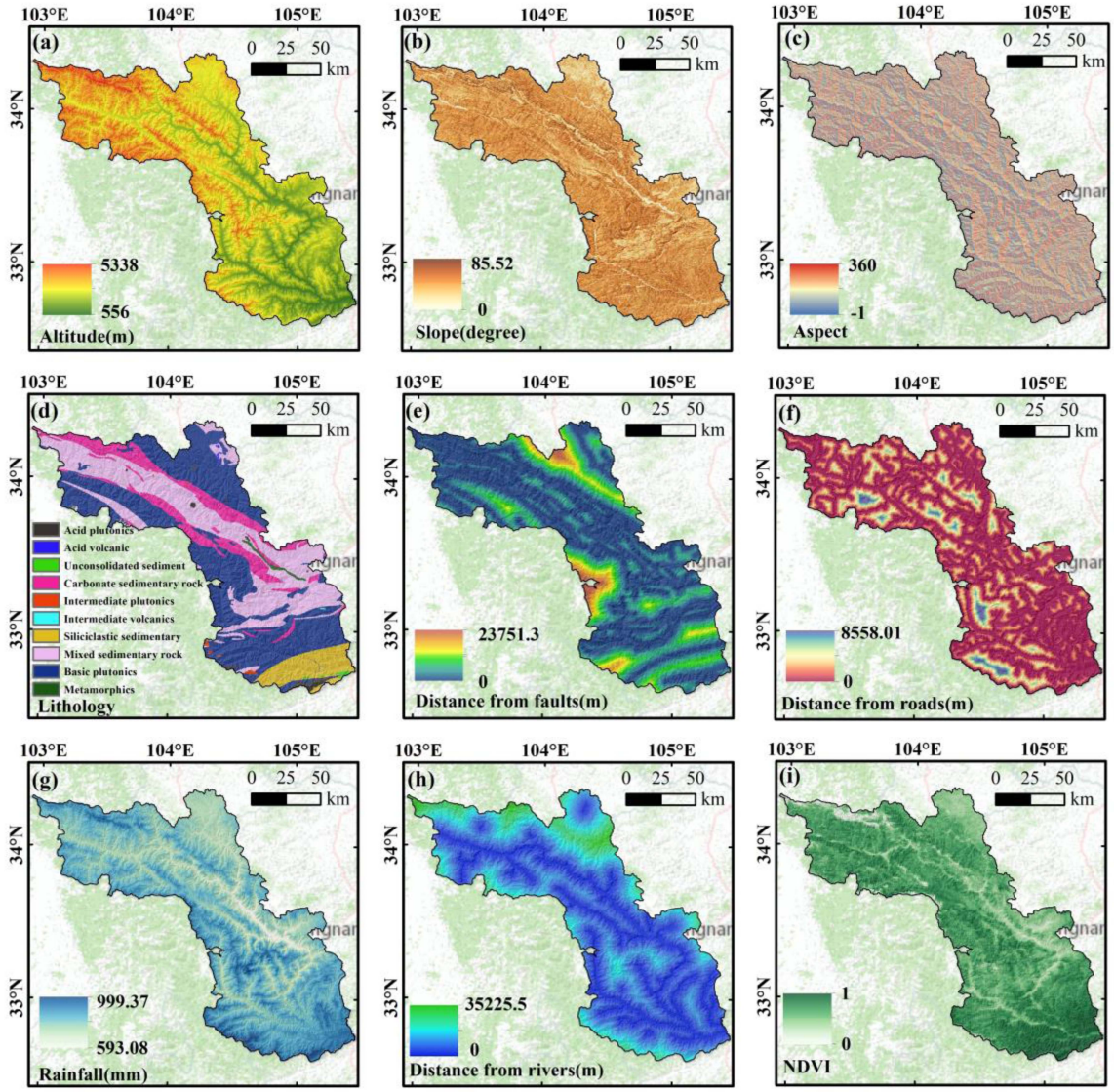


Fig. 3. Landslide contributing factors. (a) Altitude. (b) Slope. (c) Aspect. (d) Lithology. (e) Distance from faults. (f) Distance from roads. (g) Rainfall. (h) Distance from rivers. (i) NDVI.

$$+\sum_{C\neq \text{Class}(R)}\left\{\frac{p(C)}{1-p[\text{Class}(R)]}\sum_{j=1}^{k}\frac{\text{diff}[A_i,R,M_j(C)}{mk}\right\} \tag{2}$$

where C denotes the sample label, $p(C)$ is the probability of class C, Class(R) represents the label of sample R, M_j (C) is the jth sample of class C, and diff (A_i, R, H_i) and diff (A_i, R, $M_i(C)$) are the distance functions. Finally, we obtain the factor importance after repeating this process m times.

B. Base Models

GNN is a DL method based on graph structure, which is widely used in molecular chemistry, recommendation systems, and other fields. As an extension of neural networks, GNN can handle data formats represented by graph structures. In the graph, each node is defined by its characteristics as well as its neighbors and relations, and the network computes the representation vector of the target node by recursively aggregating and transforming the representation vector of the neighboring nodes. Nowadays, the mainstream GNNs include GCN, GraphSAGE, and GAT.

1) Graph Convolutional Network: GCN is a direct inference learning framework (see Fig. 5), its core idea is to use the whole adjacency matrix and convolution operation of the graph to fuse the information of the neighbor nodes, to realize the representation learning of the graph structure [36]. For each node, GCN aggregates the information of its neighbor nodes by weighted weights. The weight calculation usually considers the degree of the neighbor node to ensure the reasonable distribution of the contribution to the neighbor node

$$H^{(l+1)}=\sigma\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{3}$$

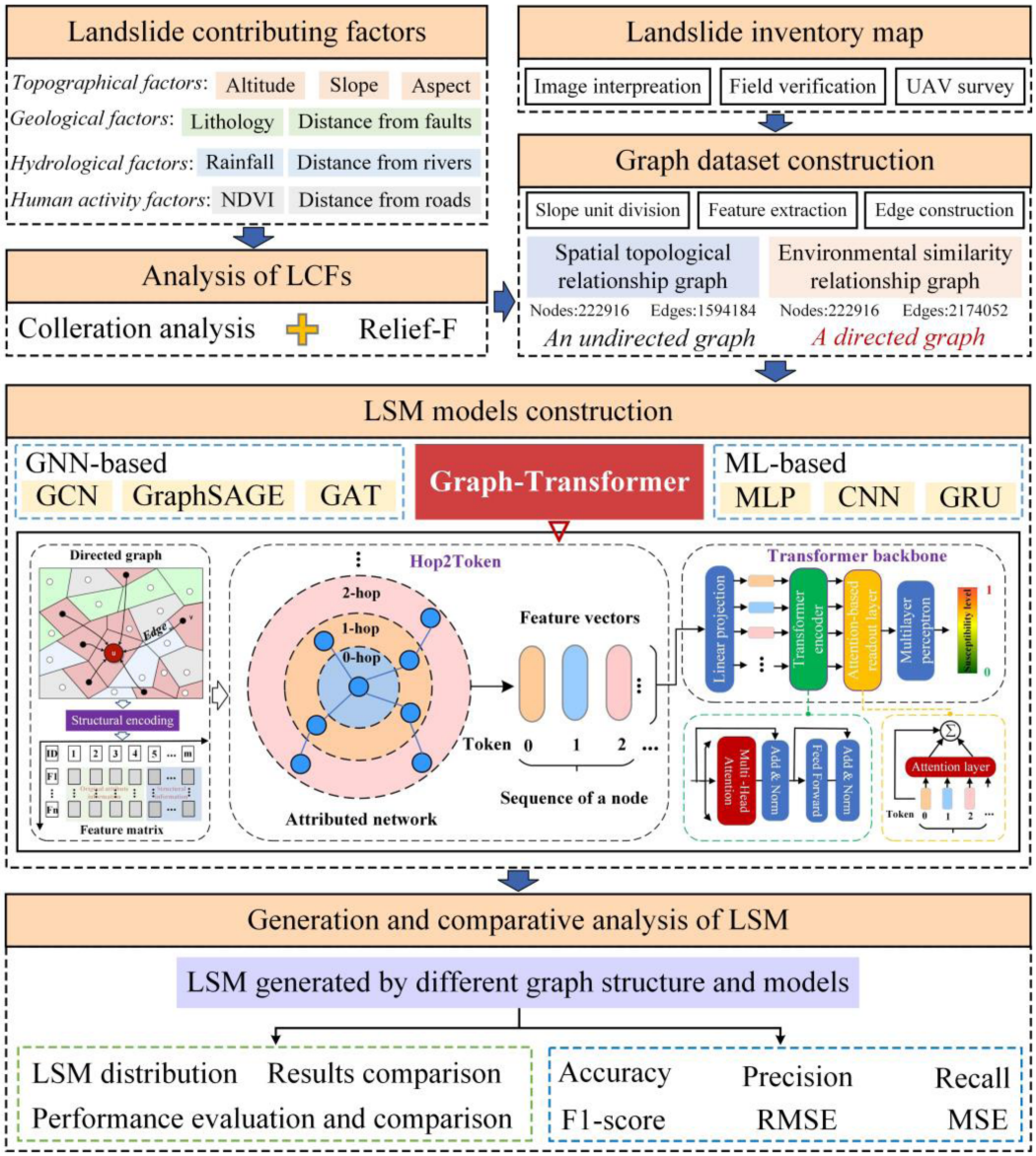


Fig. 4. Overall workflow of our approach.

where $H^{(l)}$ indicates the feature matrix in the lth layer, $H^{(0)} = X$; $\widetilde{D} = D + \mathrm{I}$; $\widetilde{A} = \mathrm{A}+\mathrm{I}$, and I is the identity matrix; $W^{(l)}$ is the weight matrix of the filter in the lth layer; σ denotes a nonlinear activation function.

2) GraphSAGE: GraphSAGE is a typical inductive learning framework that extracts the feature information of graph neighborhood nodes through neighbor sampling and information aggregation (see Fig. 6). The core idea is to generate the embedding vector of the target node by learning a function that represents the neighbor node in aggregate [37]. First, the node u is uniformly sampled with a fixed neighborhood size ($N(u)$). Then, the feature information of the neighbor node is aggregated by the aggregation function, that is, the representation of the neighbor node $N(u)$ of node u in layer $l+1$ (4). Finally, the *concat* operator is used to concatenate it with the current representation (h_u^l)of the node, multiply it with the weight matrix W, and then feed it to the fully connected layer with the nonlinear activation function σ to generate the representation h_u^{l+1} for the next step (5)

$$h_{N(u)}^{(l+1)} = \text{aggregate}(\{h_u^l, \forall v \in N(u)\}) \tag{4}$$

$$h_u^{(l+1)} = \sigma(W \cdot \text{concat}(h_u^l, h_{N(u)}^{l+1})). \tag{5}$$

3) Graph Attention Network: The core idea of GAT is to weigh the neighbor nodes of each node through the attention mechanism, and then aggregate the weighted neighbor nodes to get a new representation of the node (see Fig. 7). This allows the model to dynamically focus on different parts of neighbor nodes, making the representation of nodes more representational

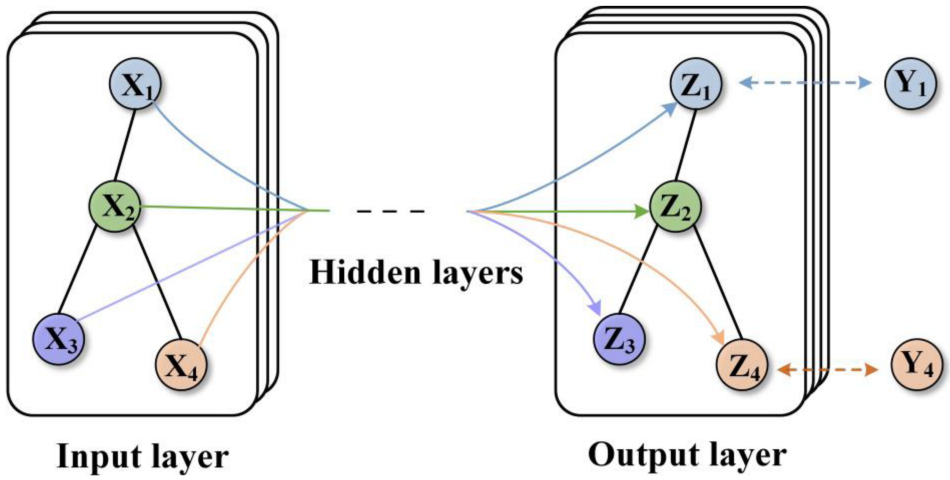


Fig. 5. GCN model.

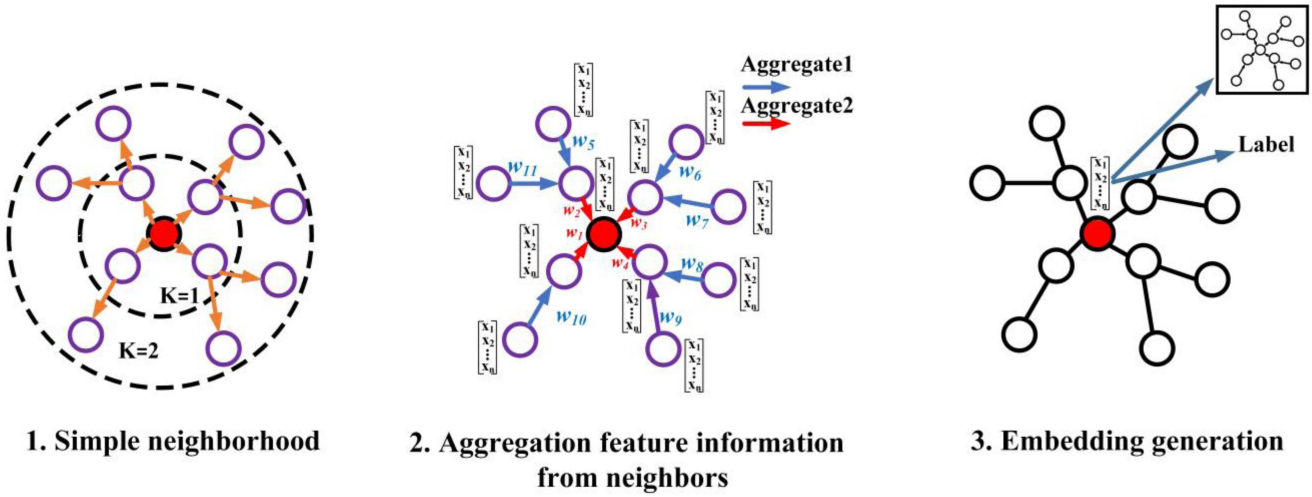


Fig. 6. GraphSAGE model.

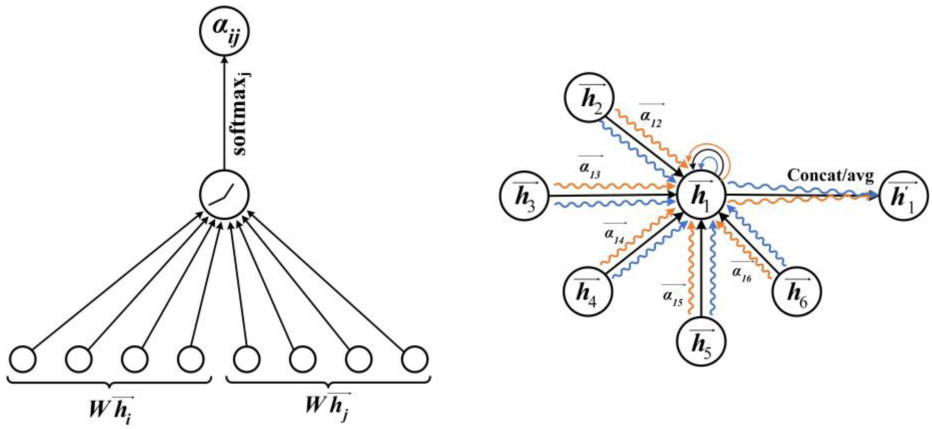


Fig. 7. GAT model.

[38]. First, the attention coefficient α_{ij} between neighbor node i and neighbor node j is calculated. Then a new representation ${h_i}^{l+1}$ of the neighbor node is obtained by weighted summation of the representation using the attention coefficient. Finally, the pooling layer or fully connected layer is superimposed for downstream learning tasks

$$e_{ij} = \alpha\left(W\overrightarrow{h_i}, W\overrightarrow{h_j}\right) \tag{6}$$

$$\alpha_{ij} = \text{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k\in N_i}\exp(e_{ik})} \tag{7}$$

$$h_i^{l+1} = \sigma\left(\sum_{j\in N_i} a_{ij} W\overrightarrow{h}_j\right). \tag{8}$$

C. Graph Construction

In the graph structure, nodes represent evaluation units and edges represent the relationship between evaluation units. To ensure an effective correlation between evaluation units, a reasonable graph $G = (V, E)$ should be constructed.

1) Evaluation Units: The slope unit has been widely used in the study of LSM [39], [40]. It is divided according to the real terrain and geomorphology, which has clear geological characteristics and can reflect the different terrain characteristics of the landslide area relatively accurately [41]. Therefore, we use the slope unit as the evaluation unit of LSM, and it is also a node in the graph structure. In this article, the division of slope units is realized by the curvature watershed method [42]. Specific division process are as follows.

1) The elevation is filtered by means, and the mean curvature value of each grid is solved with the filtered elevation data.
2) The flow direction data is solved according to the mean curvature. Then, according to the flow direction data, the depression unit is solved. Then the flow direction is taken as the base map, and the depression is taken as the water collection point to solve the basin. After vectorizing the raster data of the basin, the concave geomorphic element boundary is obtained.
3) Inverting the mean curvature data and repeating step 2 to obtain convex geomorphic element boundaries.
4) The vector merge tool is used to merge the two types of boundaries, that is, to obtain the slope unit.

When slope units are used for LSM, it is necessary to extract the attribute value of each slope unit. The statistical method we used is as follows: based on the raster data of LCFs, the regional analysis function of ArcGIS software's "Spatial analysis" tool is used to extract the mean value (continuous factors) or the majority value (discrete factors) of the raster value of LCFs within the range of each slope unit as the attribute value of this slope unit [43].

2) Connection Relationship Between Nodes:

a) Spatial Topological Relationships Graph Structure: Through the Polygon Neighbors function of ArcGIS software "Analysis Tool," the slope units can be inputted to obtain the neighbor units of each slope unit, and the graph based on the spatial topological relationships can be generated (see Fig. 8).

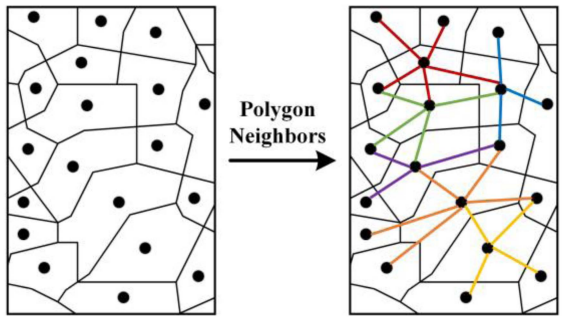


Fig. 8. Process of graph construction based on spatial topological relationships.

b) Environment Similarity Relationship Graph Structure: First, according to the composition and induced conditions of the landslide, the LCFs are divided into two categories: environmental factors and node characteristics. Second, to measure the consistency of the environment of nodes u and v, cosine similarity [24], [44] is used to calculate the similarity ($\text{Sim}_E^u(v)$)of the environment factors eigenvector x_u and x_v. Each geographic node u is only allowed to join nodes with the most consistent environment class, i.e., when $\text{Sim}_E^u(v)$ equal to MAX ($\text{Sim}_E^u(v)$, v $\in V$), v is added to the candidate set $N(u)$ of connected nodes of u. After the environmental consistency constraint, $N(u)$ is still a large number of candidate nodes for node u, and nodes with different characteristics are meaningless to the total landslide characteristics. Therefore, it is necessary to further calculate the similarity of node characteristics to filter candidate nodes. Finally, cosine similarity is used to calculate the similarity of the node eigenvector $x_{u'}$ and $x_{v'}$ to filter the redundant nodes. For each node, only the candidate node with the highest similarity is retained. That is, when $\text{Sim}_N^u(v)$ equal to MAX($\text{Sim}_N^u(v)$,$\forall v \in N(u)$), v can be added to the final neighbor set N'(u) until the size of N'(u) reaches η threshold, and finally the edge from v, $\forall v \in N'(u)$ to u is generated (see Fig. 9). In this research, the η value was set to 10.

The graph based on spatial topological relationships can be directly generated by the topological relations of slope units, and its neighborhood size is fixed, which is easy to calculate, but the landslide characteristics of neighboring units are not necessarily related. The graph based on environment similarity relationship can cluster nodes with similar features, and can also convey landslide information in different geographical environments. Therefore, we choose the environment similarity relationship to express the connection relationship between nodes.

D. Construction of LSM Model Considering Global Information

1) Transformer: The transformer model is widely used for natural language processing tasks such as language translation and generation. The transformer model is designed to address some of the problems in traditional RNN, such as the inefficiency of long sequence processing and the disappearance of gradients [45]. Compared with RNN, the transformer model adopts a novel structure, namely a self-attention mechanism.

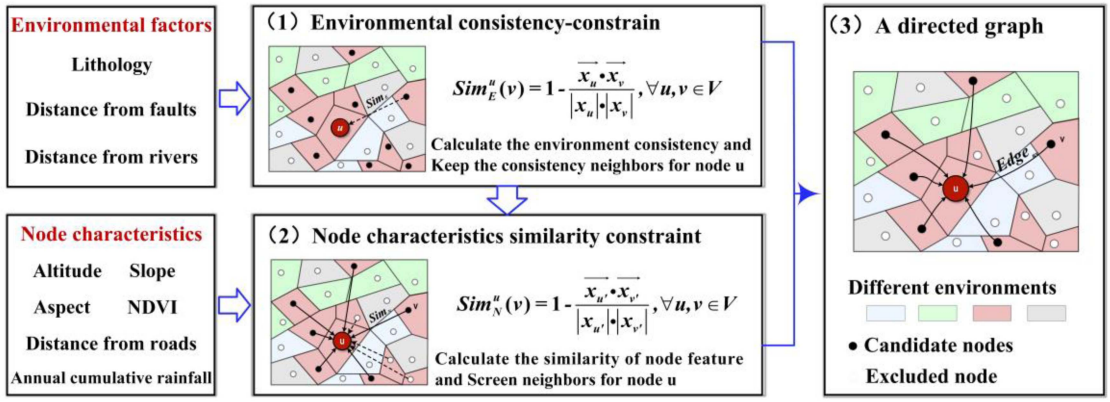


Fig. 9. Process of graph construction based on environment similarity relationship.

Self-attention mechanism is a technique used to calculate the degree of correlation of different positions in an input sequence. In transformer, the self-attention mechanism can be thought of as a way to calculate the relationships between each element in a sequence. It can compare and converge each element in the sequence with the others without the need to loop or convolve the sequence (9). The encoder in the transformer model is made up of multiple identical modules, each consisting of two sublayers: multihead self-attention (MSA) mechanism and feedforward neural network. The MSA mechanism can model the input sequence, interacting information from each location with information from other locations to get a more global representation (11). The feedforward neural network can further process the output of the MSA mechanism, providing a nonlinear transformation for the representation of each position (13).

$$\text{Attention}(Q,K,V)=\text{sofmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{9}$$

where Q, K, and V are query vectors, key vectors, and value vectors respectively.

$$\text{head}_i=\text{Attention}(QW_i^Q,KW_i^K,VW_i^V) \tag{10}$$

$$\text{MultiHead}(Q,K,V)=\text{Concat}(\text{head}_1,\text{head}_2,\ldots,\text{head}_h)W^O \tag{11}$$

where MultiHead represents the multihead attention mechanism, Concat represents the concatenation of the results of the multihead attention mechanism, and W^O is the weight matrix for linear transformation of the concatenated results. W_i^Q, W_i^K, and W_i^V is the weight moment of the query, key, and value of each head in the multihead attention mechanism

$$\text{Sublayer}(x)=\text{FeedForward}(x)\text{orMultiHead}(x,x,x) \tag{12}$$

$$\text{LayerNorm}(x+\text{Sublayer}(x))=\text{LayerNorm}(x+\text{dropout}(\text{Sublayer}(x))) \tag{13}$$

where LayerNorm (LN) represents the layer normalization operation and Sublayer represents the sublayer, each of which can be FeedForward or MultiHead*(x,x,x)* with a residual concatenation.

2) Graph–Transformer: To solve the limitation of the feature extraction capability of the LSM model based on GNN, we proposed a hybrid model for LSM, graph–transformer, which can extract global information to improve the spatial prediction accuracy of landslides. The specific structure of graph–transformer is shown in Fig. 10, which includes four key modules: structural encoding module that considers graph structure characteristics, Hop2Token module that transfers node domain information to sequence information, transformer encoder module, and attention-based readout function module that considers the influence of different neighbor nodes.

Structural encoding: In graph structure, in addition to the attribute information of nodes themselves, the structure information between nodes is also important for model learning. Adjacent nodes have similar location features, while distant nodes have different location features. The position relationship and connection strength between nodes in the graph can be comprehensively reflected by the Laplace matrix (Δ), and the feature vector obtained by the decomposition of the Laplace matrix can accurately represent the position information and feature distribution of nodes. Therefore, we use the feature vector obtained by the decomposition of the Laplace matrix to express the structure position information of the graph and concatenate it with the original attribute information matrix (X') of the graph structure to obtain the feature matrix (X)

$$\Delta=I-D^{-1/2}AD^{-1/2}=U^T\Lambda U \tag{14}$$

where A is the n $\times$ n adjacency matrix, D is the degree matrix, and Λ, U corresponds to the eigenvalues and eigenvectors respectively

$$X=X'\parallel U \tag{15}$$

where X is the feature matrix, and X' is the original attribute information matrix.

Hop2Token: After obtaining the feature matrix of the graph structure, the feature information of the neighbors of different hops of the central node can be aggregated by multiplying with the adjacency matrix. Assuming that the maximum hop of the aggregated neighbors is K, the neighbor information obtained

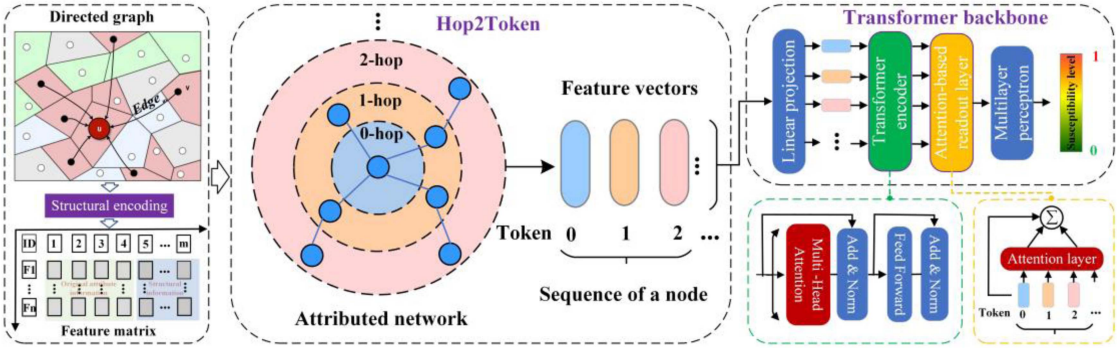


Fig. 10. Framework of graph–transformer.

Algorithm 1: The Hop2Token Algorithm.

Input: Normalized adjacency matrix A; Feature matrix X; Propagation step K

Output: Sequences of all nodes X_G

```
for k = 0 to K do
  for i = 0 to n do
    X_G[i, k] = X[i];
  end for
  X = AX;
end for
return Sequences of all nodes X_G;
```

by aggregation is $A^K\mathrm{X}$. The implementation process is shown in Algorithm 1.

$\mathrm{X_G}$ represents the characteristic information matrix after each node aggregates K-hops neighbor node information. For node u, the aggregated K-hops neighbor information can be expressed as$S_u = (x_u^0, x_u^1, \ldots, x_u^k)$. Then we map S_u to the hidden dimension d_m of the Transformer with a learnable linear projection

$$Z_u^{(0)} = [x_u^0\mathrm{E}; x_u^1\mathrm{E}; \ldots; x_u^K\mathrm{E}] \tag{16}$$

where $\mathrm{E} \in R^{d' \times d_m}$ and $Z_u^{(0)} \in R^{(K+1)\times d_m}$.

Transformer encoder: We feed the projected sequence into the Transformer encoder. The building blocks of the Transformer contain MSA and position-wise feed-forward network (FFN). LN is applied before each block. The FFN consists of two linear layers with a GELU nonlinearity

$$Z_u'^{(l)} = \mathrm{MSA}(\mathrm{LN}(Z_u^{(l-1)})) + Z_u^{(l-1)} \tag{17}$$

$$Z_u^{(l)} = \mathrm{FFN}(\mathrm{LN}(Z_u'^{(l)})) + Z_u^{(l)}. \tag{18}$$

Attention-based readout function: Considering that the neighbor node has a different influence on the central node, we propose an attention-based readout function to obtain the final representation of the central node. First, the attention coefficient between the central and neighbor node is calculated, then the attention coefficient is multiplied by each neighbor node, and finally, the final representation of the central node is obtained by adding the result of the multiplication with the original representation of the central node

$$\alpha_k = \frac{\exp((Z_0 \parallel Z_k)W_a^T)}{\sum_{i=1}^{K}\exp((Z_0 \parallel Z_i)W_a^T)} \tag{19}$$

where Z_0 is the node itself, Z_k is its k-hop representation, $W_a \in R^{1\times 2d_m}$ denotes the learnable projection, and $i = 1, \ldots, K$.

The node representation is finally aggregated as follows:

$$Z_{\mathrm{out}} = Z_0 + \sum_{k=1}^{K}\alpha_k Z_k. \tag{20}$$

TABLE I
PERFORMANCE EVALUATION METRICS OF LSM

Evaluation metric	Formula	Description
Accuracy	$\frac{TP+TN}{TP+FN+FP+TN}$	TP indicates the number of landslide samples that are correctly predicted; FN represents the number of non-landslide samples that are correctly predicted; FP is the number of landslide samples that are not correctly predicted; TN means the number of non-landslide samples that are not correctly predicted.
Precision	$\frac{TP}{TP+FP}$	
Recall	$\frac{TP}{TP+FN}$	
F1-score	$2\times\frac{Precision\times Recall}{Precision+Recall}$	

E. Model Validation Index

Confusion matrix is a common index to summarize the model classification prediction results in ML [46]. It can be used to compare the model prediction results with the real history landslide, so as to form an evaluation of the model learning results [47]. True positive (TP), false negative (FN), false positive (FP), and true negative (TN) represent the number of model prediction results divided according to the threshold in four cases, respectively. Using these values, the following indicators can be calculated (see Table I).

In addition, the receiver operating characteristic curve (ROC) is a synthesis of the confusion matrix under different thresholds,

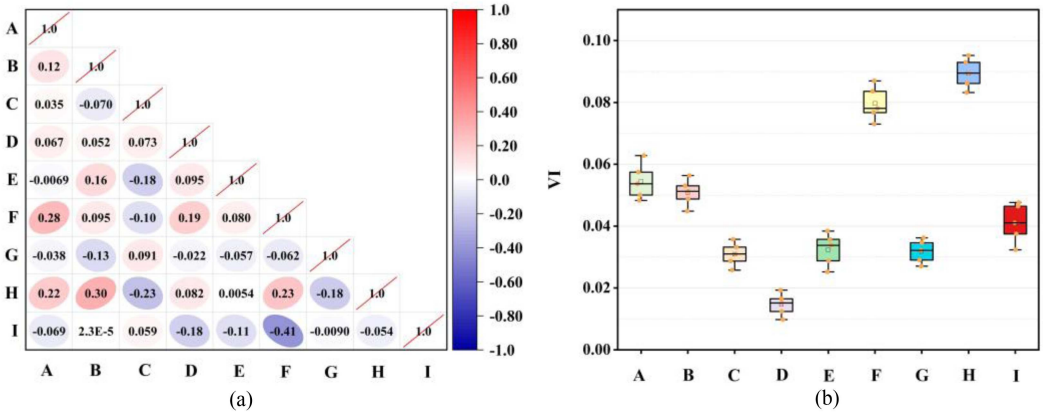


Fig. 11. Results of LCFs analysis. (a) Results of Pearson correlation analysis. (b) Variable importance of LCFs using Relief-F.

which is widely used in the evaluation of LSM. ROC takes the FP rate as the horizontal coordinate and the TP rate as the vertical coordinate, reflecting the continuous change of data specificity and sensitivity. The area under the curve (AUC) of ROC can directly reflect the result, and the closer the value of AUC is to 1, the better the model effect is [48], [49].

IV. Results

A. Analysis of Contributing Factors

Pearson correlation method was used to analyze the correlation of LCFs, and the results showed that the correlation coefficients among LCFs were less than 0.5, indicating that there was a weak correlation among LCFs [see Fig. 11(a)]. In addition, the Relief-F method was used to analyze the importance of factors, and the results showed that the importance index of LCFs was greater than 0, indicating that the selected factors had positive effects on the development of landslide [see Fig. 11(b)]. Rainfall, which has the highest importance index, is consistent with the leading factor of landslide development in BRB. The above indicates that the LCFs selected in this article are both independent and important. It can be used in the subsequent evaluation and calculation of LSM.

B. Landslide Distribution in Different Graph Structures and Models

To directly reflect the rationality of the constructed graph structures and the performance of the model, we applied the same dataset to different graph structures and models and measured the advantages and disadvantages of different graph structures and models by comparing the prediction results. Fig. 12 shows the LSM on different graph structures and models, and the natural breakpoint method [50] is used to classify landslide susceptibility into five levels: very low susceptibility (VLS), low susceptibility (LS), moderate susceptibility (MS), high susceptibility (HS), and very high susceptibility (VHS). In addition, to quantify the difference in results under different graph structures and models, we calculated the proportion of each susceptibility region in the LSM and the proportion distribution of historical landslides in each susceptibility region.

Under the spatial topological relationships graph, the MS and HS areas predicted by the GCN model were widely distributed, accounting for 19.59% and 19.48% of the study area, respectively, but only 44.74% of landslides were located in the VHS areas [see Fig. 12(a)], which could not provide effective landslide reference information. In the GraphSAGE model, about 66.36% of the landslides were located in the VHS region, but about 15% of the landslides were still in the MS and LS regions [see Fig. 12(c)], and the prediction results had some false alarms. In the prediction results of the GAT model, most of the susceptibility areas are LS (29.92%), and the VHS areas are dispersed, only 49.77% of landslides coincide with the VHS areas [see Fig. 12(e)]. In the prediction results of the proposed model, the VLS and VHS areas are concentrated in large areas, accounting for 57.80% and 11.26% of the study area, respectively, and about 62.01% of the landslides are located in the VHS areas [see Fig. 12(g)]. In contrast, under the environment similarity relationship graph, the LSM of the four models has been greatly improved, and the predicted VHS areas are more consistent with the real distribution situation. Among them, the proportion of historical landslides in the VHS areas of the four models is 92.11%, 74.49%, 78.60%, and 89.70%, respectively [see Fig. 12(b), (d), (f), (h)]. The distribution trend of the VHS and HS areas along the river was also present. This indicates that the environment similarity relationship graph is more reasonable, can more effectively mine the correlation between landslide and DPE, and the LSM obtained can better reflect the actual landslide distribution, which is of great significance for improving the landslide susceptibility in unknown areas.

Under the same graph structure, the number of historical landslides shows an increasing trend with the increase of landslide susceptibility level of the four models, which indicates that the four models have certain positive effects on the prediction of landslide susceptibility. However, from the distribution of HS and VHS areas, the LSM of the proposed model is the best, the predicted landslide area is more precise, and most of the landslides fall in the VHS areas. In addition, compared with the LSM of the other three models, the landslide susceptibility

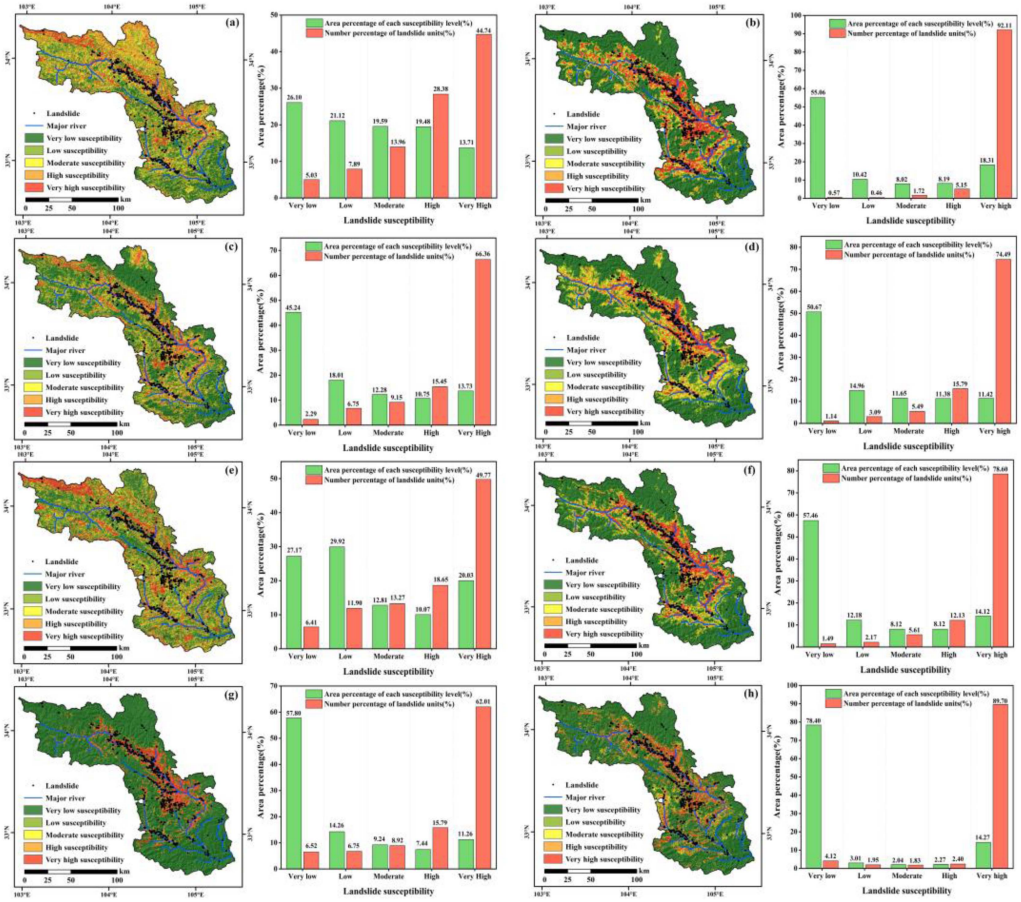


Fig. 12. Landslide distribution in different graph structures and models for both spatial topological relationships graph (left) and environment similarity relationship graph (right). (a)–(b) GCN. (c)–(d) GraphSAGE. (e)–(f) GAT. (g)–(h) Proposed model.

TABLE II
COMPARISON OF THE GRAPH STRUCTURE AND LSM METHODS

	spatial topological relationships				environment similarity relationship			
	GCN	GraphSAGE	GAT	Proposed	GCN	GraphSAGE	GAT	Proposed
Accuracy	0.7138	0.7529	0.6935	**0.7974**	0.8612	0.8545	0.8669	**0.8724**
Precision	0.6861	0.7059	0.6521	**0.7878**	0.7932	0.7954	0.7954	**0.8434**
Recall	0.7881	0.8176	0.7175	**0.8441**	0.9011	0.9163	0.9163	**0.9372**
F1-score	0.7336	0.7703	0.6832	**0.8007**	0.8713	0.8516	0.8516	**0.8756**
AUC	0.7736	0.8127	0.7133	**0.8674**	0.9223	0.9256	0.9096	**0.9446**

distribution of the proposed model is more reasonable under different graph structures, and the VHS region divided by the model is the most appropriate to the distribution of historical landslides. This shows that the model constructed in this article has excellent robustness and can perform well under two different graph structures. It can fully explore the hidden relationship between landslides and DPE, effectively reducing the false alarm rate of landslides. The results can provide references for future landslide risk management.

C. Model Validation and Comparison

To reflect the influence of different graph structures on model performance and quantify the differences between model learning abilities, we calculated the evaluation indexes of four models under different graph structures. Table II shows that under the environment similarity relationship graph, the evaluation index results (accuracy, precision, recall, F1-score, and AUC) of the four models are all higher than those under the spatial topological relationships graph. Among them, the GCN model increased by 20.65%, 15.61%, 14.34%, 18.77%, and 19.22%, and the GraphSAGE model increased by 13.49%, 12.68%, 12.07%, 10.55%, and 13.89%, respectively. The GAT model was improved by 25.12%, 21.98%, 27.71%, 24.65%, and 27.52%, respectively, and the proposed model was improved by 9.41%, 7.06%, 11.03%, 9.35%, and 8.90%, respectively. This shows that the environment similarity relationship graph can promote the learning of the model, help the model learn the most effective features, and effectively improve the prediction accuracy of the model. In addition, under the same

graph structure, the model proposed in this article is superior to the other three models. The calculated evaluation index results such as accuracy, precision, recall, F1-score, and AUC, etc., are increased under the spatial topological relationships graph by at least 5.91%, 11.60%, 3.24%, 3.95%, and 6.73%, respectively. Under the environment similarity relationship graph, the improvement is at least 0.63%, 6.03%, 2.28%, 0.49%, and 2.05%, respectively. This indicates that the proposed model in this article has strong feature learning ability, can deeply explore the hidden relationship between landslide and LCFs, and can realize the full correlation between landslide and DPE.

Statistical error distribution is a method to evaluate the performance of a model, which can reflect the accuracy and degree of bias of the model during the prediction process. For this purpose, we calculate the mean squared error (MSE) and root mean squared error (RMSE) of the four models on the testing dataset under different graph structures. Fig. 13 shows that under the spatial topological relationships graph, the maximum values of MSE and RMSE are obtained by the GCN model (0.1982, 0.4452) [see Fig. 13(a)], while the minimum values are the model proposed in this article (0.1413, 0.3579) [see Fig. 13(g)]. Under the environment similarity relationship graph, the maximum values of MSE and RMSE are obtained by the GAT model (0.1025, 0.3201) [see Fig. 13(f)], and the minimum values are the model proposed in this article (0.0595, 0.2441) [see Fig. 13(h)]. This shows that the proposed model has good performance. In addition, the MSE and RMSE of the four models under the environment similarity relationship graph are lower than those under the spatial topological relationships graph. It can be further demonstrated that the environment similarity relationship graph can effectively improve the prediction accuracy of the model and reduce the false alarm rate of the model.

To prove the reliability of the proposed model in this article, it is possible to verify the landslide disaster in the HS area predicted by the proposed model through field investigation. In order to better display the observation results, a landslide area that does not appear in the landslide catalog is selected to evaluate the accuracy of the model. Fig. 14 shows LSM and UAV photos of the selected area. The landslide occurred in June 2024. Under the spatial topological relationship graph structure, the proposed model predicts that this region is VHS, and the other three models predict that the region is MS and HS. Under the environment similarity relationship graph structure, the proposed model predicts that this region is VHS, while the other three models predict that this region is HS. The proposed model can effectively predict the location of landslides, which is of great significance for landslide prevention and management. In addition, under the environment similarity relationship graph structure, the prediction results of the four models are all better than that under the spatial topological relationship graph structure, which further proves that the environment similarity relationship graph structure can weaken the impact of environmental differences on model performance and effectively improve the prediction accuracy of the model.

V. Discussion

A. Advantage of the Proposed Graph Structure

In complex environmental scenes, the lack of effective correlation between landslide and DPE is the main reason for the low reliability of LSM. In this study, compared with the environment similarity relationship graph, the spatial topological relationships graph shows lower precision (see Table II). Among the prediction results of the four models, the LSM based on the environment similarity relationship graph is more reasonable than that based on the spatial topological relationships graph (see Fig. 12). Most landslides are located in VHS and HS areas, and only a few landslides are located in VLS and LS areas. In addition, the models supported by the environment similarity relationship graph show high model accuracy.

Based on the common sense of geography that "the more similar the geographical environment is, the more similar the geographical characteristics are" [51], it is believed that the same type of landslide samples are close and adjacent in the environmental characteristic space, and different types of landslide samples are separated. In this understanding, similar geographical and geological environments, regardless of geographical location, will form landslide samples with similar characteristics [24], [26]. Therefore, the correlation between subenvironmental regions with similar geological conditions should be considered in LSM. The characteristics of landslides can influence each other among similar subenvironment areas. In this interrelated and influencing environment, the clustering and correlation of similar landslide characteristics are formed, which can better help to summarize landslide characteristics. Therefore, based on the environment similarity relationship, this study established a geographic association network with similarity constraints of DPE, and defined the graph structure as the similarity of environment and features between nodes, so as to improve the ability of landslide characteristics induction. Specifically, first, nodes of different geographical subenvironments are initially divided by using the consistency of features of different geographical environmental factors, and the correlation between subenvironment categories is established by using the similarity of environmental characteristics. These environmental category divisions limit the information transmission of nodes in different geographical environments, so that landslide characteristics can spread in similar environments as much as possible. Then, in each subenvironment, a specific node association relationship is established by using the similarity of node feature factors, so that nodes with similar features are clustered together.

The formation of a landslide is mainly on a specific slope, which is determined by its own geographical and geological conditions. Although there is a certain spatial aggregation phenomenon in a large range, it is affected by the heterogeneity of the geographical environment, that is, the geographical environment in different locations is inconsistent and the geographical and geological conditions in adjacent spaces are inconsistent. Therefore, there is no similar relationship between neighboring units in the geographical space of landslides. Spatial topological

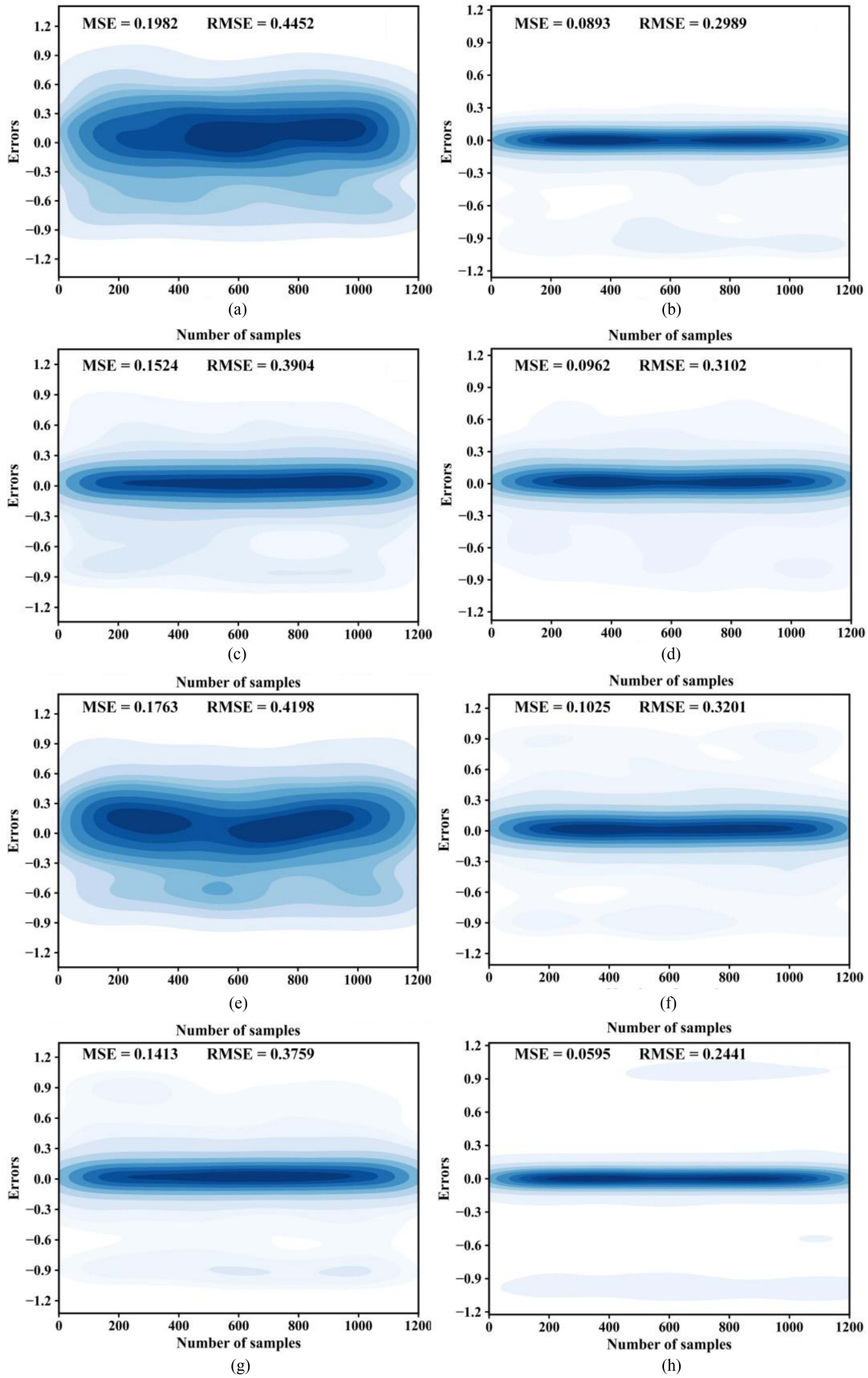


Fig. 13. Errors in MSE and RMSE for both the spatial topological relationships graph (left) and environment similarity relationship graph (right): (a)–(b) GCN. (c)–(d) GraphSAGE. (e)–(f) GAT. (g)–(h) Proposed model.

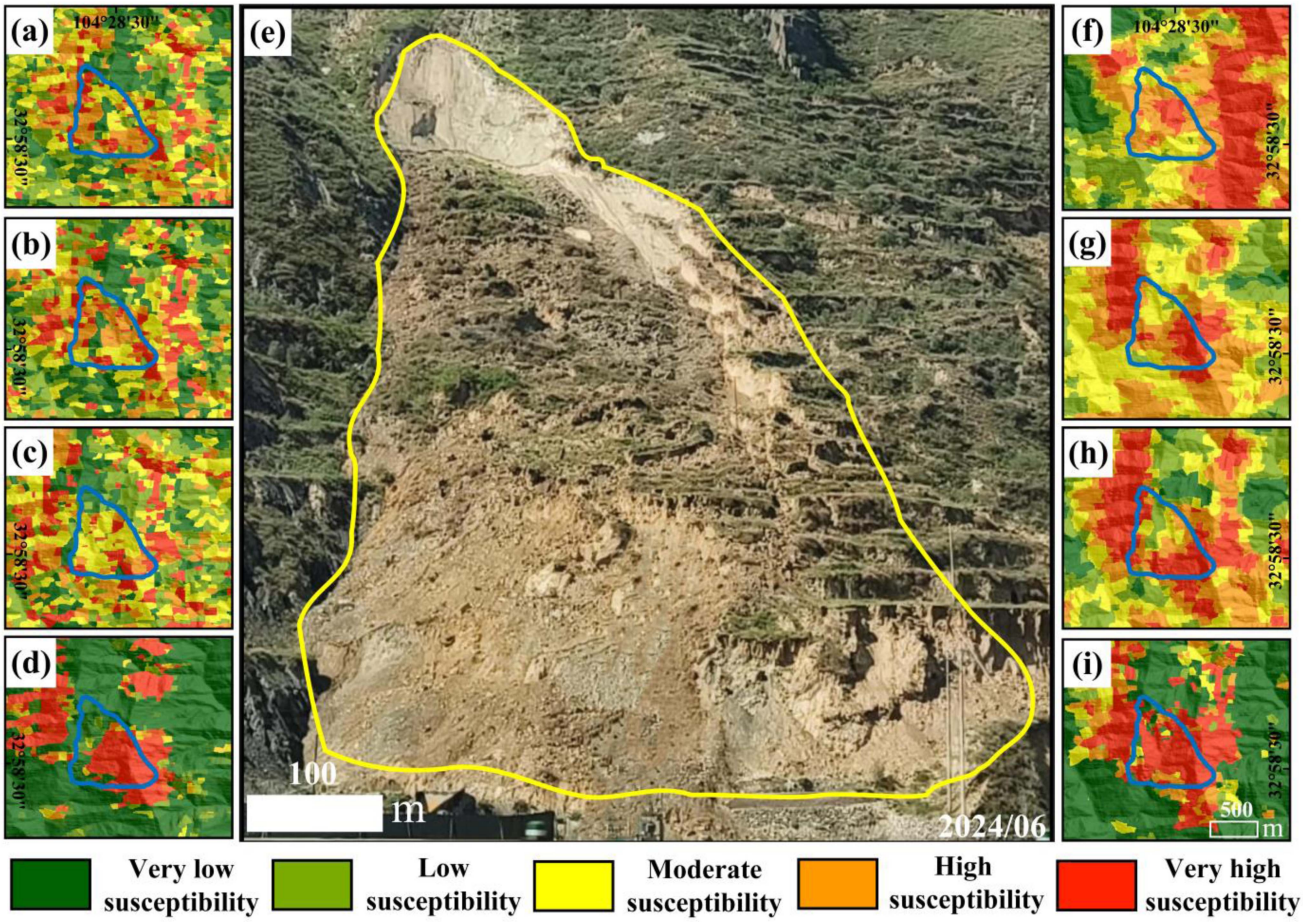


Fig. 14. Analysis of model prediction results based on fieldwork. (a)–(d) LSM generated by GCN, GraphSAGE, GAT and the proposed model based on the spatial topological relationship. (e) UAV photo. (f)–(i) LSM generated by GCN, GraphSAGE, GAT, and the proposed model based on the environment similarity relationship. The blue and yellow lines represent the landslide boundaries.

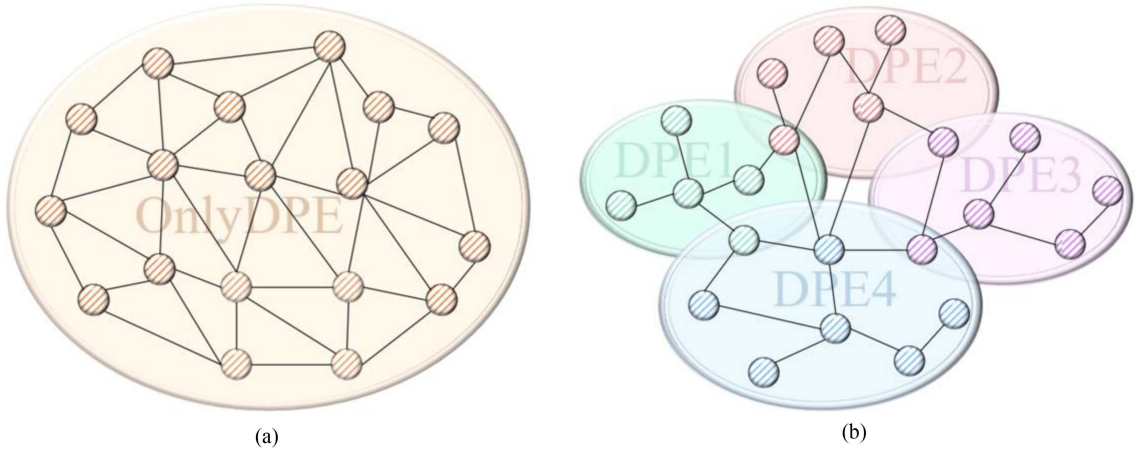


Fig. 15. Schematic diagram of different graph structures. (a) Spatial topological relationships graph. (b) Environment similarity relationship graph.

relationships make use of the similarity of location characteristics in geographic space [see Fig. 15(a)]. Its advantage is that it can correlate the information of nodes around samples and reduce the influence of sample deviation error. However, the scale at which the similarity of neighboring geographical units is very small. On a large scale, especially when the nodes of the simplified topographic units are far apart, it is more affected by the heterogeneous geographical environment, and the features of many neighboring nodes are not correlated. In contrast, the proposed environment similarity relationship graph can break

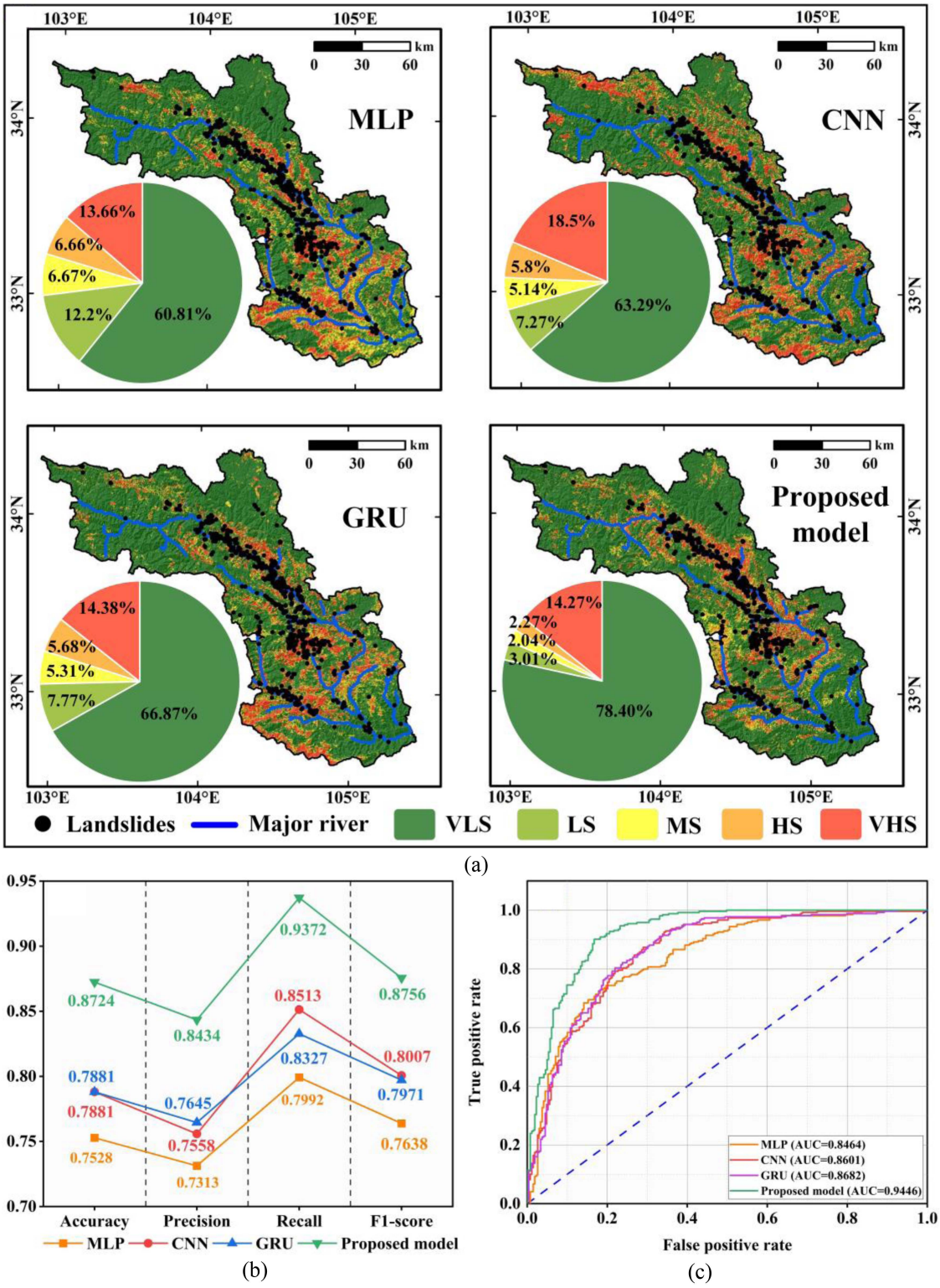


Fig. 16. Landslide susceptibility originating from the MLP, CNN, GRU, and the proposed model. (a) LSM. (b) Evaluation indexes results. (c) AUC values.

through spatial constraints and cluster related nodes together. At the same time, it is not only the correlation of complete characteristics but also the correlation after environmental classification with purpose, which is more conducive to the classification and induction of regional landslide characteristics [see Fig. 15(b)].

B. *Advantage of the Proposed Model*

In this study, we propose a new LSM model that can consider the characteristics of DPE and global information. The model combines the advantages of graph and transformer to achieve a

wide range of reliable LSM. In the previous studies on LSM, scholars mostly used traditional ML models. These ML models only pay attention to the morphological characteristics of the landslide itself and do not fully consider the characteristics of the distant but directly related DPE, resulting in a high false alarm rate of the landslide. In addition, these models can obtain reliable LSM in simple small scenes, but for complex environmental scenes, there are still limitations of relying on sample size and insufficient feature extraction ability, resulting in low accuracy and reliability of LSM. Transformer is a neural network model based on the self-attention mechanism, which can calculate weight and position coding to obtain global information [15], [16], [27]. Therefore, we propose to use graph and transformer to realize comprehensive and reliable LSM. Specifically, graph and transformer work together to enhance each other. By using transformer, the receptive field of the graph can be extended, including those related nodes that are far from the central node, thus enhancing the information transmission capability of the graph. Conversely, the graph can also help the transformer capture complex topological relationships and efficiently aggregate related nodes from neighboring areas.

In this study, compared with the other three kinds of GNN models, the proposed model in this article shows excellent performance for spatial topological relationships and environment similarity relationship graphs (see Table II). To further demonstrate the advantages of the proposed model, we compare it with other DL models widely used in LSM, such as MLP, CNN, and GRU. The results of the comparison are shown below (see Fig. 16). The LSM of the four models shows the same distribution trend on the whole, but there are great differences in the specific distribution. The VHS region accounted for more than 13% and the VLS region accounted for more than 60% of the four models [see Fig. 16(a)]. Among them, the VHS region of the CNN model occupies the highest proportion, which is 18.5%. The VLS region of the proposed model occupies the highest proportion, which is 78.4%. The MS region of MLP, CNN, and GRU models is above 5%. In contrast, the LSM of the proposed model is more reasonable. The VHS areas show a trend of distribution along the river, most historical landslides fall into the VHS areas, and the sum of VHS and VLS areas account for more than 92% of the study area, which indicates that the proposed model has excellent recognition ability and can accurately identify the landslides and nonlandslides. In addition, to quantify the differences in the learning ability of different models, we statistically calculated the evaluation indicators of the four models. Compared with MLP, CNN, and GRU, accuracy, precision, recall, and F1-score are 10.7%, 10.32%, 10.09%, and 9.35% higher, respectively [see Fig. 16(b)]. The AUC value of the proposed model is as high as 0.94 [see Fig. 16(c)], which is more than 8.8% higher than that of the other three models, indicating that the proposed model has excellent learning ability and can deeply explore the relationship between landslide and LCFs, effectively improving the reliability of large-scale LSM.

MLP and GRU models accept sequential features [10], while CNN models accept spatial features [8]. From the LSM and evaluation indicator results of the four models, the proposed model has the best feature learning ability, which indicates that compared with the feature learning methods of sequence and spatial features, the proposed method of using graph and transformer to learn landslide characteristics is more effective and comprehensive. This can also indirectly prove the rationality of the proposed graph structure. The graph structure is used to correlate landslide and DPE, which significantly improves the reliability of the LSM and the accuracy of the model.

VI. Conclusion

In this article, we propose an LSM model with graph–transformer that can consider the correlation between landslide and DPE, and global information. The main goal of this study is to reduce the impact of environmental differences and the limitations of model feature extraction ability on LSM. Some novel and significant conclusions can be drawn as follows.

1) Constructing the environment similarity relationship graph is an effective method to realize the full correlation between landslide and DPE. Compared with the models based on the spatial topological relationships graph (GCN, GraphSAGE, GAT, and Proposed model), the AUC values of the models based on the environment similarity relationship graph are increased by 19.22%, 13.89%, 27.52%, and 8.9%, respectively, indicating that the environment similarity relationship graph can effectively improve the model accuracy and reduce the impact of environmental differences on the model.
2) The introduction of transformer module that considers global information is an important means to improve model performance. Compared with GCN, GraphSAGE, and GAT models, the AUC value of the proposed model is more than 6.73% higher under the spatial topological relationships graph, and more than 2.05% higher under the environment similarity relationship graph. In addition, the AUC value of the proposed model is more than 8.8% higher than that of traditional ML models (MLP, CNN, and GRU). This shows that the proposed model has strong feature mining ability and can learn landslide features more comprehensively.

References

[1] Y. He et al., "An integrated neural network method for landslide susceptibility assessment based on time-series InSAR deformation dynamic features," *Int. J. Digit. Earth*, vol. 17, no. 1, Dec. 2024, Art. no. 2295408, doi: 10.1080/17538947.2023.2295408.

[2] C. Martinello, M. Delchiaro, G. Iacobucci, C. Cappadonia, E. Rotigliano, and D. Piacentini, "Exploring the geomorphological adequacy of the landslide susceptibility maps: A test for different types of landslides in the Bidente river basin (northern Italy)," *CATENA*, vol. 238, Apr. 2024, Art. no. 107835, doi: 10.1016/j.catena.2024.107835.

[3] X. Wei et al., "Improving pixel-based regional landslide susceptibility mapping," *Geosci. Front.*, vol. 15, no. 4, Jul. 2024, Art. no. 101782, doi: 10.1016/j.gsf.2024.101782.

[4] L. Yu, Y. Wang, and B. Pradhan, "Enhancing landslide susceptibility mapping incorporating landslide typology via stacking ensemble machine learning in Three Gorges Reservoir, China," *Geosci. Front.*, vol. 15, no. 4, Jul. 2024, Art. no. 101802, doi: 10.1016/j.gsf.2024.101802.

[5] Q. Zhu, L. Chen, H. Hu, S. Pirasteh, H. Li, and X. Xie, "Unsupervised feature learning to improve transferability of landslide susceptibility representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3917–3930, 2020, doi: 10.1109/JSTARS.2020.3006192.

[6] F. Huang et al., "Uncertainty pattern in landslide susceptibility prediction modelling: Effects of different landslide boundaries and spatial shape expressions," *Geosci. Front.*, vol. 13, no. 2, Mar. 2022, Art. no. 101317, doi: 10.1016/j.gsf.2021.101317.

[7] Y. He et al., "Thaw slump susceptibility mapping based on sample optimization and ensemble learning techniques in Qinghai-Tibet railway corridor," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5443–5459, 2024, doi: 10.1109/JSTARS.2024.3368039.

[8] Y. He et al., "A unified network of information considering superimposed landslide factors sequence and pixel spatial neighbourhood for landslide susceptibility mapping," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 104, Dec. 2021, Art. no. 102508, doi: 10.1016/j.jag.2021.102508.

[9] B. Gao et al., "Landslide risk evaluation in shenzhen based on stacking ensemble learning and InSAR," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1–18, 2023, doi: 10.1109/JSTARS.2023.3291490.

[10] Z. Zhao et al., "A comparative study of different neural network models for landslide susceptibility mapping," *Adv. Space Res.*, vol. 70, no. 2, pp. 383–401, Jul. 2022, doi: 10.1016/j.asr.2022.04.055.

[11] R. Wei, C. Ye, T. Sui, Y. Ge, Y. Li, and J. Li, "Combining spatial response features and machine learning classifiers for landslide susceptibility mapping," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 107, Mar. 2022, Art. no. 102681, doi: 10.1016/j.jag.2022.102681.

[12] R. Yuan and J. Chen, "A novel method based on deep learning model for national-scale landslide hazard assessment," *Landslides*, vol. 20, pp. 2379–2403, Jul. 2023, doi: 10.1007/s10346-023-02101-y.

[13] Y. He et al., "An identification method of potential landslide zones using InSAR data and landslide susceptibility," *Geomatics, Natural Hazards Risk*, vol. 14, no. 1, Dec. 2023, Art. no. 2185120, doi: 10.1080/19475705.2023.2185120.

[14] T. Chen et al., "BisDeNet: A new lightweight deep Learning-Based framework for efficient landslide detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 3648–3663, 2024, doi: 10.1109/JSTARS.2024.3351873.

[15] Z. Zhao, T. Chen, J. Dou, G. Liu, and A. Plaza, "Landslide susceptibility mapping considering landslide Local-Global features based on CNN and transformer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 7475–7489, 2024, doi: 10.1109/JSTARS.2024.3379350.

[16] T. Chen, Q. Wang, Z. Zhao, G. Liu, J. Dou, and A. Plaza, "LCFSTE: Landslide conditioning factors and swin transformer ensemble for landslide susceptibility assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 6444–6454, 2024, doi: 10.1109/JSTARS.2024.3373029.

[17] X. Wang, W. Deng, Z. Meng, and D. Chen, "Hybrid-attention mechanism based heterogeneous graph representation learning," *Expert Syst. Appl.*, vol. 250, Sep. 2024, Art. no. 123963, doi: 10.1016/j.eswa.2024.123963.

[18] X. Lu, L. Xie, L. Xu, R. Mao, X. Xu, and S. Chang, "Multimodal fused deep learning for drug property prediction: Integrating chemical language and molecular graph," *Comput. Struct. Biotechnol. J.*, vol. 23, Apr. 2024, Art. no. S2001037024001132, doi: 10.1016/j.csbj.2024.04.030.

[19] F. Jin, R. Li, and H. Wu, "Graph neural network-based similarity relationship construction model for geospatial services," *Geo-Spat. Inf. Sci.*, pp. 1–15, Nov. 2023, doi: 10.1080/10095020.2023.2273820.

[20] Y. (Lance) Liu et al., "Interpretable Chirality-Aware graph neural network for quantitative structure activity relationship modeling in drug discovery," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2023, vol. 37, no. 12, pp. 14356–14364, doi: 10.1609/aaai.v37i12.26679.

[21] M. Tzes, N. Bousias, E. Chatzipantazis, and G. J. Pappas, "Graph neural networks for Multi-Robot active information acquisition," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2023, pp. 3497–3503, doi: 10.1109/ICRA48891.2023.10160723.

[22] T. Liu and H. Meidani, "Optimizing seismic retrofit of Bridges: Integrating efficient graph neural network surrogates and transportation equity," in *Proc. Cyber-Phys. Syst. Internet Things Week*, May 2023, pp. 367–372, doi: 10.1145/3576914.3587503.

[23] X. Wang et al., "Landslide susceptibility evaluation based on active deformation and graph convolutional network algorithm," *Front. Earth Sci.*, vol. 11, Jan. 2023, Art. no. 1132722, doi: 10.3389/feart.2023.1132722.

[24] H. Zeng et al., "Graph neural networks with constraints of environmental consistency for landslide susceptibility evaluation," *Int. J. Geographical Inf. Sci.*, vol. 36, no. 11, pp. 2270–2295, Nov. 2022, doi: 10.1080/13658816.2022.2103819.

[25] Q. Yang et al., "A novel CGBoost deep learning algorithm for coseismic landslide susceptibility prediction," *Geosci. Front.*, vol. 15, no. 2, Mar. 2024, Art. no. 101770, doi: 10.1016/j.gsf.2023.101770.

[26] A. X. Zhu et al., "A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods," *CATENA*, vol. 183, Dec. 2019, Art. no. 104188, doi: 10.1016/j.catena.2019.104188.

[27] L. Wu et al., "Landslide mapping based on a hybrid CNN-transformer network and deep transfer learning using remote sensing images with topographic and spectral features," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 126, Feb. 2024, Art. no. 103612, doi: 10.1016/j.jag.2023.103612.

[28] Y. Yang, T. Chen, T. Lei, B. Du, A. K. Nandi, and A. Plaza, "SEDANet: A new siamese ensemble difference attention network for building change detection in remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4405316, doi: 10.1109/TGRS.2024.3381751.

[29] T. Qi et al., "Formation and distribution of landslides controlled by thrust-strike-slip fault zones and fluvial erosion in the Western Qinling Mountains, China," *Eng. Geol.*, vol. 323, Sep. 2023, Art. no. 107209, doi: 10.1016/j.enggeo.2023.107209.

[30] Q. Sun, L. Zhang, X. L. Ding, J. Hu, Z. W. Li, and J. J. Zhu, "Slope deformation prior to Zhouqu, China landslide from InSAR time series analysis," *Remote Sens. Environ.*, vol. 156, pp. 45–57, Jan. 2015, doi: 10.1016/j.rse.2014.09.029.

[31] J. Jin et al., "Prediction of river damming susceptibility by landslides based on a logistic regression model and InSAR techniques: A case study of the Bailong River Basin, China," *Eng. Geol.*, vol. 299, Mar. 2022, Art. no. 106562, doi: 10.1016/j.enggeo.2022.106562.

[32] Y. Fang et al., "The Displacement Analysis and Prediction of a Creeping Ancient Landslide at Suoertou, Zhouqu County, China," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 4139–4163, 2024, doi: 10.1109/JSTARS.2024.3357520.

[33] X. Fu, Y. Liu, Q. Zhu, D. Ge, Y. Li, and H. Zeng, "Reliable assessment approach of landslide susceptibility in broad areas based on optimal slope units and negative samples involving priori knowledge," *Int. J. Digit. Earth*, vol. 15, no. 1, pp. 2495–2510, Dec. 2022, doi: 10.1080/17538947.2022.2159549.

[34] B. Gao, Y. He, X. Chen, H. Chen, W. Yang, and L. Zhang, "A deep neural network framework for landslide susceptibility mapping by considering Time-Series rainfall," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5946–5969, 2024, doi: 10.1109/JSTARS.2024.3370218.

[35] Z. Fang, Y. Wang, L. Peng, and H. Hong, "A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping," *Int. J. Geographical Inf. Sci.*, vol. 35, no. 2, pp. 321–347, Feb. 2021, doi: 10.1080/13658816.2020.1808897.

[36] T. N. Kipf and M. Welling, "Semi-Supervised classification with graph convolutional networks," Feb. 22, 2017, *arXiv: arXiv:1609.02907*, Accessed: Aug. 30, 2023. [Online]. Available: http://arxiv.org/abs/1609.02907

[37] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs," Sep. 10, 2018, *arXiv: arXiv:1706.02216*, Accessed: Oct. 08, 2023. [Online]. Available: http://arxiv.org/abs/1706.02216

[38] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," Feb. 04, 2018, *arXiv: arXiv:1710.10903*.

[39] Y. Hua, X. Wang, Y. Li, P. Xu, and W. Xia, "Dynamic development of landslide susceptibility based on slope unit and deep neural networks," *Landslides*, vol. 18, no. 1, pp. 281–302, Jan. 2021, doi: 10.1007/s10346-020-01444-0.

[40] D. Sun et al., "Assessment of landslide susceptibility along mountain highways based on different machine learning algorithms and mapping units by hybrid factors screening and sample optimization," *Gondwana Res.*, vol. 123, pp. 89–106, Nov. 2023, doi: 10.1016/j.gr.2022.07.013.

[41] G. Yan, D. Lu, S. Li, S. Liang, L. Xiong, and G. Tang, "Optimizing slope unit-based landslide susceptibility mapping using the priority-flood flow direction algorithm," *CATENA*, vol. 235, Feb. 2024, Art. no. 107657, doi: 10.1016/j.catena.2023.107657.

[42] B. Romstad and B. Etzelmüller, "Mean-curvature watersheds: A simple method for segmentation of a digital elevation model into terrain units," *Geomorphology*, vol. 139–140, pp. 293–302, Feb. 2012, doi: 10.1016/j.geomorph.2011.10.031.

[43] V. Canavesi et al., "Different approaches to use morphometric attributes in landslide susceptibility mapping based on Meso-Scale spatial units: A case study in rio de janeiro (Brazil)," *Remote Sens.*, vol. 12, no. 11, Jun. 2020, Art. no. 1826, doi: 10.3390/rs12111826.

[44] Q. Zhang et al., "A landslide susceptibility assessment method considering the similarity of geographic environments based on graph neural network," *Gondwana Res.*, vol. 132, pp. 323–342, Aug. 2024, doi: 10.1016/j.gr.2024.04.013.

[45] A. Vaswani et al., "Attention Is All You Need," Jun. 12, 2017. [Online]. Available: http://arxiv.org/abs/1706.03762

[46] P. Dou, C. Huang, W. Han, J. Hou, Y. Zhang, and J. Gu, "Remote sensing image classification using an ensemble framework without multiple classifiers," *ISPRS J. Photogramm. Remote Sens.*, vol. 208, pp. 190–209, Feb. 2024, doi: 10.1016/j.isprsjprs.2023.12.012.

[47] J. Dou et al., "Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning," *Sci. Total Environ.*, vol. 720, Jun. 2020, Art. no. 137320, doi: 10.1016/j.scitotenv.2020.137320.
[48] J. Lu et al., "Ensemble learning landslide susceptibility assessment with optimized non-landslide samples selection," *Geomatics, Natural Hazards Risk*, vol. 15, no. 1, Dec. 2024, Art. no. 2378176, doi: 10.1080/19475705.2024.2378176.
[49] P. Dou, H. Shen, C. Huang, Z. Li, Y. Mao, and X. Li, "Large-scale land use/land cover extraction from Landsat imagery using feature relationships matrix based deep-shallow learning," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 129, May 2024, Art. no. 103866, doi: 10.1016/j.jag.2024.103866.
[50] L. J. Wang, M. Guo, K. Sawada, J. Lin, and J. Zhang, "Landslide susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models," *CATENA*, vol. 135, pp. 271–282, 2015, doi: https://doi.org/10.1016/j.catena.2015.08.007.
[51] A. X. Zhu and M. Turner, "How is the Third Law of Geography different?," *Ann. GIS*, vol. 28, no. 1, pp. 57–67, Jan. 2022, doi: 10.1080/19475683.2022.2026467.

Qing Zhang received the B.E. degree in surveying and mapping engineering, in 2022, from the Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in mapping science and technology with the Faculty of Geomatics.

His research interests include InSAR data processing technology, machine learning, and remote sensing image processing.

Yi He (Member, IEEE) received the B.S. degree in geographic information system and the Ph.D. degree in earth system science from the Lanzhou Jiaotong University, Lanzhou, China, in 2011 and 2016, respectively.

He has been a Postdoctoral Researcher with the school of environment and municipal engineering, Lanzhou Jiaotong University. He is currently a Professor with the Faculty of Geomatics, Lanzhou Jiaotong University. His research interests include disaster remote sensing, image processing, and time series InSAR prediction based on deep learning.

Yalei Zhang received the B.E. degree in surveying and mapping engineering from the Lanzhou Jiaotong University, Lanzhou, China, in 2022.

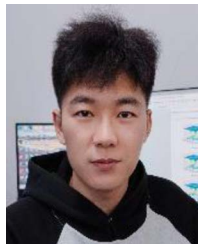
She is currently working toward the Zhangye Mineral Exploration Institute of Gansu Nonferrous Metals Geological Exploration Bureau, Zhangye, Gansu, China. Her research interests include GIS and Planning and Natural Resource Informatization.

Jiangang Lu received the B.E. degree in remote sensing science and technology from the Shandong University of Science and Technology, Qingdao, China, in 2020. He is currently working toward the M.S. degree in mapping science and technology with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, China.

His research interests include landslide hazard risk assessment, machine learning, and remote sensing image information extraction.

Lifeng Zhang received the M.S. degree in cartography and geographic information system and the Ph.D. degree in environmental science and engineering from the Lanzhou Jiaotong University, Lanzhou, China, in 2010 and 2017, respectively.

He is currently a Professor with the Faculty of Geomatics, Lanzhou Jiaotong University. His research interests include ecological environment remote sensing monitoring and analysis of land use change.



Tianbao Huo received the B.E. degree in surveying and mapping engineering, in 2022, from the Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in mapping science and technology with the Faculty of Geomatics.

His research interests include land subsidence monitoring and time series InSAR prediction based on neural networks.

Jiapeng Tang received the B.E. degree in remote sensing science and technology, in 2021, from the Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in mapping science and technology with the Faculty of Geomatics.

His research interests include InSAR data processing technology and geological hazard identification of highway corridor in loess area.

Yumin Fang received the B.E. degree in remote sensing science and technology, in 2021, from the Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in mapping engineering with the Faculty of Geomatics.

His research interests include InSAR data processing technology, deep learning, and remote sensing image information extraction.

Yunhao Zhang received the B.E. degree in surveying and mapping engineering, in 2022, from the Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in resources and environmental engineering with the Faculty of Geomatics.

His research interests include landslide displacement mechanism analysis and trend prediction.