A synergistic CNN-DF method for landslide susceptibility assessment

Jiangang Lu, Yi He*, Lifeng Zhang, Qing Zhang, Jiapeng Tang, Tianbao Huo, Yunhao Zhang, IEEE

***Abstract*—The complex structures and intricate hyperparameters of existing deep learning models make achieving higher accuracy in landslide susceptibility assessment time-consuming and labor-intensive. Deep forest (DF) is a decision tree-based deep learning framework that uses a cascade structure to process features, with model depth adapting to the input data. To explore a more ideal landslide susceptibility model, this study designed a landslide susceptibility model combining convolutional neural networks (CNN) and DF, referred to as CNN-DF. The Bailong River Basin, a region severely affected by landslides, was chosen as the study area. Firstly, the landslide inventory and influencing factors of the study area were obtained. Secondly, an equal number of landslide and non-landslide samples were selected under similar environmental constraints to establish the dataset. Thirdly, CNN was used to extract high-level features from the raw data, which were then input into the DF model for training and testing. Finally, the trained model was used to predict landslide susceptibility. The results showed that the CNN-DF model achieved high prediction accuracy, with an AUC of 0.9061 on the testing set, outperforming DF, CNN, and other commonly used machine learning models. In landslide susceptibility maps, the proportion of historical landslides in the very high susceptibility category of CNN-DF was also higher than that of other models. CNN-DF is feasible for landslide susceptibility assessment, offering higher efficiency and more accurate results. Additionally, the SHAP algorithm was used to quantify the contribution of features to the prediction results both globally and locally, further explaining the model. The landslide susceptibility map based on CNN-DF can provide a scientific basis for landslide prevention and disaster management in the target area.**

***Index Terms*—landslide susceptibility, deep forest, CNN, hybrid model**

I. Introduction

As one of the most common geological hazards worldwide, landslides are highly harmful and destructive, resulting in substantial casualties and property losses each year[1], [2]. The most effective way to prevent and manage landslides is regional landslide susceptibility assessment (LSA), which explore the nonlinear relationships between historical landslides and landslide influencing factors(LIFs) to predict the probability of landslides in specific areas [3], [4]. High-precision susceptibility predictions are crucial for land-use planning and disaster risk management [5].

With the advancement of computer and data mining technologies, traditional machine learning algorithms have been widely applied in LSA studies, including support vector machines (SVM), random forests (RF), Extreme Gradient Boosting (XGBoost), and artificial neural network (ANN) [6], [7], [8], [9]. These algorithms exhibit high predictive accuracy and resistance to overfitting when applied to small datasets. However, their simple model structures hinder the retention of knowledge from previous tasks and the ability to generalize, limiting further improvement in prediction accuracy [10]. With the rise of artificial intelligence, deep learning (DL) algorithms have outperformed traditional machine learning in fields such as image classification, image recognition, and speech recognition [11], [12]. Deep learning overcomes the limitations of traditional machine learning algorithms, with multi-layer structures that can fully exploit high-level features and add nonlinear representations to fit abstract features. Particularly, convolutional neural networks (CNN) are notable for their strong feature extraction capabilities. The application of CNN in LSA is still developing [13]. [14] were the first to apply CNN to LSA, demonstrating that CNN significantly improved accuracy compared to traditional machine learning methods, thus proving the applicability of CNN in LSA. Improving model accuracy is a key development direction for CNN in LSA, often achieved by deepening the network layers [15]. However, deeper networks are prone to gradient vanishing and exploding, making the model difficult to converge [16]. More layers mean increased parameters, leading to training difficulties and overfitting. Combining CNN with other models is another approach to improving accuracy. [5] combined CNN with Gated Recurrent Units (GRU) to extract neighborhood and sequential features of landslides. [17] proposed heterogeneous ensemble learning methods incorporating various DL models, achieving better results than single models. Although model combinations offer comprehensive feature learning, they often increase parameter and computational load, complicating training and slowing prediction speed. Achieving optimal results often requires significant time and effort. Nevertheless, CNN possess powerful feature extraction capabilities. By leveraging the spatial feature extraction advantages of convolutional kernels, CNN can obtain deep representations of raw data. The extracted features have high robustness and stability [18], [19]. Therefore, balancing accuracy and efficiency in LSA remains an urgent problem to solve.

Inspired by the layer-by-layer processing of deep learning algorithms, [20] proposed the deep forest (DF) algorithm, which uses a cascade structure to process raw data layer by layer for deep feature extraction. DF has four main

This work was supported in part by the Gansu Science and Technology Program under Grant 23JRRA881, in part by the National Natural Scientific Foundation of China under Grant 42201459, and in part by the Key Research and Development Project of Lanzhou Jiao Tong University under Grant LZJTU-ZDYF2301. (*Corresponding author: Yi He*)

Jiangang Lu, Yi He, Lifeng Zhang, Qing Zhang, Jiapeng Tang, Tianbao Huo, Yunhao Zhang are with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730700, China, also with the National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730700, China, and also with the Gansu Provincial Key Laboratory of Science and Technology in Surveying & Mapping, Lanzhou 730700, China (e-mail: rs_lujg@qq.com; heyi@mail.lzjtu.cn).

characteristics: (1) Compared to deep neural networks, DF does not require complex hyperparameter tuning and adaptively determines model complexity based on input data, making the training process simpler. (2) DF performs well with small sample sizes, which is particularly important in fields where data acquisition is challenging. (3) Due to its ensemble learning nature, DF has stronger noise resistance and generalization performance. (4) DF avoids gradient descent optimization, resulting in faster training speeds and less reliance on large-scale computing resources. Currently, DF has been widely used in image processing, time series prediction, and other fields, achieving excellent research results [21], [22], [23]. [24] utilized DF to classify multi-source remote sensing data, demonstrating that DF outperformed other methods in classification accuracy. In comparison to CNN, DF required significantly less time for hyperparameter tuning and greatly reduced computational complexity. While DF performs well in classification tasks, its application in LSA had not been explored prior to our work.

In conclusion, the performance of existing deep learning models in LSA is closely tied to model structure and hyperparameters, resulting in varying quality of landslide susceptibility maps. Finding an optimal landslide susceptibility model remains a hot topic. DF offers stronger interpretability and requires fewer hyperparameters, which can provide significant advantages in LSA. However, if the raw data do not adequately represent landslide information, training DF may not yield optimal prediction results. CNN excels in deep feature extraction, while DF adapts model complexity based on input data. We designed a LSA model combining CNN and DF. Additionally, to clarify the impact of features on prediction outcomes, SHapley Additive exPlanations (SHAP) was used to explain the DF model prediction [25]. SHAP can generate various plots to enhance model interpretability. This study validates and discusses the proposed model in the Bailong River Basin, providing insights for landslide prevention and risk management in the study area and exploring DF's applicability in LSA.

II. Materials

A. Study area

The Bailong River basin is located in the southern part of Gansu Province, covering an area of more than 20,000 km^2 (Fig. 1). The trunk stream of the Bailong River flows through Diebu County, Zhouqu County, Dangchang County, Wudu District, and Wenxian County. The basin has numerous tributaries, including the Minjiang River, Gongba River, Baishui River, and Rangshui River. Severe fluvial erosion and widespread landslides characterize the region. The main mountain ranges include the Minshan and Western Qinling ranges. The area features alternating high mountains, valleys, and basins, with abundant rainfall in the mid-to-high mountain zones. The annual average rainfall reaches up to 1000mm, mostly concentrated between May and September. The study area is located in the Chaidamu-West Qinling block, known for its complex regional tectonics and intense neotectonic activity. The intricate lithology and intersecting faults not only shape the diverse topography but also govern the occurrence, development, and distribution of geological hazards. Earthquakes, both within and outside the study area, have impacted the region. Notable earthquakes include the 4.5-magnitude Wenxian earthquake in 2013, the 6.6-magnitude Minxian-Zhangxian earthquake in 2013, and the 7.0-magnitude Jiuzhaigou earthquake in 2017. Earthquakes cause the fragmentation of rock masses and the accumulation of loose debris, providing material conditions for landslides. Towns and transportation routes in study area are mostly built along river valleys, where engineering and economic activities are relatively intense. As a result, the study area is highly susceptible to geological hazards.

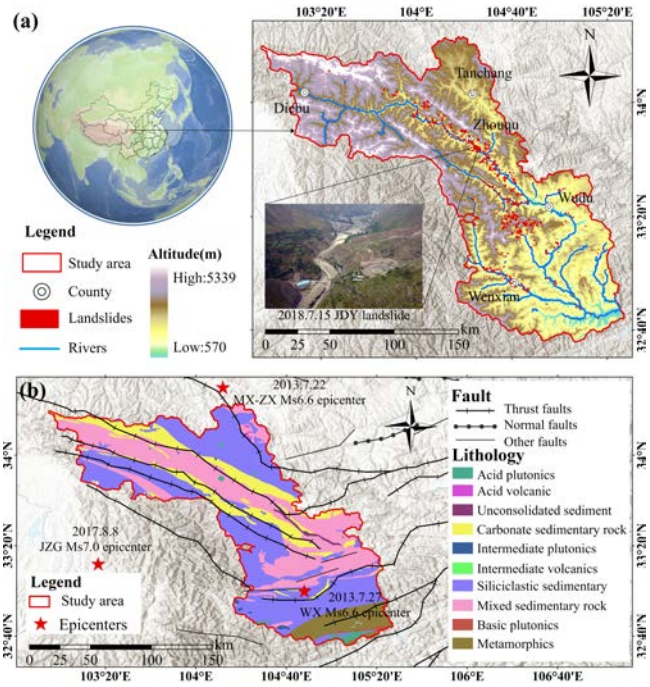

Fig. 1 Study area.

B. Landslide inventory and Landslide influencing factor

The landslide inventory map of the study area (Fig. 1) was constructed by referencing historical landslide data, visual interpretation of Google Earth imagery, and field survey data. A total of 638 landslide polygons were identified, with the largest landslide covering an area of 4.48 km², occurring in Jiangdingya, Nanyu Township on July 12, 2018. The landslides are categorized into types such as loess landslides, debris landslides, and fractured rock landslides. Most landslides were triggered by seasonal rainfall and earthquakes. The map shows a linear distribution of landslides, predominantly along riverbanks and the sides of valleys close to rivers. In addition, landslides are also concentrated at intersections and misaligned fault zones.

Through field investigations and analysis of the causes and spatial distribution characteristics of landslides in the study area, 12 Landslide Influencing Factors (LIFs) were identified. These factors include elevation, slope, profile curvature, Topographic Wetness Index (TWI), Normalized Difference Vegetation Index (NDVI), rainfall, land use, lithology, soil type, distance to roads, distance to rivers, and distance to faults. Lithology and faults were extracted from geological maps, with distance to faults calculated accordingly, and the river system was used to calculate distance to rivers. Table I lists the sources of the LIFs.

To visually demonstrate the relationship between landslides and LIFs, the landslide inventory map was overlaid on the raster images of the LIFs (Fig. 2). The distribution of landslides shows a clear pattern with respect to lithology and soil types. We consider these two factors as the internal geological environment for landslide occurrence. Regarding distance to rivers and distance to faults, landslides are mostly distributed within low-distance value ranges. The locations of faults and rivers are closely related to the internal geological environment. Therefore, the combination of these four factors is regarded as the fixed environmental characteristics for landslide occurrence, referred to as disaster-pregnant environmental factors in this study.

TABLE I

THE SOURCES OF DATA FOR LIFS USED IN THIS STUDY

Data name	Data source	Type	Resolution	LIFs
DEM	https://vertex.daac.asf.alaska.edu	Raster	12.5m	Altitude Slope Profile curvature TWI
Land use	http://globallandcover.com/	Raster	30m	Land use
River	https://www.openhistoricalmap.org	Vector	1:5000	Distance to rivers
Vegetation normalization index	http://www.geodata.cn/	Raster	1000m	NDVI
Soil type	http://www.fao.org/soils-portal	Raster	1000m	Soil type
Geological map	http://geodb.cgs.gov.cn/	Vector	1:1000,000	Lithology Distance to faults
Road	http://data.tpdc.ac.cn	Vector	1:1000,000	Distance to roads
Rainfall	http://www.geodata.cn/	Raster	1000m	Annual average cumulative rainfall

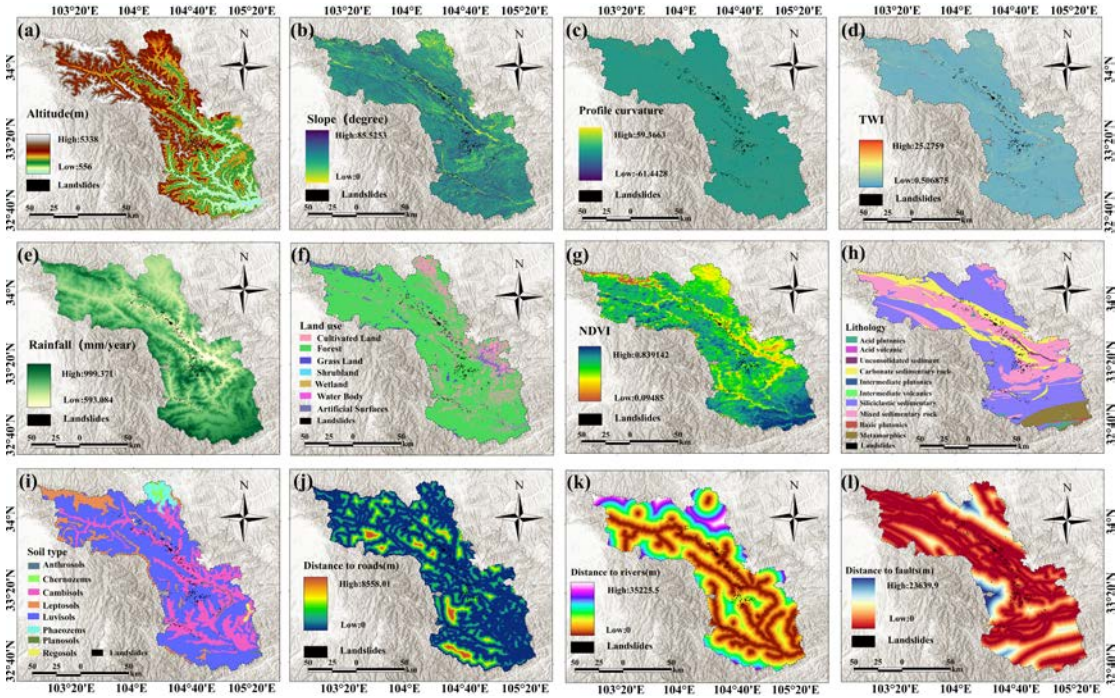

Fig. 2 The spatial distribution of LIFs.

III. METHODOLOGY

This study constructs a landslide susceptibility model integrating CNN and DF. Fig. 3 shows specific workflow. Firstly, multicollinearity tests are conducted on the obtained LIFs, and the frequency ratio of LIFs is calculated based on landslide inventory. Next, a landslide dataset is constructed based on similar disaster-prone environments, of which 70% is the training set and 30% is the testing set. Subsequently, the landslide susceptibility prediction is achieved by leveraging the feature extraction capability of CNN and the classification ability of DF. Finally, the performance of different models is evaluated by comparing accuracy metrics and landslide susceptibility maps (LSMs). Additionally, SHAP is used to explain the contribution of features to the results in the CNN-DF model.

A. Factor analysis

1) Multicollinearity

Multicollinearity refers to the presence of a high degree correlation among independent variables. If multicollinearity exists, it can introduce redundant information into the model and negatively impact the accuracy of the results [26]. Since landslides result from the combined action of LIFs, the independence of these factors is crucial for accurate susceptibility assessments. A common method for evaluating multicollinearity is the variance inflation factor (VIF). The formula for calculating VIF is given in Equation (1).

$$VIF = \frac{1}{1 - A^2} = \frac{1}{TOL} \tag{1}$$

Where A^2 is the multiple correlation coefficient of a given

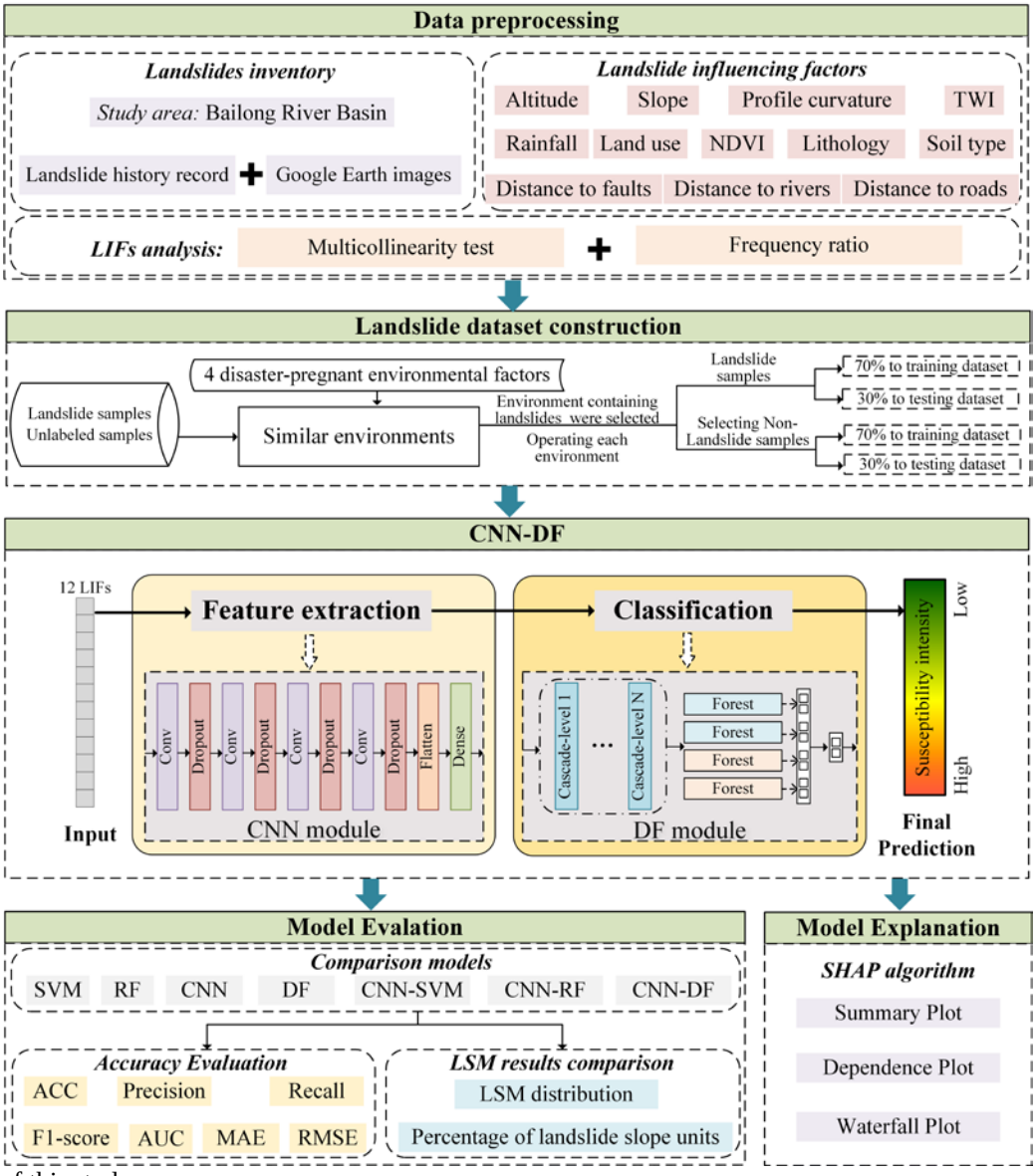

Fig. 3 Flowchart of this study.

LIF with the other LIFs in a regression analysis, while TOL stands for tolerance. The higher the correlation of this LIF with the remaining LIFs, the closer A^2 approaches 1, resulting in a larger VIF value. When VIF >5 or TOL<0.2, it indicates a significant multicollinearity problem among LIFs [27].

2) Frequency ratio

The frequency ratio (FR) is a widely used statistical approach. In landslide susceptibility assessments, it is commonly applied to quantitatively describe the correlation between classified LIFs and historical landslides[28]. The FR value is used to represent the contribution of specific intervals or subclasses of LIFs to landslide occurrences. A higher FR value indicates a greater influence of that category on landslide occurrence. The formula for calculating FR is given in Equation 2.

$$FR_{ij} = \frac{N_{ij} / N_r}{A_{ij} / A_r} \quad (2)$$

Where N_{ij} is the number of landslide grid units within the interval of the environmental factor; N_r is the total number of landslide grid units in the study area; A_{ij} is the number of grid units within the interval of the environmental factor; A_r is the total number of grid units in the study area; FR_{ij} is the frequency ratio of the j-th category of the i-th environmental factor.

B. Dataset construction

In landslide hazard assessment, common evaluation units include grid cells and slope units. Due to the highly variable terrain and complex geological environment of the study area, slope units are superior to grid units for describing the mechanical mechanisms of slopes, lithology, and environmental boundaries. Slope unit statistically aggregates the corresponding grid units, eliminating the influence of outliers and more accurately representing the terrain and its associated factor characteristics [29] . The curvature watershed method identifies terrain breakpoints based on curvature and inverse curvature to delineate watersheds, which are then

overlaid to generate slope units. This method produces relatively uniform units that capture comprehensive terrain information [30] . In this study, the study area was divided into 222,916 slope units using this method. The attributes of LIFs for each slope unit were obtained using zonal statistics tool of ArcGIS.

Similar environments are constructed based on disaster-pregnant environment factors, which enables the incorporation of prior knowledge in the non-landslide sample selection process[31]. This completes the construction of the dataset. The specific steps are as follows:

(1) FR normalization of disaster-pregnant factors. The formula for normalization is given in Equation 3.

$$FR_{ij}{}' = \frac{FR_{ij}}{\max(FR_{ij}) - \min(FR_{ij})} \tag{3}$$

Where $FR_{ij}{}'$ represents the weight between category j of LIF i and the typical category of landslide occurrences under LIF i.

(2) Similar environment classification. Each slope unit generates a weighted one-hot encoding for a total of 29 graded categories of the 4 disaster-pregnant factors. Based on similarity, all slope units are classified into 205 similar environments.

(3) Balanced sampling method based on similar environments. After processing, 638 historical landslide sites yielded 864 landslide slope units. The process for selecting non-landslide slope units based on landslide slope units and similar environments is detailed in Fig. 4. Similar environments sampling can further improve the identification of potential landslides.

Algorithm: Balanced sampling based on similar environments

Input: 864 landslide samples; 222,052 unlabeled samples; 205 similar environments

Procedure:

1. **for** ENV=1,2,…,205 **do**
2. **if** the number of landslide samples is 0, skip this environment
3. **else if** the number of landslide samples is 1, include it in the training set
4. **else** the number of landslide samples is greater than 1, they are split into training and testing sets at a ratio of 7:3
5. **if** the number of landslide samples is less than the number of unlabeled samples, an equal number of unlabeled samples are randomly selected from the unlabeled pool to serve as non-landslide samples
6. **if** the number of non-landslide samples is 1, include it in the training set
7. **else** the number of non-landslide samples is greater than 1, they are split into training and testing sets at a ratio of 7:3
8. **end for**

Output: training set; testing set

Fig. 4 Balanced sampling of landslide and non-landslide samples based on similar environments.

C. Landslide susceptibility assessment model

This study presents a landslide susceptibility model that integrates CNN with DF, termed the CNN-DF model. The structure of the model is illustrated in Fig. 5. The CNN module employs convolutional operations to learn high-level feature representations from the raw data, expanding these into a higher-dimensional space. These expanded features then serve as inputs to the Deep Forest module. The DF module leverages its strong classification capabilities to predict landslide susceptibility. The main components of the CNN-DF model are detailed as follows.

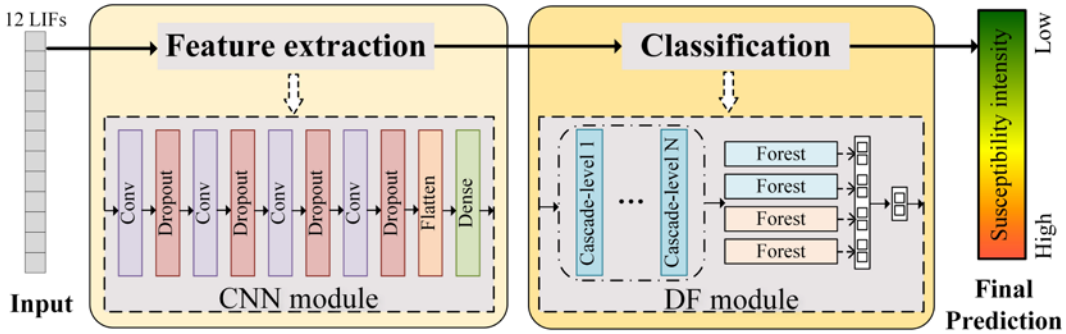

Fig. 5 Structure of the CNN-DF model.

1) CNN module

CNN is a specialized feedforward deep learning algorithm. Its network structure primarily includes an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer[11]. Due to the unique characteristics of the convolutional layers, CNN possesses strong feature extraction and nonlinear computation capabilities, allowing it to directly

extract salient features from raw data. This capability effectively avoids the need for complex mathematical methods to extract data features, which is often required in traditional deep learning methods[32]. The structure of CNN constructed in this study is shown in Fig. 6. The activation functions in the convolutional layers and non-output dense layers significantly enhance the nonlinearity of the neural network, making this parameter crucial. We chose the widely used ReLU function for these layers [33]. In the output layer, since we aim to obtain the susceptibility probability of the evaluation units, the sigmoid function was employed as the activation function. Additionally, after experimentation, we found that increasing the number of CNN layers does not necessarily improve performance. Therefore, we adopted a relatively shallow CNN model to achieve the predictions. The trained CNN model is employed for feature extraction, where raw features are input, and the output of the first dense layer is used as the extracted high-level features.

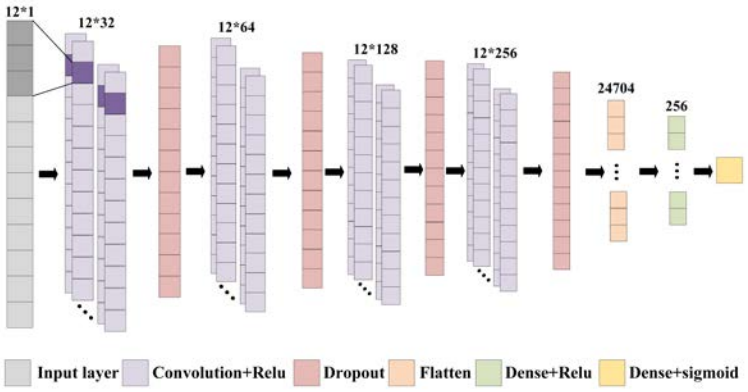

Fig. 6 Structure of CNN model.

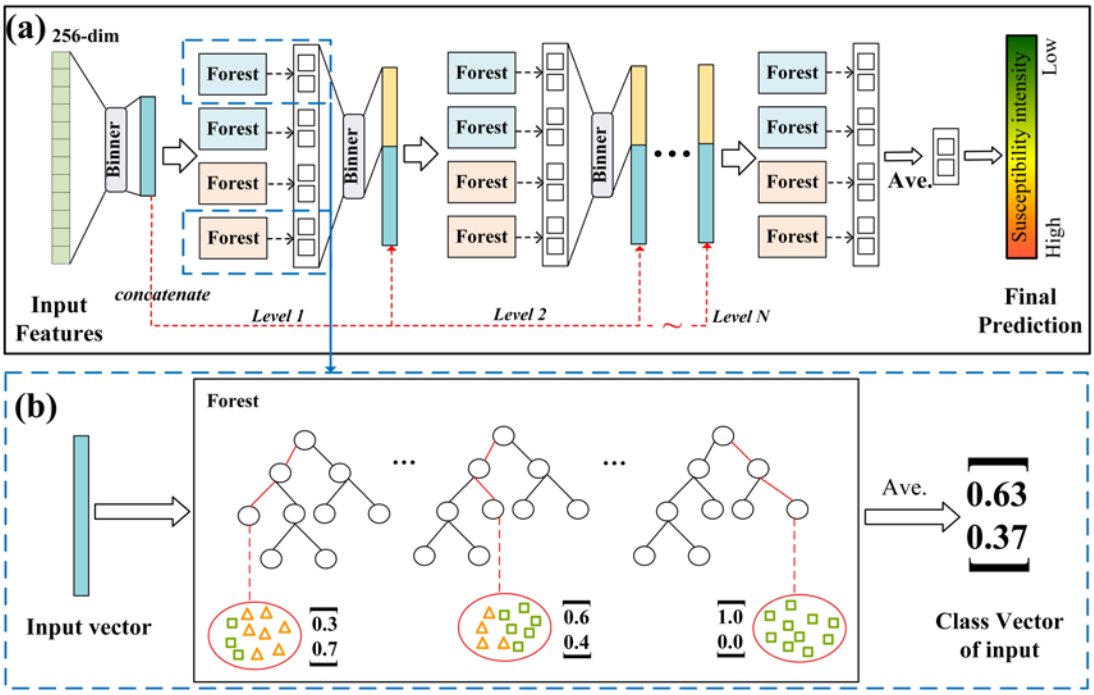

Fig. 7 (a) Structure of deep forest model. (b) Illustration of landslide class vector generation.

2) DF module

DF is a deep learning model based on RF. Similar to the layer-by-layer processing of raw features in deep neural networks, DF employs a cascade structure for feature learning, thereby fully utilizing deep features to enhance classification performance[20]. As shown in Fig. 7a, the input to the DF model consists of a sequence of features, with a length of 256, extracted by the CNN. The final output is the probability that the input sample x belongs to class 1, representing landslide susceptibility. In the Binner module, feature values are discretized into multiple unique values, which accelerates the construction of decision trees. The model has a depth of N layers, with each layer comprising two Random Forests and two Completely Random Forests. The subsequent cascade layer receives the feature information processed by the previous layer and concatenates this output with the input vector before passing it to the next layer. When transferring to a new level, the model's performance is evaluated using a validation set. If there is no significant performance gain, the training process is terminated. Consequently, the number of cascade layers is determined automatically by the model.

The process of forming class vectors for each forest is illustrated in Fig. 7b. The Completely Random Forest is

composed of multiple trees, each containing all features but randomly selecting one feature as the split node for the split tree. Splitting continues until each leaf node contains a single class. Random Forests also consist of multiple trees, with each tree randomly selecting $\sqrt{d}$ features (d is the total number of features), and then determining the split node based on the Gini score, as defined in Equation 4. To reduce the risk of overfitting, the class vectors generated by each forest are produced through K-fold cross-validation. Then the class distribution of all leaf nodes is averaged to get the class vector.

$$Gini = 1 - \sum_{i}^{k} {p_i}^2 \tag{4}$$

Where k is the number of sample classes, which is 2 in this study; p_i denotes the proportion of samples of class i in the dataset at the current node.

Confusion Matrix		Predicted Class: Positive	Predicted Class: Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
Actual Class	Negative	False Positive (FP)	True Negative (TN)

$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$	
$Precision = \frac{TP}{TP + FP}$	$Recall = \frac{TP}{TP + FN}$
$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	
$FPR = \frac{FP}{FP + TN}$	$TPR = \frac{TP}{TP + FN}$

Fig. 8 Calculation of evaluation metrics.

D. *Evaluation methods*

The confusion matrix is a matrix representation used to measure the performance of a classification model. The model's predictions on the testing set are classified as landslide if the predicted value is greater than or equal to 0.5, and as non-landslide if it is less than 0.5. Using the true classes and predicted classes, a confusion matrix can be constructed [34]. As shown in Fig. 8, based on the confusion matrix, evaluation metrics such as Accuracy, Precision, Recall, F1 score, FPR (False Positive Rate), and TPR (True Positive Rate) can be computed. Here, FPR represents the X-axis and TPR represents the Y-axis of the ROC (Receiver Operating Characteristic) curve, and the AUC (Area Under the Curve) value is also a commonly used performance metric. The larger the AUC value, the better the model's performance and accuracy.

E. *SHapley Additive exPlanations (SHAP)*

The core concept of SHAP is rooted in the Shapley value from cooperative game theory, originally employed to quantify players' contributions in cooperative games[35]. When applied to machine learning models, SHAP values are utilized to quantify the contribution of each feature to model predictions, computed according to Equation 5. SHAP explains the DF prediction outcome as the sum of contribution values from 256 input features, formulated according to Equation 6. For local explanations, the SHAP algorithm processes the model predictions according to Equation 7.

$$\phi_i = \sum_{S \subseteq N \backslash \{i\}} \frac{|S|! \cdot (n - |S| - 1)}{n!} [f(S \cup \{i\}) - f(S)] \tag{5}$$

Where N represents the set of all features; $S \subseteq N \backslash \{i\}$ denotes a subset of features excluding feature i ; $|S|$ denotes the size of subset S ; $f(S)$ represents the model prediction on subset S; $(S \cup \{i\})$ denotes the model prediction when feature i is added to subset S.

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i' \tag{6}$$

Where $g(z')$ represents the model's output being explained; z_i' denotes the binary representation of the corresponding feature; ϕ_0 denotes the mean prediction across all samples; M represents the total number of features.

$$f(x) = \ln \frac{P}{1 - P} \tag{7}$$

Where $f(x)$ represents the processed value; P denotes the model prediction result.

IV. Results and Analysis

A. *Landslide influencing factor analysis*

The multicollinearity test results for the 12 LIFs calculated using Python are shown in Table II. The VIF values of the selected LIFs are all below 5, indicating no significant multicollinearity among the 12 LIFs. Rainfall has the highest VIF value and the lowest TOL value, at 3.0043 and 0.3329, respectively, so no adjustments to the LIFs are necessary. They can be used as explanatory variables for landslides susceptibility modeling in this study.

The frequency ratio statistics of the classified influencing factors can further analyze the spatial development patterns of landslides. The frequency ratio results for all LIFs are shown in Fig. 9. The elevation results show that landslides mostly occur below 3000m, with the highest propensity between 1500-2000m, corresponding to the elevation along riverbanks. The slope range with the highest landslide propensity is 30°-45°. The profile curvature results show similar landslide propensities within the ranges (-20,0) and (0,20). The TWI values show a generally negative correlation with landslide propensity. The annual average rainfall results show that landslides mostly occur in the range of 600-900 mm/year; rainfall can destabilize slopes and trigger landslides. The land use results show that forest areas have the highest landslide propensity, as the strong transpiration of forests can induce heavy rainfall, increasing landslide risk. The NDVI results show that landslide propensity is higher within the range of 0.4-0.7, indicating that vegetation cover indirectly affects slope stability. The lithology results show higher FR values under conditions of mixed sedimentary rocks and carbonate sedimentary rocks, indicating a higher landslide propensity on soft rock slopes. The soil type results show higher FR values under conditions of Cambisols and Luvisols, indicating a higher landslide propensity in fine-textured soils. These types of

lithology and soil, with higher landslide propensity, are mostly distributed along riverbanks. The results for distance to roads, rivers, and faults show a negative correlation with landslide propensity. Compared to other factors, the FR results for lithology, soil type, distance to rivers, and distance to faults show very clear landslide propensities, confirming the rationality of the chosen disaster-pregnant environmental factors.

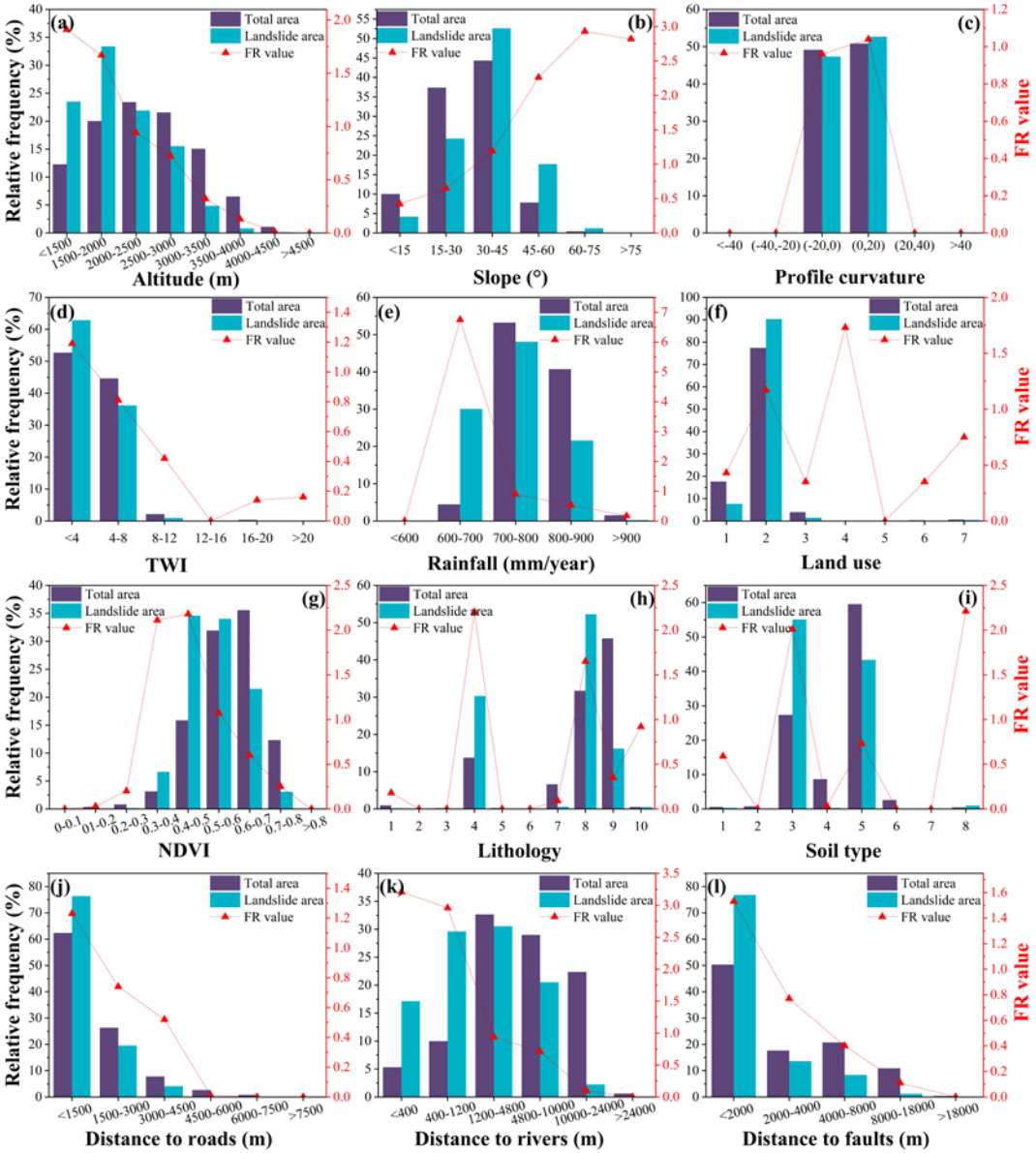

Fig. 9 Area proportion, landslide area proportion and FR value of each LIF subcategory. (i)The land use types include: Cultivated Land (1), Forest (2), Grass Land (3), Shrubland (4), Wetland (5), Water Body (6), Artificial Surfaces (7). (g) The lithology types include: Acid plutonics (1), Acid volcanic (2), Unconsolidated sediment (3), Carbonate sedimentary rock (4), Intermediate plutonics (5), Intermediate volcanics (6), Siliciclastic sedimentary (7), Mixed sedimentary rock (8), Basic plutonics (9), Metamorphics (10). (h) The soil types include: Anthrosols (1), Chernozems (2), Cambisols (3), Leptosols (4), Luvisols (5), Phaeozems (6), Planosols (7), Regosols (8).

TABLE II
THE VIF AND TOL VALUE OF LIFS MULTICOLLINEARITY ANALYSIS

LIFs	VIF	TOL
Altitude	2.74	0.37
Slope	1.91	0.52
Profile curvature	1.05	0.96
TWI	1.92	0.52
Rainfall	3.00	0.33
Land use	1.08	0.93
NDVI	1.68	0.60
Lithology	1.18	0.85
Soil type	2.0812	0.4805
Distance to roads	1.3837	0.7227
Distance to rivers	2.5869	0.3866
Distance to faults	1.0893	0.918

B. Evaluation of model performance

Hybrid models are an important approach to improving the performance of machine learning models [17], [36]. The performance of the DF model significantly influences the

performance of hybrid models in this study. To validate the performance of the DF model, we also employed RF, SVM and CNN for susceptibility modeling, with model parameters determined through trial and error. Similar to the CNN-DF model, we also established CNN-RF and CNN-SVM models. To ascertain whether the features derived from convolutional operations of CNN improve the predictive results of the DF model compared to the original features, the testing set evaluation metrics are presented in Table III and Fig.10. The CNN-DF model not only outperforms the DF and CNN models but also surpasses all other single and hybrid models in accuracy. The DF model shows significantly higher evaluation metrics compared to the SVM, RF, and CNN models, indicating that both the DF model and the CNN-DF model exhibit good fitting accuracy and predictive performance. When the features processed by CNN are input into the DF classifier, the model's predictive performance improves, producing more accurate susceptibility results. The DF classifier achieves better recognition with the new features compared to the original features. In Fig. 10b, the CNN-SVM model shows an increase in AUC by 0.0013 compared to the SVM model, but a decrease by 0.0124 compared to the CNN model. The hybrid model of SVM did not achieve the expected improvement in evaluation metrics. In contrast, the CNN-RF model shows an increase in AUC by 0.0033 and 0.0011 compared to the RF and CNN models, respectively, consistent with the trend of AUC improvement seen in the CNN-DF model. This indicates that the strategy of hybrid modeling is feasible.

TABLE III
ACCURATE METRICS FOR DIFFERENT MODELS USING TESTING DATASET

Models	ACC	Precision	Recall	F1-score
SVM	80.29%	0.7622	0.8699	0.8125
RF	80.48%	0.7949	0.8216	0.8080
CNN	80.86%	0.7712	0.8773	0.8209
DF	82.71%	0.8077	0.8587	0.8324
CNN-SVM	80.30%	0.7828	0.8439	0.8122
CNN-RF	80.48%	0.7790	0.8513	0.8135
CNN-DF	**82.90%**	**0.7960**	**0.8848**	**0.8380**

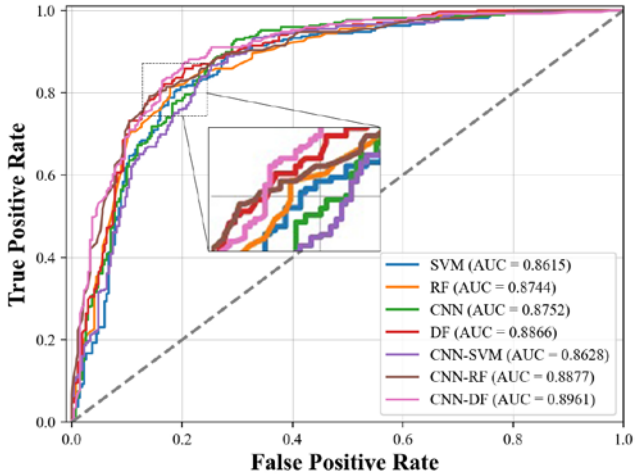

Fig. 10 ROC curves of different models using testing dataset.

C. LSM

The trained single models and hybrid models were applied to predict and visualize landslide susceptibility across the entire Bailong River Basin. The final LSMs clearly illustrate the susceptibility in different areas, aiding government and the public in disaster prevention planning. Due to the uneven distribution of susceptibility probabilities, natural breaks classification was used to categorize the results into very high, high, medium, low, and very low susceptibility[37], [38]. The natural breaks method identifies natural groupings inherent in the data, maximizing differences between categories while ensuring values within each category are as similar as possible. The LSMs for each model are shown in Fig. 11. To highlight differences in model susceptibility results, specific areas were enlarged to show more detailed features. Generally, the majority of very high susceptibility (VHS) areas in the CNN, DF, and CNN-DF models are distributed along riverbanks, with concentrated VHS regions in the central part of the study area, corresponding to the spatial distribution of rainfall. In the CNN results, concentrated VHS areas also appear in the northwest and southeast boundaries, correlating with high rainfall regions, indicating the CNN model's sensitivity to rainfall features. In the CNN-DF results, these concentrated areas are reduced, indicating the improvement in model accuracy. In terms of area proportions for each susceptibility level, the DF model's VHS areas account for only 25.1%, and VHS areas for 6.22%, which is not conducive to landslide prevention. In contrast, the CNN model shows relatively high proportions of very high susceptibility (VLS) and VHS areas, which is also not ideal. The CNN-DF model resolves this issue, with its VLS and VHS areas falling in between the proportions of the two single models, enhancing reliability. Similar phenomena are observed in the CNN-SVM and CNN-RF results, where the proportion of VLS and VHS areas is significantly reduced, demonstrating the effectiveness of the hybrid model improvement strategy. Comparing the DF model's susceptibility results with those of RF, SVM, and CNN models, RF and DF results are similar, whereas SVM and CNN show large VHS areas in the northwest boundary, affecting result reliability.

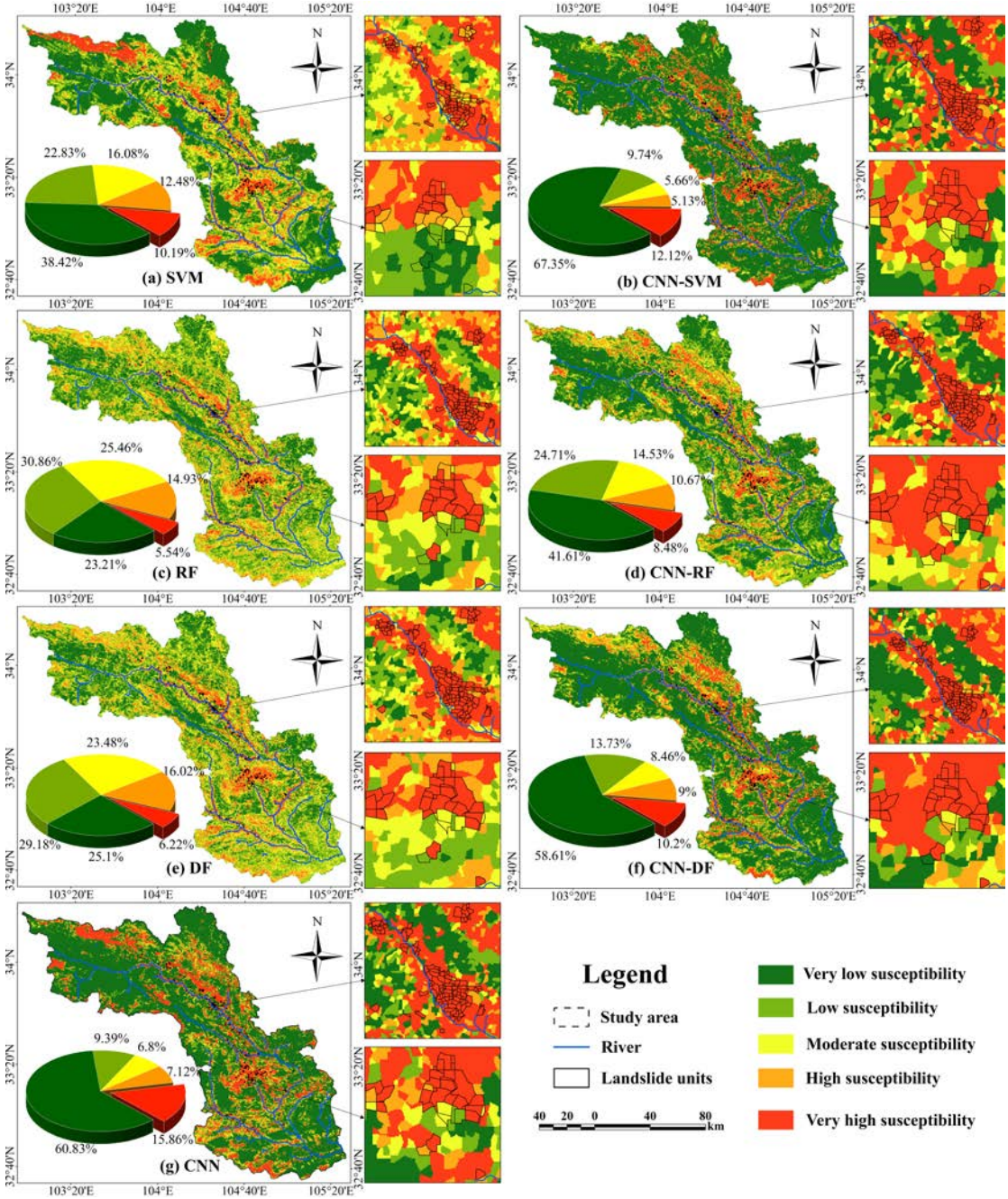

Fig. 11 LSMs generated using (a) SVM, (b) CNN-SVM, (c) RF, (d) CNN-RF, (e)DF, (f) CNN-DF, (g) CNN.

The proportion of slope units with historical landslides in the VHS level can also validate LSM reliability [39]. As shown in Fig. 12, we counted the proportion of landslide slope units in each susceptibility level for the seven models. In the CNN-DF model, 90.76% of slope units with historical landslides fall into the VHS area, the highest among all models, exceeding the DF and CNN models by 0.58 and 3.51, respectively. Similar improvement trends were observed in the CNN-SVM and CNN-RF models. These phenomena are also visible in detailed figures, indicating the hybrid models' enhanced prediction accuracy. Compared to single models, the DF model has a higher proportion of historical landslides in the VHS area than the SVM and RF models, demonstrating DF's excellent predictive capability, closely related to the superior performance of the CNN-DF model. The statistical results of landslide slope units align with the model evaluation metrics in section 4.2, confirming that the CNN-DF model has the best performance and accuracy.

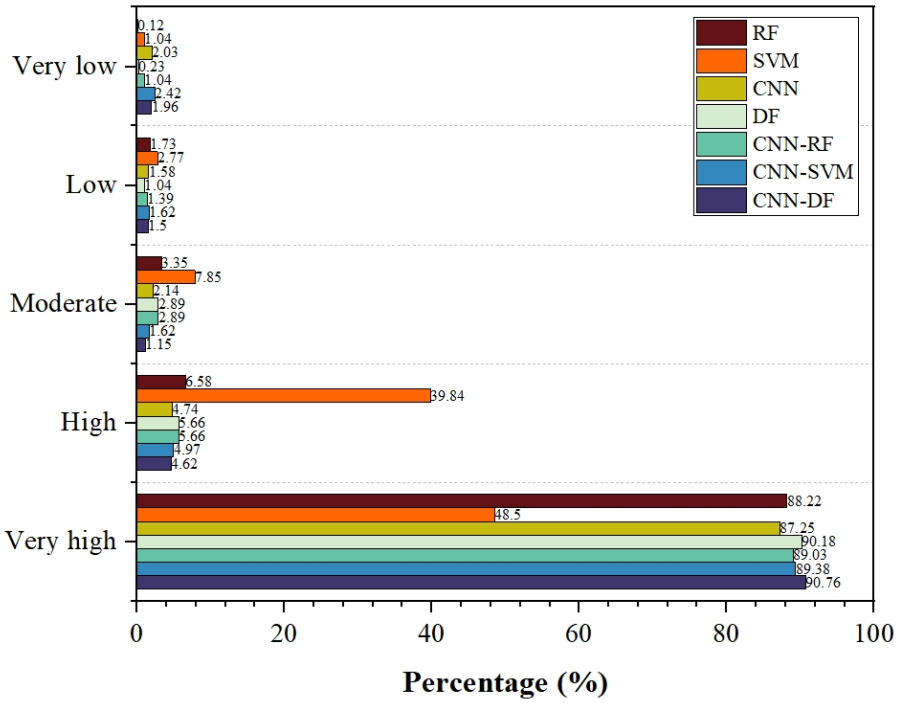

Fig. 12 Percentage of historical landslide slope units in landslide susceptibility classes.

V. Discussion

A. Performance of models

1) Advantages of Deep Forest

Despite the numerous landslide susceptibility models based on machine learning and deep learning, models like DF that can automatically determine parameters and halt training when performance no longer improves are rare. This characteristic eliminates the need for manual hyperparameter tuning, making the prediction results more rational and accurate. SVM, a classic machine learning model, classifies using a hyperplane with hyperparameters determined through grid search [40]. However, it performs the worst in terms of the proportion of historical landslide slope units, indicating its insufficient ability to capture nonlinear relationships. RF is a tree-based ensemble learning method. Comparing various accuracy metrics, the performance of DF exceeds that of RF. This is because RF only trains multiple decision trees simultaneously and uses a voting mechanism to obtain the final result, enhancing prediction accuracy [41], [42]. In contrast, DF introduces a cascade structure, which allows for more comprehensive feature learning. Additionally, the use of k-fold cross-validation helps avoid overfitting, thereby improving the accuracy of the landslide susceptibility assessment model. While CNN demonstrates good performance in feature extraction and classification, training a higher-performance model requires significant expertise in model architecture design and hyperparameter tuning [14]. The DF model has the advantage of complexity being adaptively adjusted based on the input training data, making it broadly applicable to datasets of different scales.

2) Advantages of hybrid model

Ensemble learning is gaining popularity, especially in complex environments and large-scale scenarios, where hybrid models often yield better predictions and higher accuracy than single base models. Hybrid models can provide high-quality predictions when different models yield varying results in the same region [43]. Due to the limited application of DF in landslide susceptibility, there is currently no hybrid model incorporating DF for this purpose.

Both CNN and DF represent cutting-edge deep learning models. CNN employs convolutional layers for powerful feature extraction and offers structural flexibility. It has been applied multiple times in landslide susceptibility mapping, consistently achieving excellent prediction results [19]. A trained CNN exhibits stable feature learning without overfitting, enabling the extraction of high-quality data features. The DF model has only one parameter, the number of cascade layers, which is automatically determined during the training process. By combining the advantages of both models, we innovatively propose the CNN-DF model. This hybrid model focuses solely on the structure to build a highly robust CNN. For performance evaluation, it is crucial that the training and testing sets have similar distributions. The evaluation metrics derived from such a testing set can more accurately reflect model performance [44]. In this study, the training and testing sets were selected based on similar environments, encompassing various complex geological conditions of landslides, thus meeting the requirement of similar distribution. According to the evaluation metrics, our constructed CNN model exhibits excellent predictive ability and is suitable for feature extraction. Comparing the accuracy metrics of the testing set, the

performance of the CNN-DF model is superior to that of the DF model. Further comparison of the proportion of historical landslide slope units reveals a 0.58% improvement in the CNN-DF results over the DF model, indicating that the CNN-DF model produces the most optimal results. Additionally, we selected a landslide not included in the training and testing sets for susceptibility comparison and field validation to assess the model's accuracy. As shown in Fig. 13, the location of this landslide was predicted as VHS by both the CNN and DF models in part, whereas the CNN-DF model predicted the entire area as VHS. This indicates that the CNN-DF model, by combining the strengths of both models, achieved more accurate predictions.

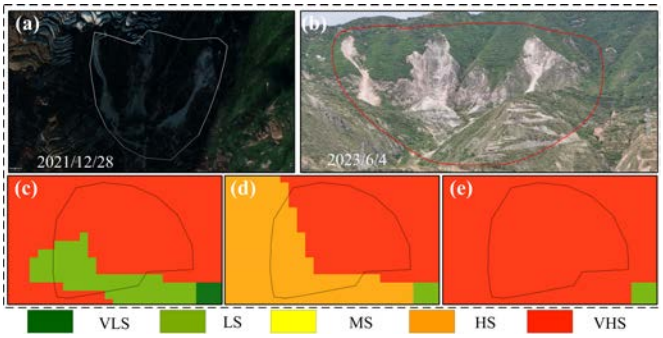

Fig. 13 (a) Google Earth image, (b) field verification photo, (c) LSM generated using CNN, (d) LSM generated using DF, (e) LSM generated using CNN-DF.

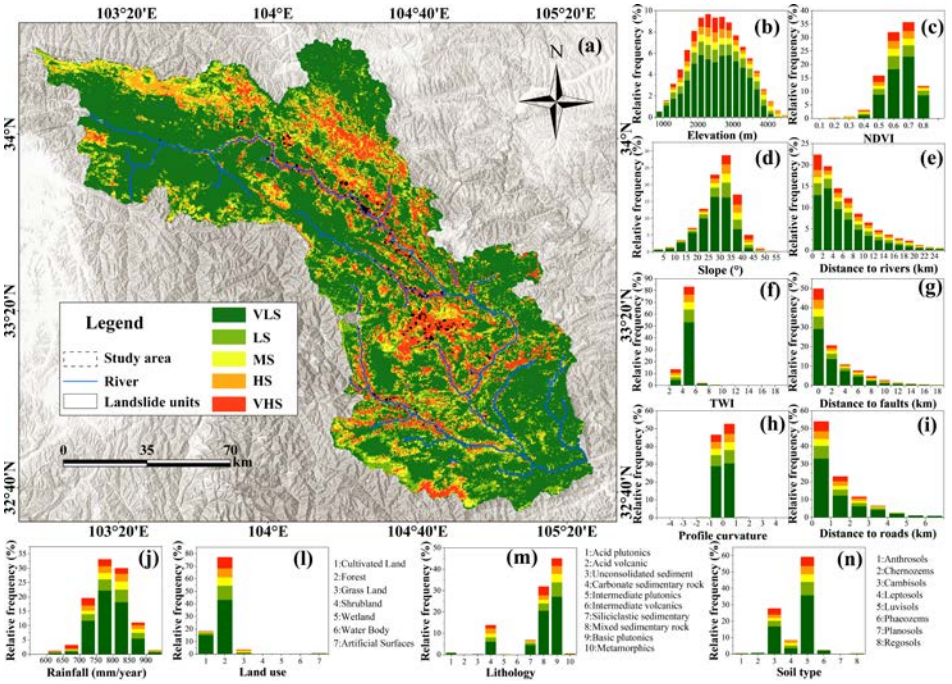

Fig. 14 (a) The LSM result of CNN-DF . (b-n) Stacked bar chart of the area proportion of the five susceptibility classes in each LIF subcategory.

3) Analysis of landslide-prone areas

Fig. 14 presents a stacked bar chart showing the area proportions of the five susceptibility levels within each LIF classification. This provides a quantitative and intuitive observation of the distribution of landslide-prone areas [45]. The distribution of VHS areas within the factors correlates clearly with the frequency ratio analysis results of the factors, further demonstrating the reliability of the hybrid model. The VHS area proportion is negatively correlated with the distance to rivers, faults, and roads. Rivers and faults are considered internal fixed features of the study area. Roads are generally constructed along human settlements, indicating that human activities contribute to the occurrence of landslides. The VHS proportion shows distinct patterns in specific lithological and soil type categories, with mixed sedimentary rock, carbonate sedimentary rock, cambisols and luvisols predominantly distributed along riverbanks, underscoring the close relationship with internal fixed features. The non-landslide samples were selected based on similar environmental constraints where landslides develop. In terms of land use, VHS areas are predominantly in forests, where strong transpiration increases the frequency of rainfall. The less evident patterns in other factors do not imply a lack of contribution, as landslide development and triggering result from the combined effect of multiple factors.

B. SHAP analysis

The SHAP algorithm calculates the marginal contributions of the 256 sequential features extracted by the CNN to the model's predictions, providing both global and local explanations for the model. SHAP values are determined by the magnitude of the feature values of the samples[46]. Based on the SHAP algorithm, the feature importance ranking of the CNN-DF model can be obtained (Fig. 15). Fig. 15a shows the global feature importance, where the mean absolute SHAP value of each feature across all given samples is considered as the importance of that feature. Fig. 15b is a summary plot, which

combines feature importance and feature value changes, showing the SHAP value of each feature for single sample. The most important feature is feature 245. Among the top 20 features displayed, except for features 134 and 126, all other features show a negative contribution as their values increase. Fig. 16 is a dependence plot, illustrating the interaction effects of features 245 and 254 on the model's prediction results. Using the SHAP algorithm to visualize local explanation information, a waterfall plot for a single sample is obtained (Fig. 17). In the displayed top 10 features, features 3 and 63 promote the occurrence of landslides, while the remaining features inhibit landslides. The local explanation plots based on the SHAP algorithm can effectively elucidate the reasons behind the prediction results of individual samples.

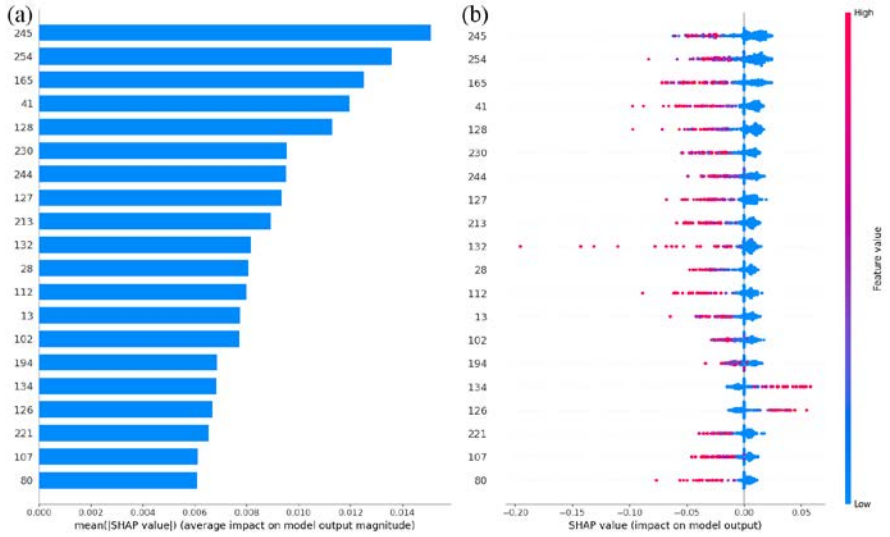

Fig. 15 Plots of feature importance ranking.

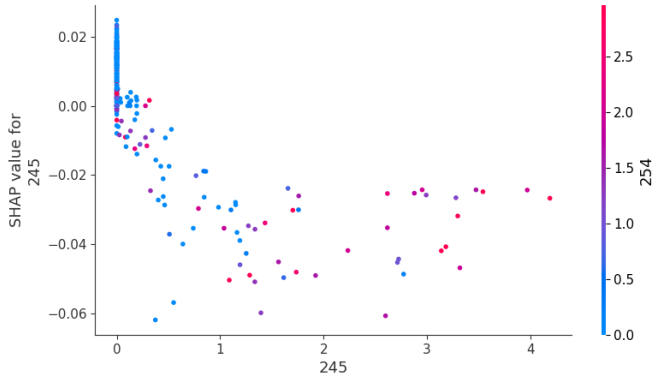

Fig. 16 SHAP dependence plot for feature 245 and 254.

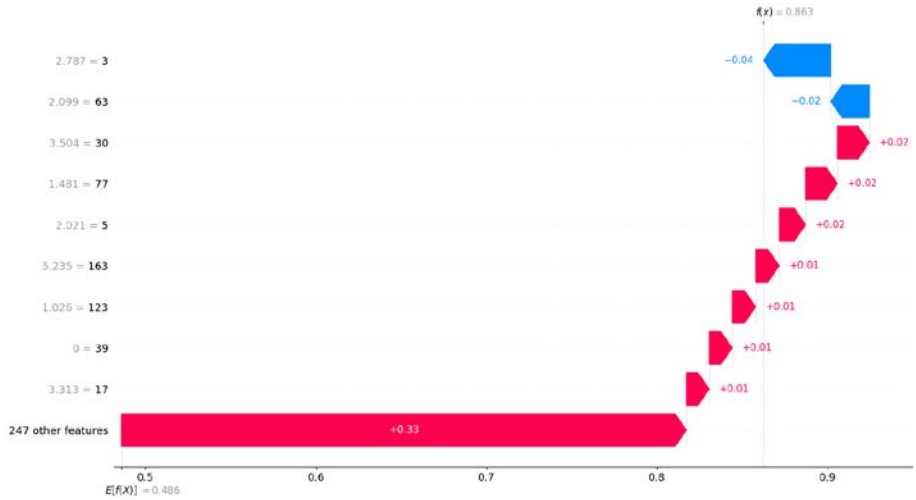

Fig. 17 The influence of single feature on model prediction. f(x) denotes the predicted value of a single sample; E[f(x)] denotes the expected value of f(x) for all samples.

C. Generalization ability of the proposed model

To validate the generalization capability of the proposed model, we conducted a test in the main urban area of Lanzhou City. Detailed information on historical landslide data and influencing factors can be found in the literature[5]. The dataset was constructed based on raster units. Non-landslide units were randomly selected in equal numbers from non-landslide areas based on the existing landslide units. The dataset was split into a training set and a validation set at a 7:3 ratio. The model structure and hyperparameter settings were kept consistent with those used in the Bailong River Basin. The test results showed that the proposed model achieved an accuracy of 96.21%, an F1-score of 0.9626, and an AUC value of 0.9898 on the validation set, outperforming the SVM, RF, and CNN models. The predicted susceptibility map is shown in Fig. 18. Compared to the LSMs produced by the SVM, RF, and CNN models, the LSM generated by the proposed model shows a smaller area of VHS around historical landslides and the largest proportion of VLS areas, which is more favorable for landslide prevention and control [31]. The experimental results demonstrate that the proposed model performs excellently across different scenarios and exhibits strong generalization capability.

In summary, the CNN-DF model leverages the optimization capabilities of DF to effectively handle classification problems in high-dimensional feature spaces. It avoids overfitting and poor classification performance issues associated with traditional CNN methods due to fully connected layers. These innovations confer the model with excellent performance and flexibility in classification tasks.

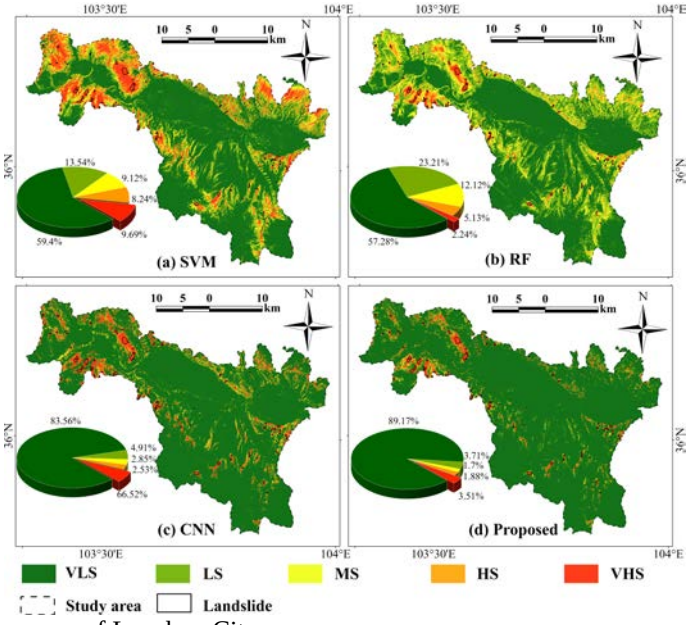

Fig. 18 LSMs of the main urban area of Lanzhou City.

VI. Conclusion

This study presents a LSA model that combines CNN and DF, applied to the Bailong River Basin in China. It is the first instance of using DF in LSA. The CNN's multiple convolutional layers are employed to extract high-level features, which are then used as inputs for the DF model instead of raw data. The performance of the CNN-DF model was validated using relevant evaluation metrics and compared with the prediction results of CNN, DF, SVM, RF, CNN-SVM, and CNN-RF. By comparing various accuracy metrics, landslide susceptibility maps, and the historical landslide slope unit proportions in VHS, the following conclusions were drawn: (1) DF outperforms CNN, SVM, and RF in prediction capability, and CNN-DF outperforms CNN-SVM and CNN-RF, indicating the superior performance of DF. (2) The prediction ability of CNN-DF is superior to that of DF, with an AUC value increase of 0.02 and an improvement of 0.58% in the historical landslide slope unit proportion. This demonstrates that the features extracted by CNN are more conducive to identification by DF classifier, further enhancing the predictive accuracy and performance of DF. Additionally, the SHAP algorithm was used to explain how the model achieves its predictions. Therefore, the CNN-DF model is feasible for LSM, offering higher efficiency and more accurate results. The LSM results of this study can provide a scientific basis for landslide prevention and disaster management in the target area. In the future, we will explore more ideal landslide susceptibility models.

Reference

[1] K. Dai *et al.*, "Entering the Era of Earth Observation-Based Landslide Warning Systems: A Novel and Exciting Framework," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 1, pp. 136–153, Mar. 2020, doi: 10.1109/MGRS.2019.2954395.

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

10.1080/19475705.2023.2185120.

[3] A. Merghadi *et al.*, "Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance," *Earth-Science Reviews*, vol. 207, p. 103225, Aug. 2020, doi: 10.1016/j.earscirev.2020.103225.

[4] A. V. Thomas *et al.*, "Landslide Susceptibility Zonation of Idukki District Using GIS in the Aftermath of 2018 Kerala Floods and Landslides: a Comparison of AHP and Frequency Ratio Methods," *J geovis spat anal*, vol. 5, no. 2, p. 21, Dec. 2021, doi: 10.1007/s41651-021-00090-x.

[5] Y. He *et al.*, "A unified network of information considering superimposed landslide factors sequence and pixel spatial neighbourhood for landslide susceptibility mapping," *International Journal of Applied Earth Observation and Geoinformation*, vol. 104, p. 102508, Dec. 2021, doi: 10.1016/j.jag.2021.102508.

[6] I. Yilmaz, "Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine," *Environ Earth Sci*, vol. 61, no. 4, pp. 821–836, Aug. 2010, doi: 10.1007/s12665-009-0394-9.

[7] R. Can, S. Kocaman, and C. Gokceoglu, "A Comprehensive Assessment of XGBoost Algorithm for Landslide Susceptibility Mapping in the Upper Basin of Ataturk Dam, Turkey," *Applied Sciences*, vol. 11, no. 11, p. 4993, May 2021, doi: 10.3390/app11114993.

[8] Z. Chang, J. Huang, F. Huang, K. Bhuyan, S. R. Meena, and F. Catani, "Uncertainty analysis of non-landslide sample selection in landslide susceptibility prediction using slope unit-based machine learning models," *Gondwana Research*, vol. 117, pp. 307–320, May 2023, doi: 10.1016/j.gr.2023.02.007.

[9] F. Huang *et al.*, "Uncertainties of landslide susceptibility prediction: Influences of random errors in landslide conditioning factors and errors reduction by low pass filter method," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 16, no. 1, pp. 213–230, Jan. 2024, doi: 10.1016/j.jrmge.2023.11.001.

[10] Y. Wang, Z. Fang, M. Wang, L. Peng, and H. Hong, "Comparative study of landslide susceptibility mapping with different recurrent neural networks," *Computers & Geosciences*, vol. 138, p. 104445, May 2020, doi: 10.1016/j.cageo.2020.104445.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[12] L. Kong and J. Cheng, "Classification and detection of COVID-19 X-Ray images based on DenseNet and VGG16 feature fusion," *Biomedical Signal Processing and Control*, vol. 77, p. 103772, Aug. 2022, doi: 10.1016/j.bspc.2022.103772.

[13] Z. Ma and G. Mei, "Deep learning for geological hazards analysis: Data, models, applications, and opportunities," *Earth-Science Reviews*, vol. 223, p. 103858, Dec. 2021, doi: 10.1016/j.earscirev.2021.103858.

[14] Y. Wang, Z. Fang, and H. Hong, "Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China," *Science of The Total Environment*, vol. 666, pp. 975–993, May 2019, doi: 10.1016/j.scitotenv.2019.02.263.

[15] P. T. Thi Ngo *et al.*, "Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran," *Geoscience Frontiers*, vol. 12, no. 2, pp. 505–519, Mar. 2021, doi: 10.1016/j.gsf.2020.06.013.

[16] O. Ghorbanzadeh and T. Blaschke, "Optimizing Sample Patches Selection of CNN to Improve the mIOU on Landslide Detection:," in *Proceedings of the 5th International Conference on Geographical Information Systems Theory, Applications and Management*, Heraklion, Crete, Greece: SCITEPRESS - Science and Technology Publications, 2019, pp. 33–40. doi: 10.5220/0007675300330040.

[17] L. Lv, T. Chen, J. Dou, and A. Plaza, "A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping," *International Journal of Applied Earth Observation and Geoinformation*, vol. 108, p. 102713, Apr. 2022, doi: 10.1016/j.jag.2022.102713.

[18] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised Deep Feature Extraction for Remote Sensing Image Classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016, doi: 10.1109/TGRS.2015.2478379.

[19] Z. Fang, Y. Wang, L. Peng, and H. Hong, "Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping," *Computers & Geosciences*, vol. 139, p. 104470, Jun. 2020, doi: 10.1016/j.cageo.2020.104470.

[20] Z.-H. Zhou and J. Feng, "Deep Forest: Towards An Alternative to Deep Neural Networks," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, Aug. 2017, pp. 3553–3559. doi: 10.24963/ijcai.2017/497.

[21] L. Yang, X.-Z. Wu, Y. Jiang, and Z.-H. Zhou, "Multi-Label Learning with Deep Forest," 2019, *arXiv*. doi: 10.48550/ARXIV.1911.06557.

[22] Z. Chen, T. Wang, H. Cai, S. K. Mondal, and J. P. Sahoo, "BLB-gcForest: A High-Performance Distributed Deep Forest with Adaptive sub-Forest Splitting," *IEEE Trans. Parallel Distrib. Syst.*, pp. 1–1, 2021, doi: 10.1109/TPDS.2021.3133544.

[23] J. Huang, P. Chen, L. Lu, Y. Deng, and Q. Zou, "WCDForest: a weighted cascade deep forest model toward the classification tasks," *Appl Intell*, vol. 53, no. 23, pp. 29169–29182, Dec. 2023, doi: 10.1007/s10489-023-04794-z.

[24] J. Xia, Z. Ming, and A. Iwasaki, "Multiple Sources Data Fusion Via Deep Forest," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia: IEEE, Jul. 2018, pp. 1722–1725. doi: 10.1109/IGARSS.2018.8517679.

[25] B. Pradhan, A. Dikshit, S. Lee, and H. Kim, "An explainable AI (XAI) model for landslide susceptibility modeling," *Applied Soft Computing*, vol. 142, p. 110324, Jul. 2023, doi: 10.1016/j.asoc.2023.110324.

[26] Y. He *et al.*, "Thaw Slump Susceptibility Mapping Based on Sample Optimization and Ensemble Learning

Techniques in Qinghai-Tibet Railway Corridor," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 17, pp. 5443–5459, 2024, doi: 10.1109/JSTARS.2024.3368039.

[27] 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. Top. Appl. Earth Observations Remote Sensing*, vol. 17, pp. 5946–5969, 2024, doi: 10.1109/JSTARS.2024.3370218.

[28] A. Jaafari, A. Najafi, H. R. Pourghasemi, J. Rezaeian, and A. Sattarian, "GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran," *Int. J. Environ. Sci. Technol.*, vol. 11, no. 4, pp. 909–926, May 2014, doi: 10.1007/s13762-013-0464-0.

[29] Z. Chang *et al.*, "Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 15, no. 5, pp. 1127–1143, May 2023, doi: 10.1016/j.jrmge.2022.07.009.

[30] G. Yan, S. Liang, and H. Zhao, "An Approach to Improving Slope Unit Division Using GIS Technique," *Scientia Geographica Sinica*, 2017, doi: 10.13249/j.cnki.sgs.2017.11.019.

[31] J. Lu *et al.*, "Ensemble learning landslide susceptibility assessment with optimized non-landslide samples selection," *Geomatics, Natural Hazards and Risk*, vol. 15, no. 1, p. 2378176, Dec. 2024, doi: 10.1080/19475705.2024.2378176.

[32] B. Gao *et al.*, "Landslide Risk Evaluation in Shenzhen Based on Stacking Ensemble Learning and InSAR," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 16, pp. 1–18, 2023, doi: 10.1109/JSTARS.2023.3291490.

[33] H. Chen *et al.*, "A landslide extraction method of channel attention mechanism U-Net network based on Sentinel-2A remote sensing images," *International Journal of Digital Earth*, vol. 16, no. 1, pp. 552–577, Oct. 2023, doi: 10.1080/17538947.2023.2177359.

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

[35] S. M. Lundberg *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nat Biomed Eng*, vol. 2, no. 10, pp. 749–760, Oct. 2018, doi: 10.1038/s41551-018-0304-0.

[36] 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. Top. Appl. Earth Observations Remote Sensing*, vol. 17, pp. 7475–7489, 2024, doi: 10.1109/JSTARS.2024.3379350.

[37] G. F. Jenks, "The Data Model Concept in Statistical mapping," *International Yearbook of Cartography*, pp. 186–190, 1967.

[38] 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," *International Journal of Applied Earth Observation and Geoinformation*, vol. 107, p. 102681, Mar. 2022, doi: 10.1016/j.jag.2022.102681.

[39] Z. Chang *et al.*, "An updating of landslide susceptibility prediction from the perspective of space and time," *Geoscience Frontiers*, vol. 14, no. 5, p. 101619, Sep. 2023, doi: 10.1016/j.gsf.2023.101619.

[40] J. Liu *et al.*, "Multi-kernel support vector machine considering sample optimization selection for analysis and evaluation of landslide disaster susceptibility," *Acta Geodaetica et Cartographica Sinica*, vol. 51, no. 10, pp. 2034–2045, 2022, doi: 10.11947/1.AGCS,2022.20220326.

[41] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[42] Y. Song *et al.*, "Landslide susceptibility assessment through multi-model stacking and meta-learning in Poyang County, China," *Geomatics, Natural Hazards and Risk*, vol. 15, no. 1, p. 2354499, Dec. 2024, doi: 10.1080/19475705.2024.2354499.

[43] T. Zeng, L. Wu, D. Peduto, T. Glade, Y. S. Hayakawa, and K. Yin, "Ensemble learning framework for landslide susceptibility mapping: Different basic classifier and ensemble strategy," *Geoscience Frontiers*, vol. 14, no. 6, p. 101645, Nov. 2023, doi: 10.1016/j.gsf.2023.101645.

[44] Y. Qiu and J. Zhou, "Short-Term Rockburst Damage Assessment in Burst-Prone Mines: An Explainable XGBOOST Hybrid Model with SCSO Algorithm," *Rock Mech Rock Eng*, vol. 56, no. 12, pp. 8745–8770, Dec. 2023, doi: 10.1007/s00603-023-03522-w.

[45] S. Zhao, F. Dai, J. Deng, H. Wen, H. Li, and F. Chen, "Insights into landslide development and susceptibility in extremely complex alpine geoenvironments along the western Sichuan–Tibet Engineering Corridor, China," *CATENA*, vol. 227, p. 107105, Jun. 2023, doi: 10.1016/j.catena.2023.107105.

[46] P. Arrogante-Funes, A. G. Bruzón, A. Álvarez-Ripado, F. Arrogante-Funes, F. Martín-González, and C. J. Novillo, "Assessment of the regeneration of landslides areas using unsupervised and supervised methods and explainable machine learning models," *Landslides*, Oct. 2023, doi: 10.1007/s10346-023-02154-z.