Landslide Susceptibility Mapping Considering Landslide Local-Global Features Based on CNN and Transformer

Zeyang Zhao, Tao Chen, *Senior Member, IEEE*, Jie Dou, Gang Liu, and Antonio Plaza, *Fellow, IEEE*

***Abstract*—Landslide susceptibility mapping (LSM) is a crucial step in quantitatively assessing landslide risk, essential for geologic hazards prevention. With the rapid development of deep learning models, convolutional neural networks (CNNs), and transformer architectures have been applied to LSM. However, these models still face the challenges of suboptimal mapping accuracy and limited capacity for multilevel landslide features extraction. In this study, we present a CNN-transformer local-global features extraction network (CTLGNet) that combines the strengths of both CNN and transformer models to effectively extract both landslide local and global features. We apply this model to LSM in two regions: the Three Gorges Reservoir area and Jiuzhaigou. To begin, nine landslide conditioning factors are selected and analyzed to construct the landslide dataset for LSM. Subsequently, the dataset is randomly split into training, validation, and test datasets in a 6:2:2 ratio to attain LSM results. Then, CTLGNet is compared to CNN, residual neural network, densely connected convolutional network, vision transformer, and fractional Fourier image transformer using various evaluation metrics. The results demonstrate that CTLGNet exhibits exceptional landslide prediction and generalization capabilities, outperforming the other five models across all evaluation metrics except Recall, with AUC values of 0.9817 and 0.9693 for the two regions, respectively. The LSM results indicate that CTLGNet can effectively extract both landslide local and global features to achieve landslide localization and detail capture. Overall, our proposed framework excels in extracting multilevel landslide features and holds great potential for widespread application.**

***Index Terms*—Convolutional neural network (CNN), landslide local-global features, landslide susceptibility mapping (LSM), transformer.**

I. Introduction

LANDSLIDES represent one of the most devastating natural phenomena worldwide, posing a grave threat to human life, property, and regional sustainable development [1], [2]. Landslide susceptibility refers to the spatial likelihood of landslides, which is determined by considering regional geological and environmental factors alongside historical landslide occurrences [3]. As a result, landslide susceptibility mapping (LSM) provides valuable data for government agencies involved in land planning and the development of effective disaster mitigation strategies [4].

With the rapid development of artificial intelligence (AI), numerous researchers have tested various machine learning (ML) algorithms in LSM tasks in the past two decades [5], [6], [7], [8]. ML algorithms possess the capability to automatically discern patterns from landslide data, allowing them to model complex, nonlinear relationships [9]. Nonetheless, ML algorithms also exhibit certain limitations, such as their reliance on feature engineering and the potential risks of overfitting and underfitting.

Deep learning (DL) algorithms, as a branch of ML algorithms, have demonstrated outstanding performance in image-related tasks [10], [11], [12]. By training in large-scale parallel computing environments, DL models have the capacity to automatically capture complex relationships between landslides and landslide conditioning factors (LCFs) and extract high-dimensional landslide features. This advantage has encouraged researchers to apply DL models in LSM studies with notable successes [13], [14], [15]. However, the two most popular DL models currently: convolutional neural network (CNN) and transformer, each have their own distinct advantages and limitations when it comes to landslide feature extraction, which impacts their accuracy in LSM and restricts the potential for widespread application [16].

In recent years, CNN and their variants, such as residual neural network (ResNet) [17] and densely connected convolutional

Manuscript received 20 December 2023; revised 21 February 2024; accepted 15 March 2024. Date of publication 19 March 2024; date of current version 4 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62371430 and Grant 62071439, in part by the Opening fund of State Key Laboratory of Geohazard Prevention and Geoenvironment Protection (Chengdu University of Technology) under Grant SKLGP2022K016, in part by the Open Fund of State Key Laboratory of Remote Sensing Science under Grant OFSLRSS202207, in part by the Open Fund of Badong National Observation and Research Station of Geohazards under Grant BNORSG-202302, in part by Opening Fund of the Key Laboratory of National Geographic Census and Monitoring, Ministry of Natural Resources under Grant 2023NGCM11, and in part by the Open Fund of Hebei Cangzhou Groundwater and Land Subsidence National Observation and Research Station under Grant CGLOS-2023-02. *(Corresponding author: Tao Chen.)*

Zeyang Zhao is with the School of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China (e-mail: zhaozy@cug.edu.cn).

Tao Chen is with the School of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China, also with the Badong National Observation and Research Station of Geohazards, China University of Geosciences, Wuhan 430074, China, also with the Key Laboratory of National Geographic Census and Monitoring, Ministry of Natural Resources, Wuhan 430079, China, and also with the State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China (e-mail: taochen@cug.edu.cn).

Jie Dou is with the Badong National Observation and Research Station of Geohazards, China University of Geosciences, Wuhan 430074, China (e-mail: doujie@cug.edu.cn).

Gang Liu is with the College of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China (e-mail: liuganggis@sina.com).

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10071 Cáceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/JSTARS.2024.3379350

network (DenseNet) [18], have achieved remarkable success in LSM tasks [19], [20]. Due to the spatial translation invariance and low inductive bias of convolutions, CNN-based models could extract landslide local features (LLFs) efficiently from multisource remote sensing images [21]. However, the limited receptive field of CNN-based models restricts their ability to process extensive contextual information, resulting in challenges in capturing landslide global features (LGFs) [22].

Transformer is an encoder-decoder sequence transformation model that utilizes the self-attention (SA) mechanism to expand the receptive field to capture global contextual information. One notable application is vision transformer (ViT), which successfully employs the pure transformer backbone for image classification tasks [23], leading to competitive performance in LSM tasks [24]. The SA mechanism calculates the correlations and weights between each pixel and all other pixels in the image, which allows the transformer to better extract LGFs compared to CNN [25]. However, transformer-based models have relatively weak LLFs extraction capabilities and typically require training with extensive large-scale data [26].

In LSM tasks, both LGFs and LLFs play crucial roles. The LGFs encompass key aspects, including the size, location, and spatial distribution of landslides. Conversely, the LLFs pertain to specific characteristics such as the edges, texture, and shape of the landslides. The strong LGFs and LLFs could enable the model to better distinguish pixels that exhibit subtle differences between landslides and nonlandslides, thereby improving the accuracy of the LSM. It is important to note that CNN and transformer have complementary advantages in extracting LGFs and LLFs [16]. Numerous researchers have introduced the combination of CNN and transformer models into various tasks, including scene classification [27], semantic segmentation [28], and change detection [29]. The findings indicate that the CNN-transformer hybrid framework can effectively extract both local and global features of an image, which improves the overall performance and reduces the transformer's dependence on large-scale training datasets [29]. However, there are no studies evaluated the effectiveness of hybrid CNN and transformer models in LSM tasks.

Taking into account the distinctive strengths of CNN in extracting LLFs and transformer in modeling LGFs, we introduce a CNN-transformer local-global features extraction network (CTLGNet) designed specifically for LSM tasks. By leveraging the combined advantages of CNN and transformer models, the network adeptly extracts both LGFs and LLFs to improve the accuracy of the LSM. The main contributions of this article are outlined as follows. We utilize the CNN structure to extract LLFs from the input data, and then use transformer structure to extract LGFs from the output features of the CNN. The CTLGNet can comprehensively extract LLFs and LGFs, improving the spatial prediction of landslides. We conduct a comparative analysis of the CTLGNet and individual models, evaluating their respective capabilities in extracting LLFs and LGFs from several evaluation perspectives. Furthermore, we compare the computational effort of the different models.

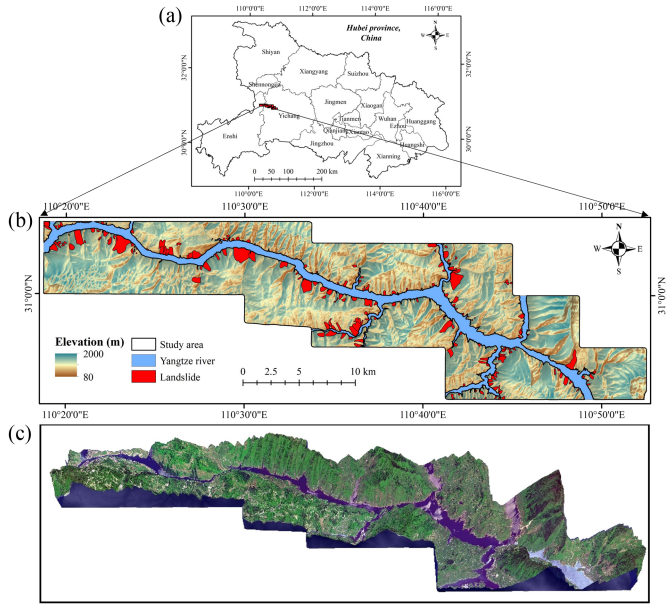

Fig. 1. Location of Site A. (a) Geographical location of Site A. (b) Distribution of landslides in Site A. (c) 3-D topography of Site A.

II. Study Area and Data

A. Study Area

In order to validate the universality and reliability of the proposed model, this study opted for two landslide-prone areas as study areas, namely the Three Gorges Reservoir area and Jiuzhaigou. These areas boast distinct geological structures and geographical features, providing ample evidence to showcase the model's performance and applicability.

1) Site A: Three Gorges Reservoir Area: Site A is located in the section from Zigui to Badong in the Three Gorges Reservoir area in western Hubei Province, China, including parts of Zigui County of Yichang City and Badong County of Enshi Prefecture [Fig. 1(a)]. The whole zone is located from 30°1'12" N to 30°55'48" N, 110°18'0" E to 110°52'12" E, with a total area of about 396 km^2. The Yangtze River flows through the site in the WNW-ESE direction.

The landslides distribution and three-dimensional (3-D) topography of Site A are shown in Fig. 1(b) and (c), respectively. Site A belongs to the subtropical monsoon climate, characterized by hot summers and cold winters, with four distinct seasons. Precipitation mainly concentrates in June to September, with an average annual rainfall of 1100 mm and prone to heavy rainfall. In terms of topography and geomorphology, Site A has a rugged surface which is very favorable to the formation and development of landslides. The faults in the area are widely developed and the lithology is unstable, so landslides have often occurred historically. As one of the most important water conservancy projects in China, it is important to assess the probability of landslide occurrence in Site A.

2) Site B: Jiuzhaigou: Site B is located in Jiuzhaigou County, Aba Tibetan and Qiang Autonomous Prefecture, Sichuan Province, China [Fig. 2(a)]. The whole zone is located from

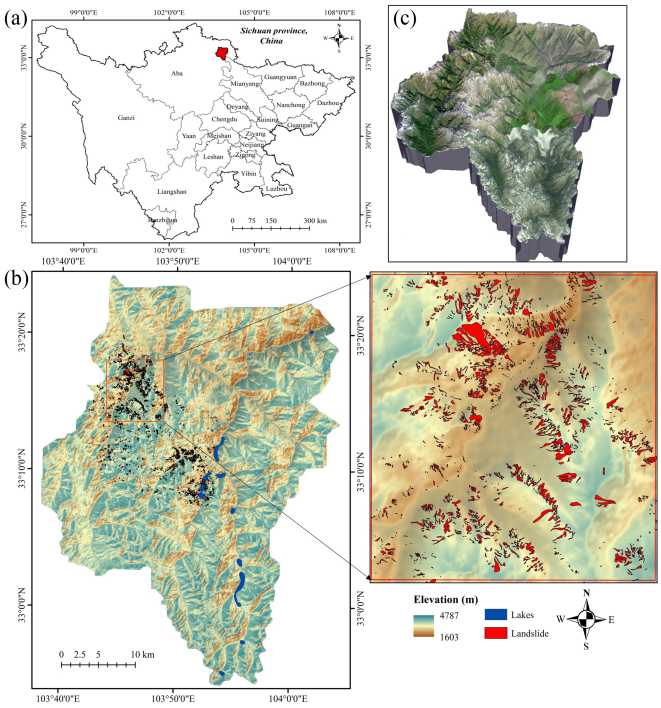

Fig. 2. Location of Site B. (a) Geographical location of Site B. (b) Distribution of landslides in Site B. (c) 3-D topography of Site B.

32°54'15" N to 33°24'19" N, 103°38'19" E to 104°04'20" E, with a total area of about 1367 km^2.

Fig. 2(b) and (c) depict the distribution of landslides and the 3-D topography of Site B, respectively. Site B belongs to the highland humid climate, which has less precipitation and is mostly concentrated in July and August. As the transition zone between the Tibetan Plateau and the Sichuan Basin, Site B is characterized by high peaks, deep valleys, and steep terrain. The topography of the area is high in the south and low in the north, with a relative elevation difference of more than 2000 m. Site B has complex geological background and strong tectonic movement, which has facilitated the development of geological hazards. On August 8, 2017, a 7.0 magnitude earthquake occurred in Jiuzhaigou County, Sichuan Province, resulting in multiple landslide disasters in Site B, posing a significant threat to local residents and tourists [Fig. 2(b)]. Therefore, accurate LSM is of great importance to people's safety in Site B.

B. Landslide Inventory Map

The drawing of landslide inventory map can visually display the distribution of landslides and create labels to input into the model for training. The landslide inventory map of Site A was mainly constructed by referring to the historical landslide data, combined with remote sensing images and on-site field survey data. And the landslide inventory data of Site B are mainly from landslide history documentation and visual interpretation of Google Earth images. Specifically, there are 202 landslides in Site A, with a total area of about 23.40 km^2, most of which are reservoir bank landslides, mainly along both sides of the mainstream of the Yangtze River. In Site B, there are more than 4000 landslides distributed, with a total area of about 9.51 km^2. The landslides are small in area and mainly concentrated in the west and central part of Site B.

TABLE I
DETAILED VALUES OF LANDSLIDE CONDITIONAL FACTORS

LCFs	Site A	Site B
Elevation(m)	[80, 2000]	[1603, 4787]
Aspect	Flat, North, Northeast, East, Southeast, South, Southwest, West, Northwest	
Slope(°)	[0, 78.42]	[0, 78.00]
Lithology	Soft rock, Soft and hard, Hard rock	Quaternary, Neogene, Triassic, Permian, Carboniferous, Devonian
DTF(m)	[0, 8753.58]	[0, 7728.60]
DTR(m)	[0, 4706.22]	[0, 6423.22]
Precipitation (mm)	[274.6, 320.9]	[104.0, 133.5]
LULC	Cropland, Forest, Shrub, Grassland, Water, Snow/Ice, Barren, Impervious, Wetland	
NDVI	[-1, 1]	[-1, 1]

C. Landslide Conditioning Factors

The basic assumption of landslide susceptibility is that landslides will occur in areas with similar geological and environmental conditions to historical landslides, so the selection of LCFs is critical for LSM. LCFs typically contain factors for topography and geomorphology, geology, hydrology, and ground cover [7]. Taking into account previous studies on the study area and the availability of data [30], we finally selected nine LCFs: elevation, aspect, slope, lithology, distance from fault (DTF), distance from river (DTR), precipitation, land use/land cover (LULC), and normalized difference vegetation index (NDVI).

The initial three LCFs were derived from the digital elevation model, sourced from http://www.gscloud.cn. Lithology and fault data were procured through vectorization from the 1:2 00 000 scale geological map database available at the National Geological Archive (http://www.ngac.org.cn/Map/List). River vector data were extracted from the 1:2 50 000 scale National Basic Geographic Database. Precipitation data for Sites A and B were sourced from the China Monthly Precipitation Dataset for July 2010 and August 2017, respectively, accessible for download from the National Earth System Science Data Center (http://www.geodata.cn). LULC data for 2010 and 2017 are obtained from the CLCD (China Land Cover Dataset) dataset, retrievable at http://zenodo.org/records/5816591. Furthermore, NDVI data were acquired using Landsat 5 TM (2010/07) and Landsat 8 OLI images (2017/02), respectively, which can be downloaded from https://earthexplorer.usgs.gov/. Table I shows detailed information of LCFs in Sites A and B, all of which were processed using ArcGIS 10.2.

III. METHODOLOGY

First, we identify nine geoenvironmental factors as LCFs and perform multicollinearity and importance analyses on the LCFs. Next, we generate landslide datasets by generating patch blocks from pixels. These datasets are then randomly divided into

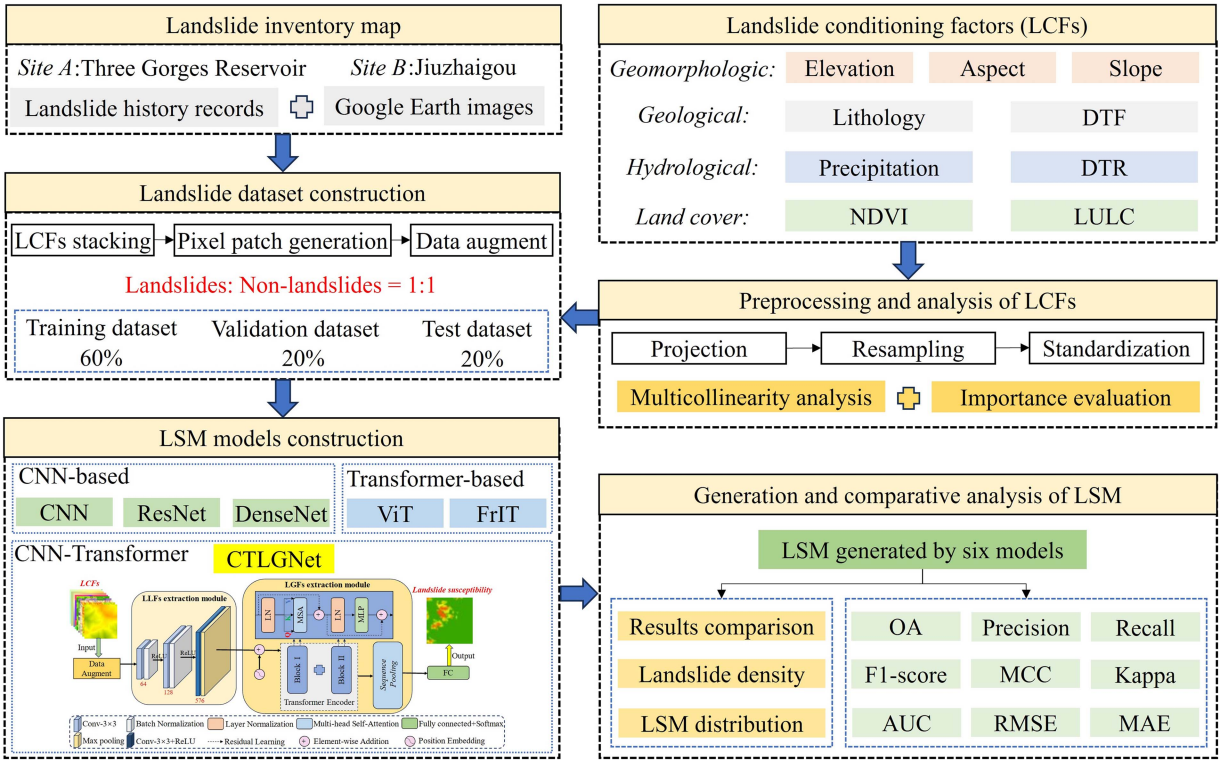

Fig. 3. Flow chart of this study.

training, validation, and test sets using a 6:2:2 ratio. Then, CNN, ResNet, DenseNet, ViT, fractional Fourier image transformer (FrIT), and CTLGNet are constructed to generate LSM. Lastly, we compare and evaluate model accuracy and generalization abilities. The detailed flowchart is shown in Fig. 3.

A. Data Preprocessing

The data preprocessing in this study involves several key steps, including coordinate system unification, resampling, and image standardization. Initially, we standardized the coordinate systems for all the data. Recognizing the variations in spatial resolution among different geospatial data types, we utilized bilinear interpolation for resampling to unify the spatial resolution [31]. This ensured consistency in the number of rows and columns in LCFs. To eliminate the difference in magnitude between discrete and continuous factors and accelerate model convergence, the Z-score method was employed to standardize the images of LCFs [32].

B. LCFs Selection

During the LSM process, it is critical to ensure the relative independence and importance of the selected LCFs. The presence of strong multicollinearity between LCFs or low importance of LCFs may adversely affect computational efficiency and lead to unstable and inaccurate models. Therefore, it is necessary to conduct multicollinearity analysis and importance evaluation on LCFs before constructing the landslide datasets [33].

1) Multicollinearity Analysis: In order to further evaluate the correlation among LCFs, we performed a multicollinearity analysis on the factors. Multicollinearity can be assessed using the variance inflation factor (VIF) and tolerance (TOL). The mathematical expression for VIF is as follows:

$$\mathrm{VIF} = \frac{1}{1 - R_i^2} \tag{1}$$

By utilizing $X = \{x_1, x_2, \ldots, x_n\}$, we establish an independent dataset for variables, where R_i^2 signifies the coefficient of determination for the ith independent variable, X_i, when it is regressed on all other predictor variables within the model. TOL represents the reciprocal of the VIF value in numerical terms. In general, if the VIF is less than 10, it suggests a low level of multicollinearity among the data. If the VIF falls between 10 and 100, it indicates the presence of strong multicollinearity. A VIF value of 100 or higher suggests severe multicollinearity.

2) Importance Evaluation: In order to measure the importance of each LCF for the occurrence of landslides, this study conducted an importance evaluation of nine LCFs in Sites A and B by calculating the Gini index in the random forest model [34]. The higher value of the Gini coefficient of an LCF indicates that the LCF is more important to the model, and the sum of the Gini index of all LCFs should be equal to 1.

C. Landslide Dataset Construction

After data preprocessing and LCFs selection, the LCFs are stacked into a 3-D matrix (H, W, and C), where H represents

height, W represents width, and C indicates the number of LCFs. In the entire region, a landslide pixel is randomly chosen as the center of a patch block. The block is then expanded in the H, W, and C directions, resulting in a landslide patch block with size of $72 \times 72 \times 9$. A total of 10 000 landslide patch blocks are generated for both Sites A and B, which constructs a landslide dataset for LSM. And an equal number of nonlandslide patch blocks are generated using the same steps for the training, validation, and test sets. Maintaining a consistent ratio of positive and negative samples, the landslide dataset is randomly divided into training, validation, and test sets in a 6:2:2 ratio. Finally, in order to further enrich sample diversity and enhance model performance, various techniques such as random horizontal flipping, random rotation, and local random scaling are applied to the training set during the data augmentation while keeping the patch block size unchanged, which were then input into all LSM models constructed in the following for training.

D. CNN-Based Models

1) CNN: The earliest CNN model, LeNet-5, was introduced in the 1990s for handwritten digit recognition [35]. CNN networks possess characteristics such as local perception, parameter sharing, and translation invariance, enabling them to effectively extract and utilize LLFs from input data [36]. As a result, CNNs have found wide-ranging applications in computer vision (CV) tasks and have achieved remarkable breakthroughs in the field of LSM [37], [38].

In this study, we constructed a ten-layer CNN network for LSM using 2-D convolutional kernels. The architecture comprises two convolutional layers with a kernel size of 3×3, two max pooling layers, two ReLU activation functions layers, a batch normalization (BN) layer, and a fully connected layer. The CNN structure could extract high-dimensional landslide features through convolution and pooling layers, obtaining relatively excellent LSM results.

2) ResNet: ResNet is designed to address the problems of network degradation and gradient vanishing encountered during the training of deep networks [17]. ResNet introduces the concept of residual learning, wherein the input feature maps of each residual module are added to the feature maps obtained after a series of convolution operations, creating a residual connection. This innovative design of skip connections enables the network to directly learn residuals without relying solely on gradual learning through multiple layers. ResNet architectures typically consist of multiple stacked residual modules [33]. Each residual module comprises two convolutional layers for feature extraction and a skip connection that adds the input features to the output features, facilitating the learning process.

In this article, we employed ResNet18 to build the LSM model, utilizing an architecture consisting of 18 layers within a deep CNN. ResNet18's residual blocks could facilitate the training of deeper networks and ensure that landslide features are propagated and retained more efficiently.

3) DenseNet: DenseNet is designed to address the challenges of gradient vanishing and feature reuse during the training of deep networks [18]. Unlike traditional CNNs, DenseNet introduces dense connections and skip connections, achieving superior network performance. DenseNet takes a different approach by establishing dense connections, where each layer is connected to all previous layers, forming a dense connection structure [34]. This architecture ensures that each layer has direct access to the feature maps of all preceding layers, enabling more comprehensive and rich information flow throughout the network. The network structure of DenseNet comprises multiple dense blocks, with each dense block consisting of multiple convolutional layers and a feature fusion layer known as the transition layer. The transition layer plays a vital role in controlling the dimensionality of feature maps and reducing computational complexity.

In this study, DenseNet-BC (DenseNet with bottleneck and compression) was used to obtain LSM results. This variant incorporates bottleneck and compression operations, effectively reducing the dimensionality of the landslide feature map and enhancing the overall performance of DenseNet.

E. Transformer-Based Model

1) ViT: Building upon the success of the transformer in NLP, Google introduced ViT in 2020, marking the extension of transformer's application to the field of CV for the first time [23]. Unlike traditional models, ViT takes a unique approach by treating images as 1-D sequences to extract features. To begin with, ViT divides the input image into a collection of patch blocks and subsequently flattens each block into a vector. These vectors serve as the input for the model. An embedding layer then maps these input vectors to a high-dimensional feature space, and these transformed vectors undergo processing through multiple layers of encoders. Within these encoders, the SA layer and the feed-forward neural network layer work in tandem to extract the image's global features. Finally, the output vector from the last position is taken as the image's feature representation, resulting in the classification outcome [39], [40].

We used the original VIT structure to extract landslide features and implement the LSM task. The configuration includes four attention heads and eight transformer layers. The VIT model proves effective in capturing LGFs across the input sequence while emphasizing key areas within the input image.

2) FrIT: In response to the lack of global contextual information in CNN and the loss of local semantic information in VIT, the FrIT was proposed. FrIT has demonstrated success in capturing both global and local contextual features [41]. Different from VIT which is based on SA mechanism, FrIT extracts global contexts based on 2-D fractional Fourier transform layers. Other than that, the structure of FrIT remaining parts is basically the same as that of VIT. We migrated the FrIT to the LSM task and conducted a comparative analysis against the network proposed in our study.

F. CTLGNet

To address the limitations of feature extraction capabilities in CNN- and transformer-based models and enhance the landslide spatial prediction accuracy by simultaneously extracting LGFs

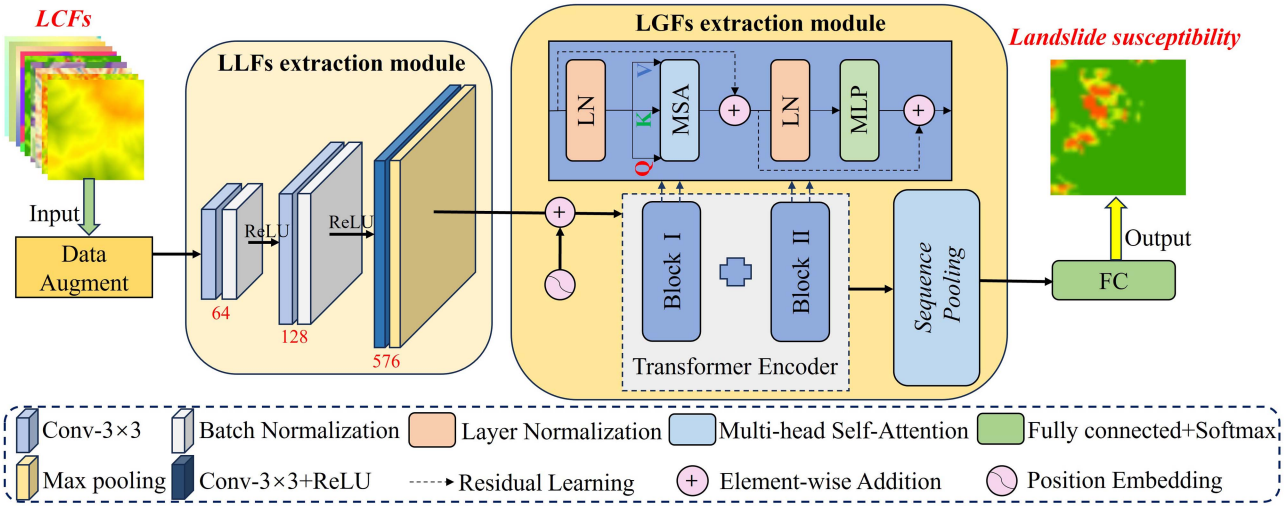

Fig. 4. Framework of LSM based on CTLGNet.

and LLFs, we propose a hybrid model called CTLGNet for LSM tasks. Fig. 4 illustrates the specific structure of CTLGNet, which comprises two key modules: the LLFs extraction module based on the convolutional architecture and the LGFs extraction module based on the improved transformer architecture.

1) LLFs Extraction Module: Given the strong local perception of convolution, we designed an LLFs extraction module that consists of three blocks to effectively extract LLFs. Each of the first two blocks is composed of convolution layers with a kernel size of 3 × 3, a BN layer, and a ReLU activation function. The first of these convolutional layers is used to capture the low-level LLFs, while the second one further extracts the more abstract LLFs. The BN and activation function layers play crucial roles in normalizing the data and introducing nonlinearities, respectively. The final block in the structure incorporates a convolutional layer with a 3 × 3 kernel size, a ReLU activation function, and a max pooling layer, facilitating the extraction of higher-level LLFs. The LLFs extraction module in our proposed model imparts sufficient inductive bias, enabling it to adeptly extract LLFs associated with landslide areas of varying sizes. Suppose we have an image input represented by $X \in R^{H\times W\times C}$

$$X_p = \text{MaxPool}\left(\text{Conv}\left(\text{Conv}\left(\text{Conv}\left(X\right)\right)\right)\right) \tag{2}$$

X will become the 1-D sequence $X_p \in R^{1\times(P^2\cdot C)}$, where p represents the resolution of each image patch block. This module empowers the model to effectively retain a broader spectrum of high-dimensional and critical LLFs.

2) LGFs Extraction Module: After patch and position embedding operations, the LGFs extraction module extracts the LGFs through improved transformer. First, each transformer encoder consists of traditional components, including multihead SA (MSA), layer normalization (LN), multilayer perceptron (MLP), and skip connections.

The MLP comprises a fully connected layer with GELU activation function. When the encoder input is X_L

$$X'_{L+1} = \text{MLP}\left(\text{LN}\left(X_L\right)\right) + X_L \tag{3}$$

$$X_{L+1} = \text{MSA}\left(\text{LN}\left(X'_{L+1}\right)\right) + X'_{L+1} \tag{4}$$

where L represents the number of transformer encoder. In models like ViT and BERT, the class token plays a crucial role in image classification [23], [42] as a specialized embedding vector. In CTLGNet, we introduce a replacement for class tokens called sequence pooling (SeqPool), ensuring that the output sequence encompasses relevant information from various regions of the input image [43]. By preserving this information, the model becomes more adept at leveraging spatially sparse data, ultimately resulting in enhanced LSM performance. SeqPool serves as an attention-based approach that enables the network to appropriately balance the sequential embeddings within the potential space generated by the transformer encoder. If we get the output $X_e \in R^{B\times N\times D}$ of the last transformer encoder, seqpool implements $R^{B\times N\times D} \mapsto R^{B\times D}$ through a series of transformations. The details are as follows:

$$x'_L = \text{softmax}\left(\text{Linear}(x_L)^T\right) \tag{5}$$

$$z = x'_L\, x_L = \text{softmax}\left(\text{Linear}(x_L)^T\right) \times x_L. \tag{6}$$

In the equation, X_e represents the output of the last encoder, B is the batch size, N is the sequence length, D represents the total embedding dimension, and z is the output.

The model's output layer features a fully connected layer, incorporating a Softmax activation function. This fully connected layer is responsible for generating landslide susceptibility values, providing a probabilistic of landslide occurrence. The proposed CTLGNet combines the strengths of CNN and transformer architectures, offering an effective solution for extracting both LLFs and LGFs in LSM tasks.

G. Model Evaluation Metrics

Evaluation metrics play a crucial role in quantitatively validating and assessing the predictive capability of a model. This study incorporates a range of model evaluation metrics to thoroughly

evaluate the performance of the model. These metrics include overall accuracy (OA), precision rate (Precision), recall rate (Recall), F1-score, Matthews correlation coefficient (MCC), Kappa coefficient (Kappa), root mean square error (RMSE), mean absolute error (MAE), area under the curve (AUC), and receiver operating characteristic (ROC). The specific formulas for calculating these metrics are as follows:

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

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

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

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

$$\text{MCC} = \frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \tag{11}$$

$$\text{Kappa} = \frac{p_0-p_e}{1-p_e} \tag{12}$$

$$p_e = \frac{(\text{TP}+\text{FN})(\text{TP}+\text{FP})+(\text{FP}+\text{TN})(\text{FN}+\text{TN})}{n^2}. \tag{13}$$

Among these, true positive (TP) refers to the number of pixels correctly classified as landslides, true negative (TN) represents the number of pixels correctly classified as nonlandslides, false positive (FP) denotes the number of pixels incorrectly classified as landslides, and false negative (FN) indicates the number of pixels incorrectly classified as nonlandslides. The closer the value of the above-mentioned metrics is to 1, the better the model performance is. RMSE and MAE are employed to quantify the deviation between the predicted landslide susceptibility values by the model and the actual values, thereby assessing the accuracy and reliability of the model. The calculation formulas for both are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(r_i-p_i)^2} \tag{14}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|r_i-p_i|. \tag{15}$$

In the formula, n represents the total number of pixels in the test set, while r_i and p_i, respectively, denote the real data and predicted data. Additionally, the smaller the values of RMSE and MAE, the more accurate the model is. In addition, the ROC curve, which illustrates the trade-off between sensitivity and specificity, is a crucial metric for assessing the reliability of classification models. The closer the ROC curve is to the upper-left corner, the more dependable the results are considered to be.

Landslide density (LD) serves as a quantitative measure of landslide distribution within the LSM results. The classification of landslide susceptibility results encompasses five distinct categories: very low (VL), low (L), moderate (M), high (H), and very high (VH). LD is defined as the ratio of the proportion of landslide area to the proportion of zone area within the susceptibility zones [19], [44]. The formula for calculating the LD value is as follows:

$$\text{LD} = \text{PL}/\text{PC}. \tag{16}$$

In the given equation, PC represents the proportion of each susceptibility zone's area relative to the entire site, while PL represents the proportion of landslide area within each susceptibility zone relative to the total landslide area of the entire site. As the landslide susceptibility class increases, the LD value is expected to increase from VL to VH. This relationship indicates a higher concentration of landslides in zones with higher susceptibility class [45].

TABLE II
COMPUTER HARDWARE INFORMATION

Item	Detailed parameter
CPU	Intel Xeon(R) Silver 4210R CPU@2.4GHZ×40
GPU	GeForce RTX 3090/PCIe/SSE2
Memory	128GB
Hard disk	8TB

H. Experimental Environment and Hyperparameter Settings

Table II provides an overview of the computer hardware utilized in the experiments conducted for this study. The software, developed within the Python 3.8 environment, was built on the PyCharm 2021 and Microsoft Visual Studio 2021 software platforms. The experiments were carried out on the Linux-Ubuntu 18.04.05 LTS operating system. The DL network was crafted using the Keras 2.4.3 AI development framework.

The model accuracy and computational efficiency of the model depend not only on the structural design of the model, but also on the hyperparameter settings in the model [41]. We used trial-and-error method to optimize all model parameters by limiting the range of hyperparameters, and finally selected the optimal combination of parameters according to the model accuracy. The stochastic gradient descent optimizer was selected for the network, initializing the learning rate at 0.001, and employing a batch size of 16 for each learning iteration. The final output layer of all models was set as the fully connected layer with the activation function "Softmax," which maps the output of multiple neurons in the upper layer to a value of 0-1, i.e., the probability of landslide occurrence. To complement the Softmax activation, we chose the cross-entropy loss function with label smoothing as the loss function for all models. The training process involved 100 epochs, and the accuracy on the validation set was continuously monitored during training to identify and select the best-performing model.

IV. RESULTS AND ANALYSIS

A. Analysis of the LCFs

The results of the multicollinearity analysis of LCFs in Sites A and B are presented in Fig. 5. It can be seen that the LCFs in Sites A and B have low multicollinearity. In Site A, all LCFs have VIF values below 2, indicating a low level of multicollinearity among

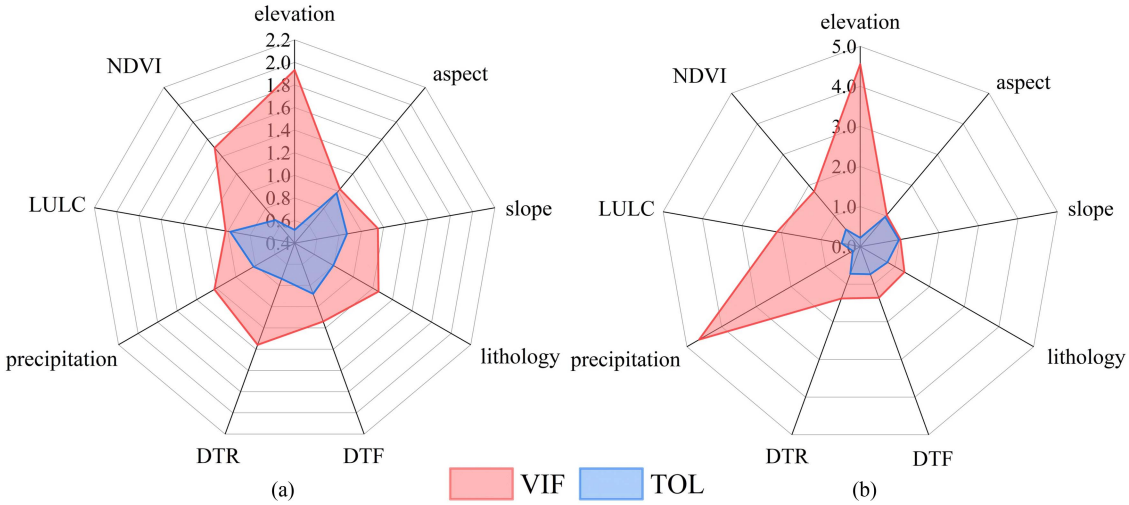

Fig. 5. Results of multicollinearity analysis of LCFs. (a) Site A. (b) Site B.

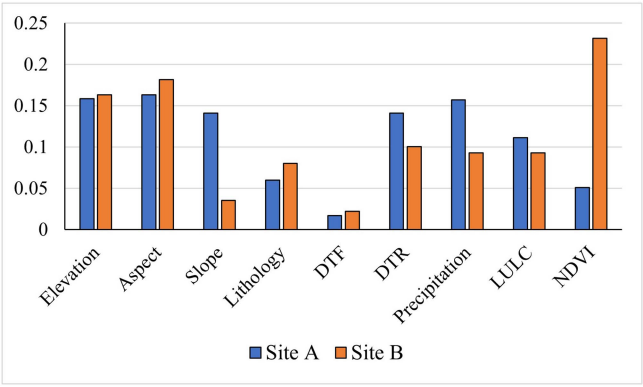

Fig. 6. Results of importance evaluation of LCFs. (a) Site A. (b) Site B.

the data. The highest VIF value is associated with elevation, which is 1.933, corresponding to a TOL value of 0.517. While in Site B, all LCFs have VIF values below 5, with elevation and precipitation having VIF values of 4.560 and 4.649, respectively, and corresponding TOL values of 0.219 and 0.215.

Fig. 6 shows the results of importance evaluation of LCFs. It is evident that all LCFs play a role in landslide occurrence, albeit to varying degrees across different regions. Notably, in both Sites A and B, DTF exhibits relatively low importance. NDVI demonstrates lower importance in Site A but holds significant relevance in Site B. Conversely, the slope is highly significant in Site A but exhibits less significance in Site B. Given the absence of strong multicollinearity among the selected LCFs and their collective importance in influencing landslide occurrences, we will utilize all LCFs to obtain LSM results.

B. Comparison of LSM Results

Using the six network architectures constructed in the last section, we obtained LSM results for Sites A and B. The Jenks natural breakpoint method was employed to classify the landslide susceptibility into five categories: VL, L, M, H, and VH [46], [47]. Figs. 7 and 8 illustrate that the LSM results obtained from each model in the same area exhibit similarities. Specifically, within Sites A and B, the majority of the area is classified as VL and L susceptibility zones, while the distribution of H and VH susceptibility zones aligns closely with the spatial locations and outlines of historical landslides. Notably, the H and VH susceptibility zones predicted by the CTLGNet model exhibit a superior alignment with historical landslides in both global location and local details. In Site A, VL and L susceptibility zones are mainly concentrated along the main stem and tributaries of the Yangtze River. In Site B, they are concentrated in the western and central parts of the region. Overall, all six models demonstrate a certain level of spatial probability prediction capability for landslides and point to areas of nonhistoric landslides with high susceptibility values that deserve attention.

The PC, PL, and LD values obtained by the six models in Site A are shown in Table III. In this study, we define areas with L and VL susceptibility as low landslide-prone areas, and areas with VH, H, and M susceptibility as landslide-prone areas [48]. In terms of PL values, CTLGNet successfully identified 97.93% of historical landslides located in landslide-prone zones. Remarkably, these landslides were confined to only 12.65% of the total area classified as landslide-prone zones. Further analysis revealed that within the VH susceptibility zones, CTLGNet accurately matched 90.73% of the historical landslide records, despite these zones accounting for just 8.78% of the total area designated as H susceptibility zones. Additionally, the PC, PL, and LD values obtained by the six models in Site B are displayed in Table III. CTLGNet predicted that 87.21% of the area is not prone to landslides, which is marginally lower than ViT's prediction of 87.95%. Surprisingly, the overall percentage of landslide-prone area was only 12.79%. However, the CTLGNet successfully predicted 97.71% of the historical landslides, with 88.95% of the region's VH susceptibility area accurately identified despite occupying only 6.27% of the total area.

Based on the definitions of LD, both values can be used to quantitatively express the distribution of landslides in the LSM results by combining the PC and PL values. From the

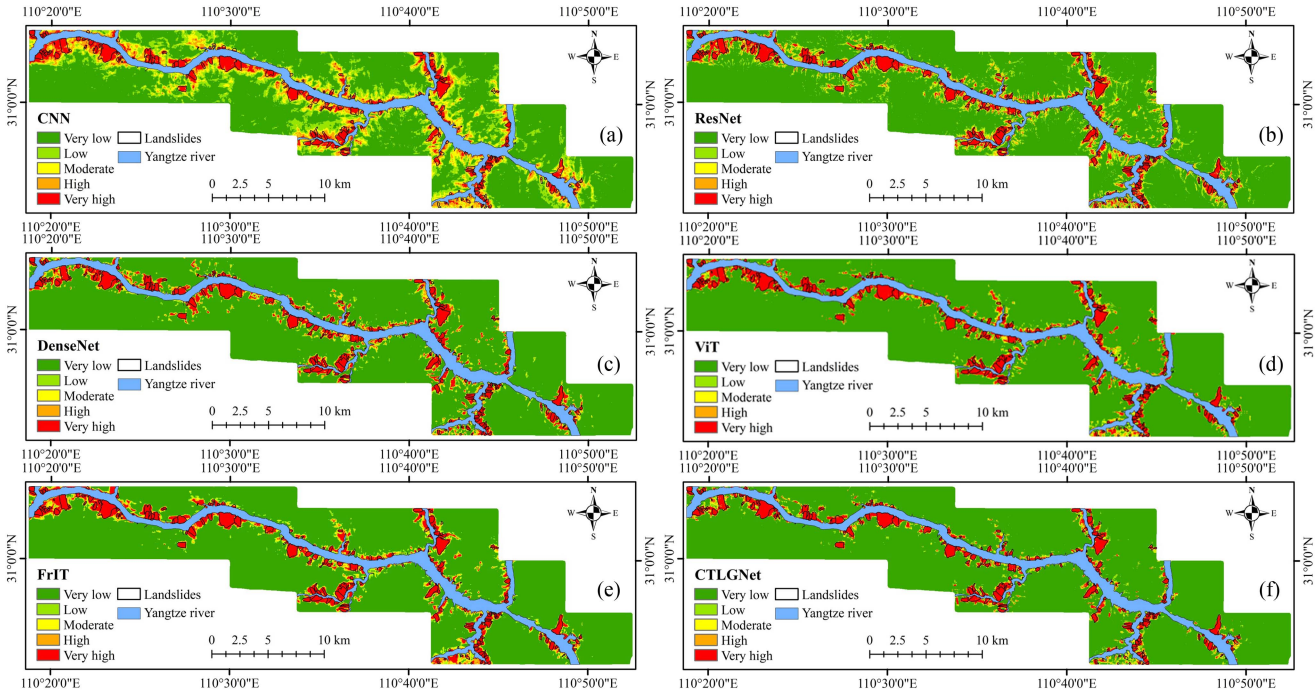

Fig. 7. Comparison of LSM results obtained from six models in Site A. (a) CNN. (b) ResNet. (c) DenseNet. (d) ViT. (e) FrIT. (f) CTLGNet.

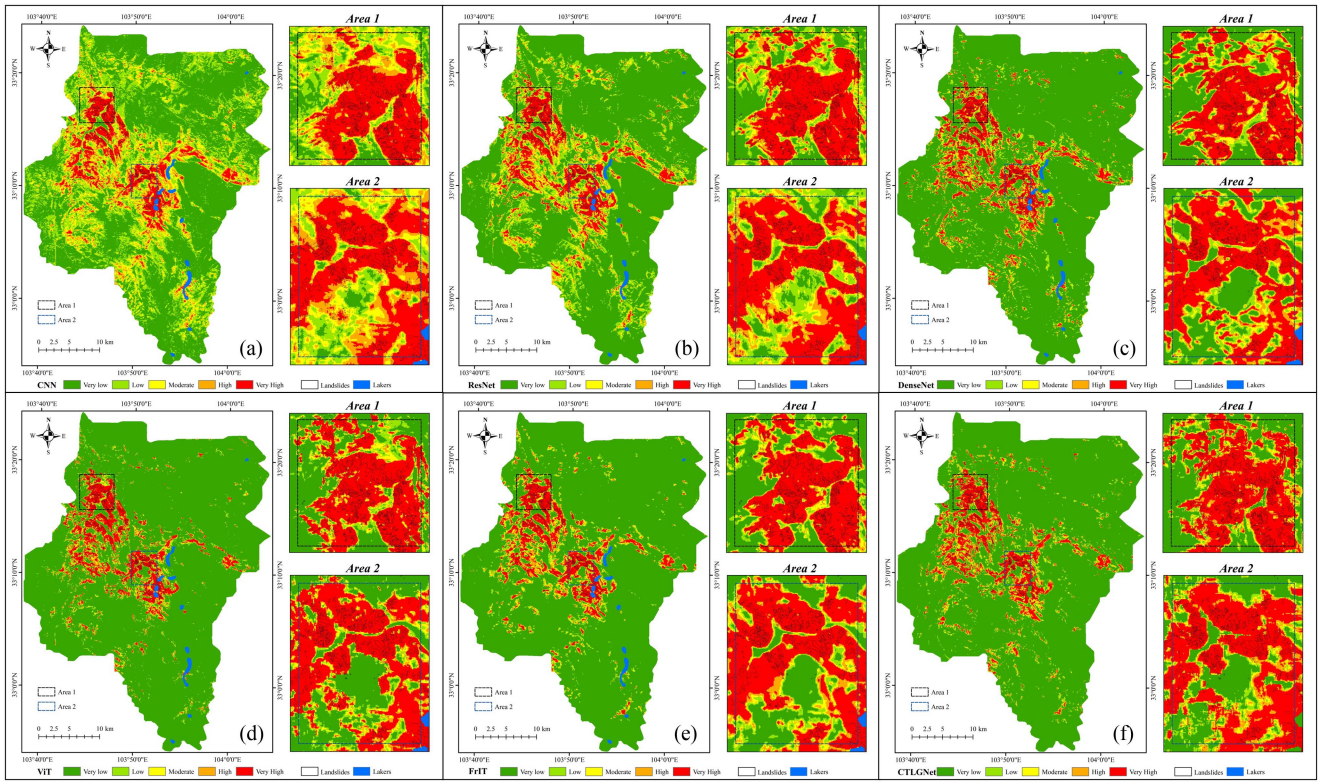

Fig. 8. Comparison of LSM results obtained from six models in Site B. (a) CNN. (b) ResNet. (c) DenseNet. (d) ViT. (e) FrIT. (f) CTLGNet.

TABLE III
LD RESULTS OBTAINED FROM EACH MODEL IN SITES A AND B

Models	LSM partition	PC_A	PL_A	LD_A	PC_B	PL_B	LD_B
CNN	Very low	0.6102	0.0018	0.0029	0.5247	0.0014	0.0027
	Low	0.1457	0.0080	0.0551	0.2304	0.0104	0.0453
	Moderate	0.0847	0.0305	0.3605	0.1048	0.0360	0.3439
	High	0.0767	0.1811	2.3618	0.0676	0.1104	1.6319
	Very high	0.0828	0.7785	9.4053	0.0725	0.8418	11.6061
ResNet	Very low	0.7542	0.0025	0.0033	0.6748	0.0069	0.0103
	Low	0.0724	0.0089	0.1223	0.1289	0.0125	0.0970
	Moderate	0.0389	0.0172	0.4426	0.0580	0.0222	0.3822
	High	0.0383	0.0638	1.6648	0.0531	0.0632	1.1916
	Very high	0.0961	0.9076	9.4459	0.0852	0.8952	10.5043
DenseNet	Very low	0.8048	0.0048	0.0060	0.7913	0.0133	0.0168
	Low	0.0349	0.0079	0.2253	0.0583	0.0175	0.2993
	Moderate	0.0251	0.0154	0.6159	0.0314	0.0193	0.6168
	High	0.0323	0.0561	1.7335	0.0342	0.0623	1.8205
	Very high	0.1029	0.9158	8.9016	0.0848	0.8876	10.4618
ViT	Very low	0.8256	0.0089	0.0108	0.8383	0.0151	0.0180
	Low	0.0273	0.0118	0.4323	0.0400	0.0130	0.3243
	Moderate	0.0192	0.0164	0.8566	0.0232	0.0192	0.8286
	High	0.0249	0.0487	1.9594	0.0228	0.0421	1.8485
	Very high	0.1031	0.9142	8.8692	0.0756	0.9106	12.0394
FrIT	Very low	0.8105	0.0080	0.0100	0.8154	0.0114	0.0141
	Low	0.0312	0.0068	0.2185	0.0448	0.0110	0.2476
	Moderate	0.0229	0.0129	0.5641	0.0283	0.0148	0.5245
	High	0.0284	0.0447	1.5718	0.0329	0.0552	1.6765
	Very high	0.1068	0.9273	8.6812	0.0784	0.9072	11.5706
CTLGNet	Very low	0.8479	0.0129	0.0152	0.7771	0.0133	0.0171
	Low	0.0256	0.0078	0.3041	0.0950	0.0096	0.1009
	Moderate	0.0175	0.0172	0.9856	0.0334	0.0189	0.5646
	High	0.0212	0.0548	2.5781	0.0317	0.0688	2.1685
	Very high	0.0878	0.9073	10.3298	0.0627	0.8895	14.1824

TABLE IV
MODEL EVALUATION METRICS OF EACH MODEL IN SITES A AND B

Site	Models	OA	Precision	Recall	F1-score	MCC	Kappa	AUC	RMSE	MAE
Site A	CNN	92.500%	88.956%	96.743%	92.686%	0.853	0.850	0.9732	0.274	0.075
	ResNet	93.800%	90.362%	97.812%	93.939%	0.879	0.876	0.9783	0.249	0.062
	DenseNet	92.625%	91.079%	**98.727%**	94.749%	0.896	0.893	0.9758	0.232	0.054
	ViT	94.050%	91.614%	96.743%	94.108%	0.882	0.881	0.9791	0.244	0.060
	FrIT	94.700%	91.481%	98.371%	94.801%	0.894	0.896	0.9787	0.230	0.053
	CTLGNet	**95.450%**	**93.552%**	97.456%	**95.464%**	**0.910**	**0.909**	**0.9817**	**0.213**	**0.045**
Site B	CNN	90.550%	88.213%	93.597%	90.825%	0.812	0.811	0.9541	0.307	0.095
	ResNet	90.075%	86.376%	95.147%	90.549%	0.806	0.802	0.9567	0.315	0.099
	DenseNet	90.850%	88.387%	94.047%	91.129%	0.819	0.817	0.9501	0.303	0.092
	ViT	92.050%	90.196%	94.347%	92.225%	0.842	0.841	0.9605	0.282	0.080
	FrIT	92.225%	88.858%	96.548%	92.543%	0.844	0.847	0.9635	0.278	0.078
	CTLGNet	**93.475%**	**91.136%**	**96.623%**	**93.799%**	**0.871**	**0.869**	**0.9693**	**0.255**	**0.065**

distribution of LD, it can be observed that the LD values of all models increase with the rise of susceptibility levels, and the LD value in the VH susceptibility area is significantly higher than in other areas. In conclusion, CTLGNet demonstrated the highest LD values in landslide-prone areas and exhibited the greatest disparity in LD values among different susceptibility classes. These findings indicate that CTLGNet possesses stronger predictive power compared to both CNN-based and transformer-based models in accurately predicting landslide-prone areas.

C. Comparative Evaluation of Model Performance

In this study, the dataset was partitioned into three subsets: training, validation, and test sets. The performance and predictive ability of six models were thoroughly evaluated using the test set. The ROC curves and corresponding AUC values of these six models were generated and presented in Fig. 9 for both Sites A and B. In terms of AUC values, CTLGNet demonstrated exceptional performance among all the models, surpassing the others in terms of AUC values for both Site A (AUC = 0.9817) and Site B (AUC = 0.9693). These outstanding results indicate that CTLGNet exhibits superior predictive capabilities, making it the most promising model for LSM task.

Table IV presents a comprehensive comparison of model evaluation metrics for the six models in both Sites A and B, where bold indicates the best. Notably, the proposed CTLGNet in this study exhibited the highest accuracy across most evaluation metrics, underscoring its effectiveness. In Site A, CTLGNet achieved an impressive OA of 95.450%, Precision of 93.552%,

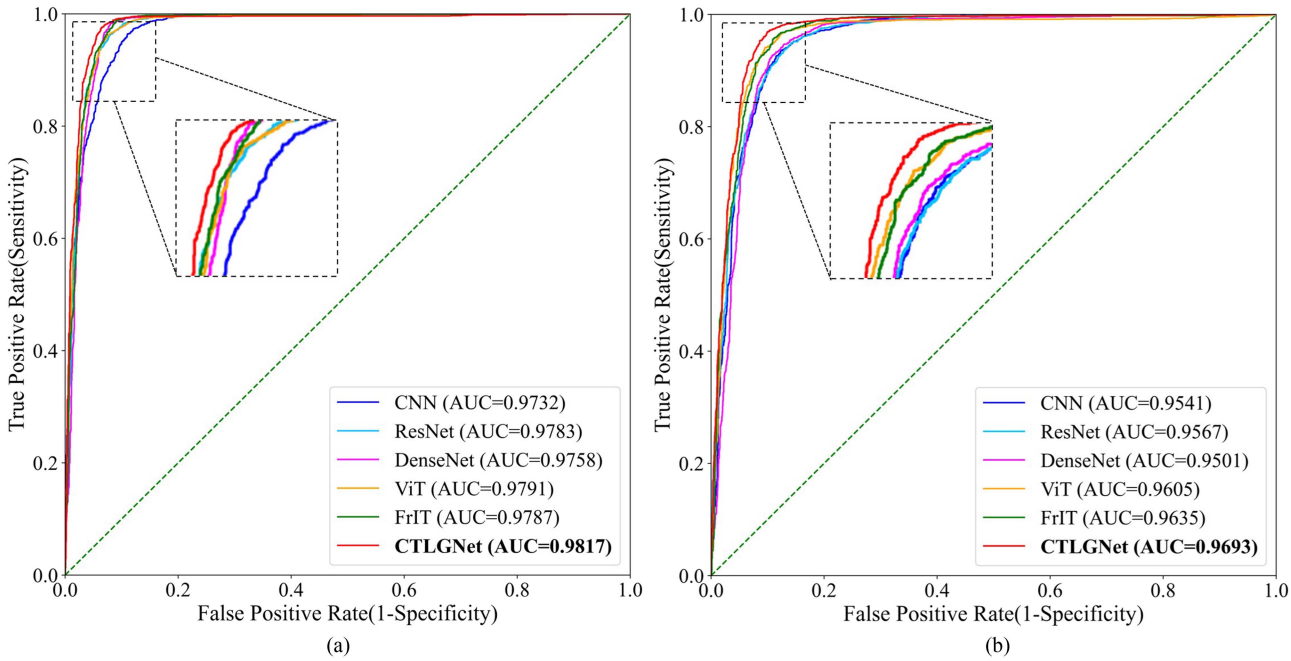

Fig. 9. ROC curves and AUC values on the test set for the six models. (a) Site A, (b) Site B.

F1-score of 95.464%, MCC of 0.910, Kappa of 0.909, RMSE of 0.213, and MAE of 0.045. It only had a slightly lower Recall (97.456%) compared with ResNet, DenseNet and FrIT. Moving on to Site B, CTLGNet maintained its superiority with an OA of 93.475%, Precision of 91.136%, Recall of 96.623%, F1-score of 93.799%, MCC of 0.871, Kappa of 0.869, RMSE of 0.255, and MAE of 0.065. These exceptional performance results further validate the efficacy of CTLGNet as the top-performing model in this study, showcasing its potential for accurate and reliable LSM in both Sites A and B.

V. Discussion

A. Comparison of the Ability of the Model to Extract LGFs and LLFs

If the model accurately extracts the spatial probability of landslide occurrence, the resulting susceptibility values should correspond to historical landslides, with susceptibility values within landslide areas approaching 1. In order to test the ability of the models in landslide localization and detail capture (i.e., LGFs and LLFs extraction), we extracted the susceptibility values for historical landslide areas predicted by six models and annotated the distribution and median line of susceptibility values (as depicted in Fig. 10). In the violin plot, wider violins indicate a denser distribution of pixels within that range, while longer protrusions suggest the presence of more outliers in the landslide susceptibility results. The region between 25% and 75% represents the middle 50% range of the dataset, representing the median concentration trend of the data.

In Site A, it is evident that CTLGNet outperforms all other models by achieving the highest mean susceptibility value among the historical landslide areas, measuring an impressive 0.876. Additionally, it exhibits exceptional performance in terms of having the lowest standard deviation (SD) and mean absolute deviation (MAD) compared with the other models. The median line and the 25%–75% percentile range of CTLGNet also surpass those of the other five models, further indicating its superior predictive capability. Moving on to Site B, CTLGNet still delivers a competitive performance with an average susceptibility value of 0.842, slightly lower than DenseNet's score of 0.857. Nevertheless, it maintains its reputation for reliability with the lowest SD and the lowest MAD in this context. When considering the combination of mean, SD, and MAD values of the susceptibility measurements, as well as the clustering of the distribution of these values, it becomes evident that CTLGNet consistently outperforms the other models.

In order to further demonstrate the advantages of CTLGNet's in extracting LGFs and LLFs, we choose Site A as a representative region and compare the spatial details of the LSM results obtained from the six models in historical landslide areas. The comparison results are shown in Fig. 11. Benefiting from LGFs and LLFs extracted by the model, CTLGNet can be very good for landslide localization and detail capture, and more accurately predict the location of landslide occurrence. The H and VH susceptibility zones obtained by the CTLGNet model can be better matched to historical landslides in terms of global location and local details. The other five models are more or less deficient in LGFs or LLFs extraction, which leads to the complete division of some historical landslide areas into VL, L, and M susceptibility areas.

B. Comparison of Model Efficiency and Computation

In order to further validate the performance of the proposed model, this study conducts a comparison focusing on model efficiency and computation. Given the inherent advantage of

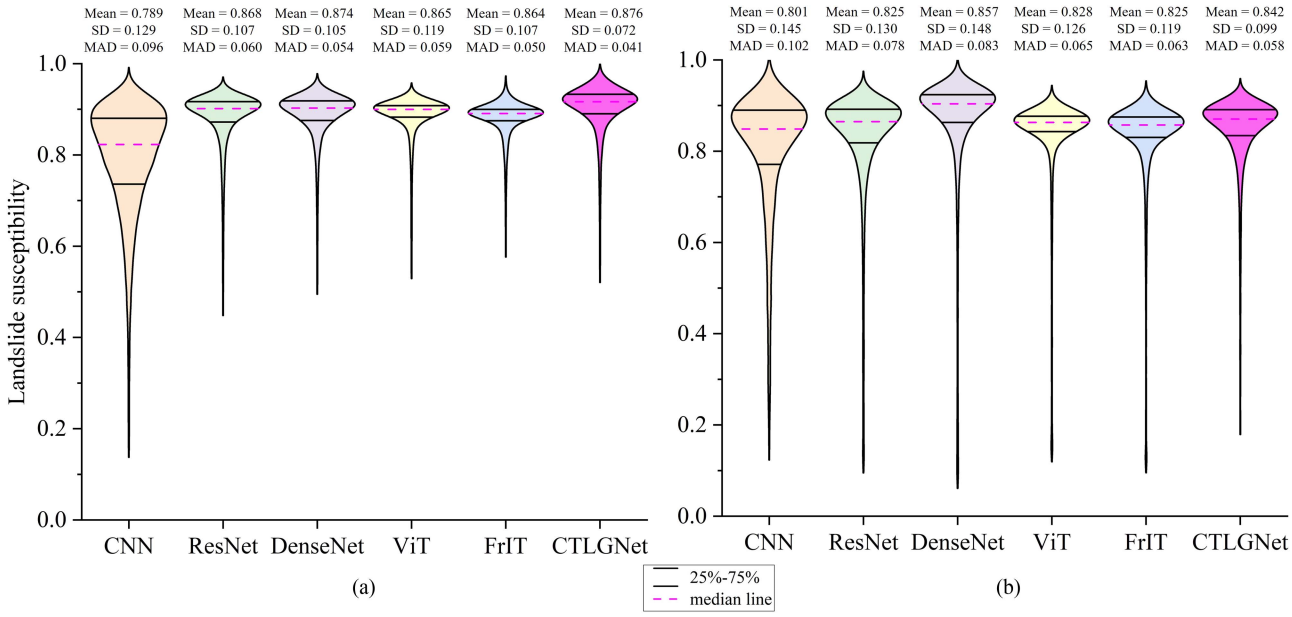

Fig. 10. Distributional relationships between landslide and landslide susceptibility. (a) Site A. (b) Site B.

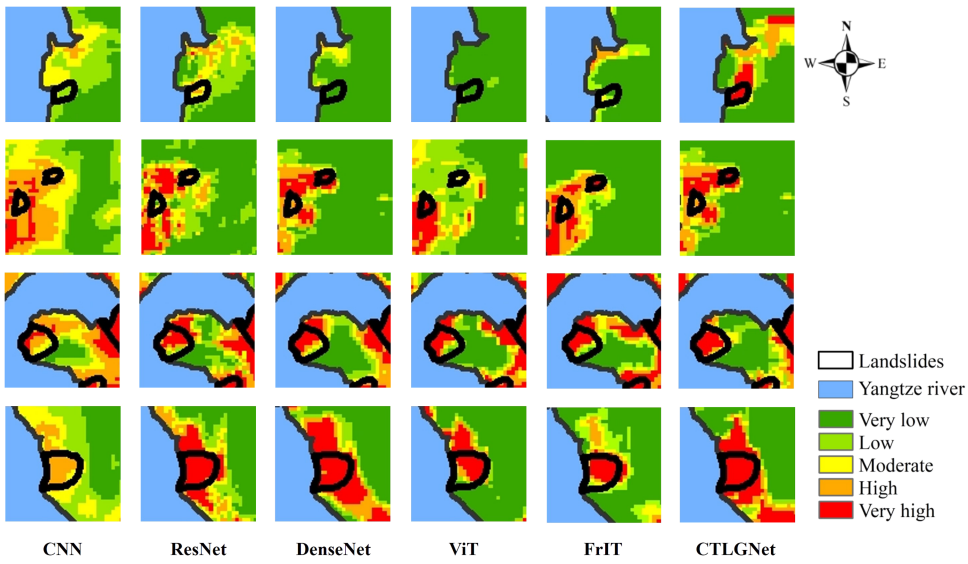

Fig. 11. Detailed comparison of LSM results of the six models in Site A.

convolution-based models in this aspect, our analysis is limited to a comparison with transformer-based models. Specifically, we computed the number of model parameters and the floating-point operations (FLOPs) for CTLGNet and compared them against two transformer-based models utilized in this study (ViT and FriT), as well as an efficient transformer-based model, MobileViT [49]. FLOPs can be interpreted as the amount of computation of the model, which is commonly used to measure the complexity of the model. MobileViT integrates the strengths of both CNN and ViT, emerging as a lightweight and general model that has demonstrated superior performance over CNN and ViT-based networks across various tasks and datasets. The number of parameters and FLOPs for each model are summarized in Table V.

TABLE V
COMPARISON OF MODEL PARAMETERS AND FLOPS

Models	Parameters	FLOPs
ViT	11 207 426	11 193 973
FrIT	10 676 482	10 669 653
MobileViT	1 307 621	72 661 289
CTLGNet	950 979	1 894 933

The analysis reveals that CTLGNet exhibits the lowest number of parameters and FLOPs among the compared models, highlighting its computational efficiency. Conversely, both ViT and FrIT exhibit higher parameter counts and FLOPs, consistent with their structural characteristics. While MobileViT has only

TABLE VI
AUC VALUES FOR THE FIVE-FOLD CROSS-VALIDATION OF THE CTLGNET MODEL

	①	②	③	④	⑤	Mean
Site A	0.9817	0.9812	0.9811	0.9815	0.9815	0.9814
Site B	0.9693	0.9689	0.9690	0.9688	0.9692	0.9690

a higher number of parameters than CTLGNet but the highest FLOPs, a trend consistent with the findings reported in the original article.

C. Robustness Analysis of the CTLGNet

To further test the robustness of the proposed model, we performed a five-fold cross-validation of the CTLGNet. First, the constructed landslide dataset was randomly and evenly divided into five parts. Then, based on the training, validation, and test dataset ratio of 3:1:1 set in this study, one of the parts is taken as the test dataset each time without repetition, and the CTLGNet is trained with the other four parts as the training and validation datasets. Finally, the AUC value of the model on the test set is calculated, providing a comprehensive measure of the accuracy of LSM results.

The AUC values obtained from the five-fold cross-validation of CTLGNet, as presented in Table VI, exhibit minor fluctuations within narrow ranges. Specifically, the AUC values range from 0.9811 to 0.9817 for Site A and from 0.9688 to 0.9693 for Site B. Calculating the mean AUC values across the five experiments yields 0.9814 for Site A and 0.9690 for Site B. These values outperform those of the comparison models, affirming the robustness and stability of CTLGNet.

D. Limitations and Future Work

First and foremost, the accuracy and credibility of LSM heavily depend on the quality of data sources. Given the rapid progress in remote sensing imaging sensors, multimodal remote sensing data can offer complementary information [50]. Effectively enhancing data quality and establishing a high-quality landslide dataset by integrating multimodal data present a significant challenge.

Furthermore, the training samples are the basis of constructing the LSM model. While this study applied fundamental data augmentation techniques, such as flipping and rotation, to augment the multilevel features of landslides, more advanced data augmentation methods have already demonstrated success in image classification tasks [51], [52]. Investigating whether these sophisticated data augmentation strategies are effective in highlighting landslides features to improve the accuracy of LSM models is a worthy goal.

Last but not least, the performance of LSM is notably influenced by the chosen model architecture [48]. Model construction using different DL models may result in variations in the extraction and effectiveness of landslide features. While this research employed a sequential structure to integrate the CNN and transformer, investigating the potential of utilizing parallel or hierarchical structures to build LSM models represents a direction for our future endeavors.

VI. CONCLUSION

In this article, we propose CTLGNet that takes into account both LLFs and LGFs for LSM. The model is applied in two regions: Three Gorges Reservoir area and Jiuzhaigou. Historical landslide inventory and nine LCFs were selected to construct a landslide dataset for LSM. Subsequently, we employ comprehensive evaluation metrics to evaluate the LSM results obtained by our proposed model and compared with five other models: CNN, ResNet, DenseNet, ViT, and FrIT.

The experimental results demonstrate that CTLGNet utilized in this study achieve relatively accurate LSM results and the regions identified as VH and H susceptibility classes most closely match the spatial distribution of historical landslides. Statistical results show that CTLGNet outperforms the other five models in all evaluation metrics except Recall, with AUC values of 0.9817 and 0.9693 for the two regions, respectively. Furthermore, the extraction of landslide susceptibility values within historical landslide areas demonstrates that CTLGNet produces results with the highest mean value and the lowest MAD and SD. This indicates that the CTLGNet captured the best localization and details of landslides and extracted LLFs and LGFs most adequately. Besides that, CTLGNet has the lowest number of parameters and FLOPs among the models related to transformer, which gives it an advantage in terms of model computations. In conclusion, our proposed method demonstrates outstanding capabilities in predicting landslides and exhibits excellent generalization ability. As a result, it holds promising potential for a wide range of LSM applications.

REFERENCES

[1] M. J. Froude and D. N. Petley, "Global fatal landslide occurrence from 2004 to 2016," *Natural Hazards Earth Syst. Sci.*, vol. 18, no. 8, pp. 2161–2181, Aug. 2018.
[2] S. L. Gariano and F. Guzzetti, "Landslides in a changing climate," *Earth Sci. Rev.*, vol. 162, pp. 227–252, Nov. 2016.
[3] F. Guzzetti, P. Reichenbach, M. Cardinali, M. Galli, and F. Ardizzone, "Probabilistic landslide hazard assessment at the basin scale," *Geomorphology*, vol. 72, no. 1–4, pp. 272–299, Dec. 2005.
[4] P. T. T. Ngo et al., "Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran," *Geosci. Front.*, vol. 12, no. 2, pp. 505–519, Aug. 2020.
[5] T. Chen, R. Niu, B. Du, and Y. Wang, "Landslide spatial susceptibility mapping by using GIS and remote sensing techniques: A case study in Zigui County, the Three Georges Reservoir, China," *Environ. Earth Sci.*, vol. 73, no. 9, pp. 5571–5583, Oct. 2014.
[6] T. Chen, R. Niu, and X. Jia, "A comparison of information value and logistic regression models in landslide susceptibility mapping by using GIS," *Environ. Earth Sci.*, vol. 75, no. 10, pp. 1–16, May 2016.
[7] A. Merghadi et al., "Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance," *Earth-Sci. Rev.*, vol. 207, Aug. 2020, Art. no. 103225.
[8] 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. Geoinformation*, vol. 107, Mar. 2022, Art. no. 102681.
[9] Y. Achour and H. R. Pourghasemi, "How do machine learning techniques help in increasing accuracy of landslide susceptibility maps?," *Geosci. Front.*, vol. 11, no. 3, pp. 871–883, May 2020.
[10] L. Zhang and L. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 270–294, Jun. 2022.
[11] M. Zhang, X. Zhao, W. Li, Y. Zhang, R. Tao, and Q. Du, "Cross-scene joint classification of multisource data with multilevel domain adaption network," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2023.3262599.

[12] L. Zhang, L. Song, B. Du, and Y. Zhang, "Nonlocal low-rank tensor completion for visual data," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 673–685, Feb. 2021.
[13] W. Huang et al., "Landslide susceptibility mapping and dynamic response along the Sichuan-Tibet transportation corridor using deep learning algorithms," *CATENA*, vol. 222, Mar. 2023, Art. no. 106866.
[14] C. Chen and L. Fan, "An attribution deep learning interpretation model for landslide susceptibility mapping in the Three Gorges Reservoir area," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Oct. 2023, Art. no. 3000515.
[15] 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, Mar. 2024.
[16] L. Wang et al., "UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 190, pp. 196–214, Aug. 2022.
[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 29th IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[18] G. Huang, Z. Liu, V. Laurens, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. 30th IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
[19] L. Lv, T. Chen, J. Dou, and A. Plaza, "A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 108, Apr. 2022, Art. no. 102713.
[20] Y. Ge, G. Liu, H. Tang, B. Zhao, and C. Xiong, "Comparative analysis of five convolutional neural networks for landslide susceptibility assessment," *Bull. Eng. Geol. Environ.*, vol. 82, no. 10, Sep. 2023, Art. no. 337.
[21] A. M. Youssef, B. Pradhan, A. Dikshit, M. M. Al-Katheri, S. S. Matar, and A. M. Mahdi, "Landslide susceptibility mapping using CNN-1D and 2D deep learning algorithms: Comparison of their performance at Asir Region, KSA," *Bull. Eng. Geol. Environ.*, vol. 81, no. 4, Mar. 2022, Art. no. 165.
[22] X. Liu, Y. Wu, W. Liang, Y. Cao, and M. Li, "High resolution SAR image classification using global-local network structure based on vision transformer and CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Feb. 2022, Art. no. 4505405.
[23] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," Oct. 2020, *arXiv:2010.11929*.
[24] D. Wang et al., "Evaluation of deep learning algorithms for landslide susceptibility mapping in an alpine-gorge area: A case study in Jiuzhaigou County," *J. Mountain Sci.*, vol. 20, no. 2, pp. 484–500, Feb. 2023.
[25] A. Jamali, S. K. Roy, and P. Ghamisi, "WetMapFormer: A unified deep CNN and vision transformer for complex wetland mapping," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 120, Jun. 2023, Art. no. 103333.
[26] A. A. Aleissaee et al., "Transformers in remote sensing: A survey," *Remote Sens.*, vol. 15, no. 7, pp. 1860–1860, Mar. 2023.
[27] J. Ma, M. Li, X. Tang, X. Zhang, F. Liu, and L. Jiao, "Homo–heterogenous transformer learning framework for RS scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2223–2239, Mar. 2022.
[28] Y. Feng, H. Xu, J. Jiang, H. Liu, and J. Zheng, "ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 4410213.
[29] G. Wang et al., "P2FEViT: Plug-and-play CNN feature embedded hybrid vision transformer for remote sensing image classification," *Remote Sens.*, vol. 15, no. 7, Jan. 2023, Art. no. 1773.
[30] T. Chen, L. Zhu, R. Niu, C. J. Trinder, L. Peng, and T. Lei, "Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models," *J. Mountain Sci.*, vol. 17, no. 3, pp. 670–685, Mar. 2020.
[31] S. Ji, D. Yu, C. Shen, W. Li, and Q. Xu, "Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks," *Landslides*, vol. 17, pp. 1337–1352, Feb. 2020.
[32] X. Gao, T. Chen, R. Niu, and A. Plaza, "Recognition and mapping of landslide using a fully convolutional DenseNet and influencing factors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7881–7894, Aug. 2021.
[33] 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, Jan. 2024.
[34] A. L. Boulesteix, A. Bender, J. L. Bermejo, and C. Strobl, "Random forest Gini importance favours SNPs with large minor allele frequency: Impact, sources and recommendations," *Brief. Bioinf.*, vol. 13, no. 3, pp. 292–304, May 2012.
[35] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[36] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, vol. 8689, pp. 818–833.
[37] J. Gu et al., "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.
[38] Y. Yi, Z. Zhang, W. Zhang, H. Jia, and J. Zhang, "Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region," *CATENA*, vol. 195, Dec. 2020, Art. no. 104851.
[39] Y. Bazi, L. Bashmal, M. M. A. Rahhal, R. A. Dayil, and N. A. Ajlan, "Vision transformers for remote sensing image classification," *Remote Sens.*, vol. 13, no. 3, p. 516, Feb. 2021.
[40] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Surv.*, vol. 54, no. 10, pp. 1–41, Jan. 2022.
[41] X. Zhao et al., "Fractional Fourier image transformer for multimodal remote sensing data classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2314–2326, Feb. 2024.
[42] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," Oct. 2018, *arXiv:1810.04805*.
[43] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, "Escaping the big data paradigm with compact transformers," Aug. 2021, *arXiv:2104.05704*.
[44] H. Shu et al., "Relation between land cover and landslide susceptibility in Val d'Aran, Pyrenees (Spain): Historical aspects, present situation and forward prediction," *Sci. Total Environ.*, vol. 693, Nov. 2019, Art. no. 133557.
[45] B. Pradhan and S. Lee, "Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models," *Environ. Earth Sci.*, vol. 60, no. 5, pp. 1037–1054, Jul. 2009.
[46] F. Huang et al., "Landslide susceptibility prediction considering neighborhood characteristics of landslide spatial datasets and hydrological slope units using remote sensing and GIS technologies," *Remote Sens.*, vol. 14, no. 18, Sep. 2022, Art. no. 4436.
[47] 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, Jul. 2023.
[48] D. Sun, H. Wen, D. Wang, and J. Xu, "A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm," *Geomorphology*, vol. 362, Aug. 2020, Art. no. 107201.
[49] S. Mehta and M. Rastegari, "MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer," Mar. 2022, *arXiv:2110.02178*.
[50] C. Cao, K. Zhu, P. Xu, B. Shan, G. Yang, and S. Song, "Refined landslide susceptibility analysis based on InSAR technology and UAV multi-source data," *J. Cleaner Prod.*, vol. 368, Sep. 2022, Art. no. 133146.
[51] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," Apr. 2018, *arXiv:1710.09412*.
[52] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. 32nd IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6022–6031.

Zeyang Zhao received the B.S. degree majoring in geographic information science from the Chengdu University of Technology, Chengdu, China, in 2021. He is currently working toward the M.S. degree in earth exploration and information technology with the China University of Geosciences, Wuhan, China.

His main current research interests include landslide susceptibility mapping and data mining of geographic information based on deep learning.

Tao Chen (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2008.

He is currently an Associate Professor with the School of Geophysics and Geomatics, China University of Geosciences, Wuhan, China. From 2015 to 2016, he was a Visiting Scholar with the University of New South Wales, Sydney, Australia. He has authored or coauthored more than 60 scientific papers including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *International Journal of Applied Earth Observation and Geoinformation*, etc., and guest edited 10 journal special issues. His research interests include image processing, machine learning, and geological remote sensing.

Dr. Chen serves as an Associate Editor for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Jie Dou received the Ph.D. degree in natural environmental studies from the University of Tokyo, Tokyo, Japan, in 2015.

He is currently a Full Professor with the China University of Geosciences, Wuhan, China. He extended his experience with the University of Tokyo, the Public Works Research Institute, etc. He has participated in research projects on earthquake/rainfall-triggered geohazards, spatial analysis with artificial intelligence (AI), and risk mitigation. He got a Japan Society for the Promotion of Science (JSPS) fellowship. He has authored or coauthored more than 100 peer-reviewed articles.

Dr. Dou served as an Associate Editor for *Frontiers in Earth Science*, and on the editorial boards of several other international journals, such as the *Journal of Mountains Science*, *Geocarto International*, and *Geomatics, Natural Hazards and Risk*. He has served as a Reviewer for more than 30 ISI-listed international journals. He has been a Steering Committee Member for several commissions/working groups of international academic societies, such as the World Landslide Forum 5, BIGS2021, BIGS2023, etc.

Gang Liu received the Ph.D. degree in cartography and geographic information engineering from Southwest Jiaotong University, Chengdu, China, in 2014.

He is currently a Professor with the College of Earth Sciences, Chengdu University of Technology, Chengdu, China. From 2018 to 2019, he was a Visiting Scholar with The University of Georgia, Athens, GA, USA. He has authored or coauthored more than 50 papers including the *International Journal of Geographical Information Science*, *Transactions in GIS*, *Computers Environment and Urban Systems*, *Sustainable Cities and Society*, and *Journal of the Geological Society*. His research interests include spatial network, spatial analysis, geodata modeling, image interpretation, etc.

Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Department of Technology of Computers and Communications, University of Extremadura, Badajoz, Spain, in 1999 and 2002, respectively.

He is currently a Full Professor and the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored more than 600 publications and guest edited 10 journal special issues. He has reviewed more than 500 manuscripts for more than 50 different journals.

Dr. Plaza served as the Editor-in-Chief for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2013 to 2017. He is included in the Highly Cited Researchers List (Clarivate Analytics) from 2018 to 2020.