**ORIGINAL PAPER**

# Application of convolutional neural networks based on Bayesian optimization to landslide susceptibility mapping of transmission tower foundation

**Mansheng Lin[1] · Shuai Teng[1] · Gongfa Chen[1] · Bo Hu[1]**

## Abstract

The stability of tower foundation slopes is an important factor to maintain the operation of a power system. However, it is time-consuming and expensive to evaluate tower foundation slopes one by one due to the large area. The aim of this study is to investigate the performance of CNNs with different architectures and training options for transmission tower foundation landslide spatial prediction (LSP) by Bayesian optimization. Accordingly, fourteen influencing factors related to landslide evaluated by gain ratio technique are considered and 424 historical landslide locations in Luoding and Xinyi Counties (Guangdong Province, China) are randomly divided into 80% for training and 20% for testing the CNNs. The CNN performances are investigated by permutating and combining different numbers of convolutional layers, pooling layers and learning rate strategy. In 59 Bayesian optimized cases, three conclusions are drawn: (a) the CNNs yielded the best result with 3 convolution layers, (b) the CNN without a pooling layer performs best, and (c) a piece-wise decay learning rate strategy yields better performance. Meanwhile, the excellent performance of the CNN obtained by Bayesian optimization (CNN$^B$) has also been validated by comparisons with gravitational search optimization algorithm and other landslide spatial models, which indicates that CNN$^B$ can be applied to generate the susceptibility maps for locating transmission tower foundations in high landslide susceptibility zones and reducing the impact of landslides on power supply by taking measures in advance.

## Introduction

The landslide assessment of the transmission tower foundations is a critical issue to ensure the construction and stable operation for power systems (Zhou et al. 2022). However, landslide assessment for tower foundations is a complicated task that involves the sciences of geotechnics, geomorphology, hydrology, and statistics (Liu et al. 2020). Although the method of establishing a physical model to evaluate the stable state of a foundation slope is reliable (Jiang et al. 2018; Wang et al. 2019a), it is only suitable for the small research area or a single slope (Park et al. 2019). Since it is time-consuming and expensive (Lee et al. 2013), using the

physical models to evaluate slope stability widely is a challenge. As the result of LSP, the landslide susceptibility map (LSM) can solve these problems.

A LSM is a way of expressing the susceptibility indices of a landslide in a region. The generation model of the LSM takes the factors that lead to landslides as the input, and the spatial distribution of the susceptibility indices of a landslide as the output. Recently, many models for generating an LSM are proposed. For example, data statistical models (e.g., entropy data mining model (Chen et al. 2017), Dampster-Shafer model (Chen et al. 2016) and frequency ratio model (Kayastha 2015)) and machine learning models (neuro-fuzzy model (Oh and Pradhan 2011), artificial neural networks (ANNs) (Tian et al. 2018b), random forest (Sun et al. 2020) and decision tree (Guo et al. 2021)). These methods can help engineers to obtain an LSM. However, these methods cannot extract critical information from the landslide influencing factors, especially, the machine learning methods are prone to overfitting, making it difficult to improve the prediction accuracy (Wang et al.

✉ Gongfa Chen
gongfa.chen@gdut.edu.cn

1  School of Civil and Transportation Engineering, Guangdong
   University of Technology, Guangzhou 510006, China

2019b). Each of the previously mentioned models for LSM has its disadvantages and advantages, which are shown in Table 1.

To address the above problems, deep learning models have received more attention. Unlike the previously mentioned approaches, it is a non-linear tool that extracts the key features from the input data (Lecun et al. 2015). In many fields, the accuracy and objectivity of deep learning models can be achieved or exceed the level of human experts (Krizhevsky et al. 2017). Among deep learning models, convolutional neural networks (CNNs) have a remarkable ability to solve the problem of pattern recognition (Anwer et al. 2018). CNNs have been also applied to the prediction task of LSMs (Azarafza et al. 2021; Nikoobakht et al. 2022). However, it is time-consuming and impractical to use the enumeration method to try the combinations of model parameters in a specific range for generating the LSMs (Hakim et al. 2022). Hence, many existing studies are more focused on comparing the impact of different models on the accuracy (Sun et al. 2020).

A suitable CNN model plays a decisive role in the prediction result (Lee et al. 2018; Wang et al. 2019c). In order to find suitable CNN hyperparameters, Ezzat et al. (2021), for example, used the gravitational search optimization to find suitable CNN hyperparameters for diagnosing COVID-19 from chest X-ray images. Kim et al. (2020) applied the harmony search (HS) algorithm to improve performance of a 1D CNN in human respiration pattern recognition. Hakim et al. (2022) applied the CNN model that is obtained by the grey wolf optimizer to the LSM prediction. In fact, a few studies have shown the disadvantages of these methods. For example, the HS algorithm is not efficient enough in performing local search in numerical optimization and is sensitive to the parameter settings (Yi et al. 2015). The grey wolf optimizer is prone to fall into local optimums and has slow convergence speed (Xie et al. 2020). To overcome these drawbacks, Bayesian optimization algorithm is proposed in this study to rapidly search suitable CNN hyperparameters and training options in prediction of LSMs. Furthermore, this study explores the performance of CNNs with different frameworks in LSP and applies the LSP technique to the stability evaluation of tower foundation slopes.

Therefore, taking the Luoding and Xinyi counties (Guangdong province, China), which has experienced various landslide hazards in recent decades, as the study area, this paper combines the LSP technique with the big data of transmission tower locations and investigates the regularity of CNN performance on LSP by permutating and combining its key layers based on Bayesian optimization for the first time. And the LSM predicted by the optimized CNN is used to evaluate the stability of transmission tower foundation slopes on large scale and provide the suitable/unsuitable geological, geographical and environmental situations for the tower construction in the study area. Specifically, the goals of this paper mainly include the following: (a) determining the best CNN hyperparameters and training options using Bayesian optimization algorithm, (b) applying the optimized CNN to generate LSMs, (c) analyzing model performance, and (d) evaluating the stability of transmission tower foundations in the study area.

## Materials

### Study area

The study area is located in Luoding and Xinyi counties, which is a mountainous region of Guangdong Province, China, extending between 22° 11′ N and 22° 57′ N latitude and 110° 40′ E and 111° 52′ E longitude and covering an area of approximately 5408 km$^2$. The altitude of the study

**Table 1** Some advantages and disadvantages of models previously used for LSM

| Method | Advantage | Disadvantage |
|---|---|---|
| Entropy data mining model (Chen et al. 2017) | Ability to use variable quantitative and qualitative data | Limited problem-solving ability |
| Dampster-Shafer model (Chen et al. 2016) | Powerful visual display of variables and relationships between them in a system | Limited transferability in different study areas |
| Frequency ratio model (Kayastha 2015) | The calculation principle is simple, and the factors weight affecting the occurrence of events is stable | Limited transferability in different study areas and low accuracy |
| Neuro-fuzzy model (Oh and Pradhan 2011) | Automatic generalization ability to generate fuzzy rules and fuzzy sets to a predetermined precision level | Expert knowledge that needs to be preset in advance |
| ANNs (Tian et al. 2018b) | Ability to handle data regardless of the measurement scale | Difficulty of network training and getting stuck at local maximums |
| Random forest (Sun et al. 2020) | Ability to provide the importance of features after training and facilitate factor evaluation | Time-consuming calculations |
| Decision tree (Guo et al. 2021) | High accuracy for small-size data samples with high-dimensional feature space | The calculation costs grow exponentially as the problem gets bigger<br>Accumulation of layer errors |
| CNN with grey wolf optimizer (Hakim et al. 2022) | Strong robustness<br>Easy handling the high-dimensional input data | Prone to fall into local optimums and has slow convergence speed |

area is high in the middle. And the mountainous area exceeds 50%. According to data from the National Meteorological Center of China (http://data.cma.cn/), the average monthly rainfall in the study area from 1960 to 2019 is 136.43 mm. Usually, the rainy season is from April to September in every year, accounting for more than 75% of the annual rainfall.

The geological conditions of the study area can be viewed from three aspects: stratum rocks, igneous rocks and metamorphic plutonic intrusion rocks. The stratum is mainly composed of Jixian-Qingbaikou System (silica biotite schist, plagioclase, etc.), Quaternary (gravel, clay, silt, etc.) and Carboniferous (sandstone, shale, dolomite, limestone, etc.). Igneous rocks are mainly composed of plutonic acid igneous rocks (mixed granites). The metamorphic plutonic intrusion rocks are mainly granite.

### Landslide inventory map

Landslide inventory is a basic but essential tool for landslide hazard management, representing a fundamental base of knowledge on the spatial distribution of existing landslides (Tian et al. 2018a). The locations, scales, and directions of the historical landslides in this paper are mainly obtained from the report of the Guangdong Provincial Geological Environment Monitoring Station of the China Geological Environment Monitoring Institute (Li et al. 2012). Finally, the locations of landslides are collected and stored in the geographic information system (GIS) database. Therefore, a

total of 424 landslide locations are collected from the study area. Landslides in the study area can be divided into small-, medium-, and large-scale according to their volume. Most of landslides are medium-scaled (less than 1,000,000 m$^3$) and small-scaled (less than 100,000 m$^3$), accounting for more than 90% of the landslides. Furthermore, in order to allow the CNN models to learn the features of the non-landslide locations, 424 locations that are less prone to landslides are also randomly selected in the study area. Additionally, a total of 1658 transmission towers in the study area are found in https://www.openstreetmap.org/. And the landslide inventory map of the study area is shown in Fig. 1.

### Preparation of influencing factors

The selection of landslide influencing factors is particularly important for landslide susceptibility mapping. According to Reichenbach et al. (2018), a total of 596 factors are found to assess landslide susceptibility from 1983 to 2016, and the average number of factors used in each model is 9. Besides, the selected landslide influencing factors should be measurable, operable, uneven, complete, and non-redundant (Ayalew and Yamagishi 2005). In addition, some studies (Aditian et al. 2018; Chang et al. 2020; Huang et al. 2020b) have shown that the number of factors between 4 and 12 are suitable for LSP. Therefore, 14 factors (Table 2) are selected as the input of LSM prediction in the study area, including altitude, aspect, distance to faults (m), distance to rivers (m), distance to roads (m), land
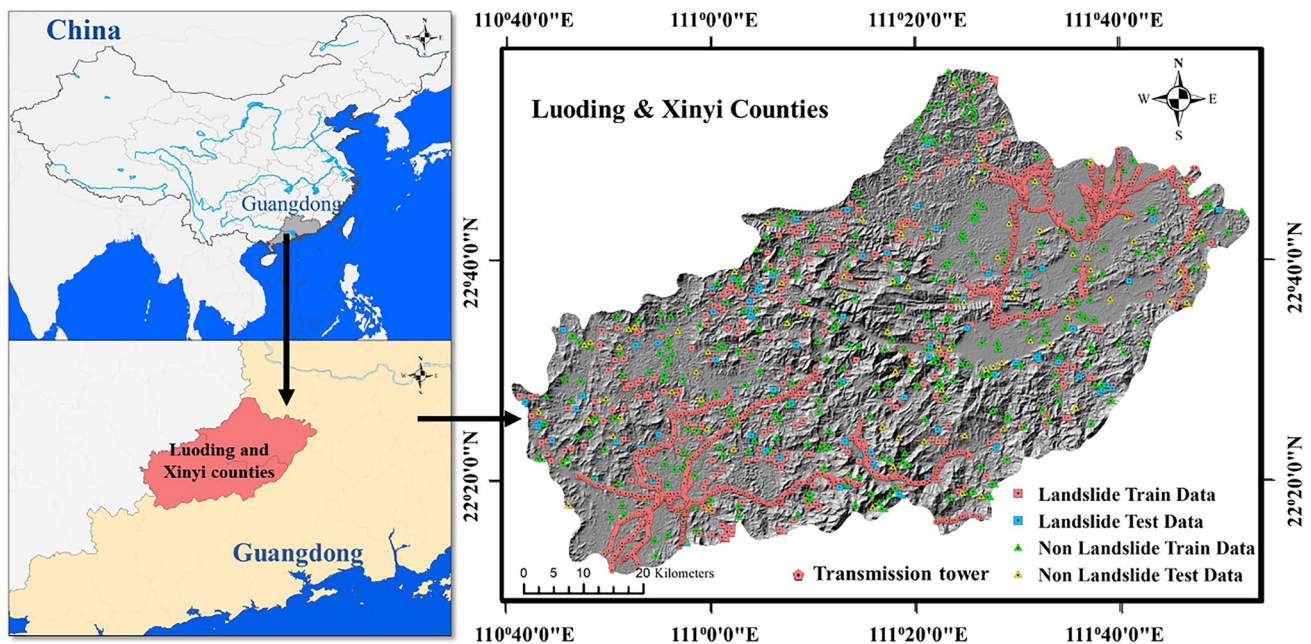


**Fig. 1** Landslide inventory map of Luoding and Xinyi counties

**Table 2** The information about landslide influencing factors used in this study

| No.# | Category | Factors | Reason for being selected | Data source |
|---|---|---|---|---|
| 1 | Topography | Altitude (Fig. 2a) | Slopes is closely related to the local altitude, so altitude is one of the factors that influence landslides (Guo et al. 2021). | The digital elevation model (DEM) with a 30 m resolution of the study area is downloaded from the website (http://www.gscloud.cn/home). |
| 2 | | Slope angle (Fig. 2n) | Slope angle directly affects slope stability and has been widely used in landslide sensitivity analysis (Wang et al. 2019b). | DEM derivatives. It is divided into 5 classes. |
| 3 | | Aspect (Fig. 2b) | Aspect is related to the landslides because slopes in different orientations are affected differently by precipitation and solar radiation (Pham et al. 2015). | DEM derivatives. |
| 4 | | Plan curvature (Fig. 2i) | It reflects the rate of change of the aspect along the contour and thus can affect the flow of water across a surface (Oh and Pradhan 2011). | DEM derivatives. Its values range from −4.43 to 5.54 and are divided into 5 classes. |
| 5 | | Profile curvature (Fig. 2j) | It influences the acceleration and deceleration of flow through slope, thus some valuable information about erosion and deposition is provided (Wu et al. 2020). | DEM derivatives. Its values range from −9.45 to 7.70 and are divided into 5 classes. |
| 6 | | SDS (Fig. 2m) | It is an index reflecting the degree of surface fluctuation and erosion intensity (Atkinson and Massari 1998). | $SDS = 1/\cos(slope)$. The SDS in this study area ranges from 1 to 2.37 and is divided into 5 classes. |
| 7 | | TWI (Fig. 2g) | TWI describes the topographic properties of hydrological processes, because both slope and local upslope contribution areas are considered (Moore et al. 1991). | $TWI = Ln(SCA/\tan\beta)$, where SCA is the specific catchment area $(m^2/m)$, and $\beta$ is the slope angle (degree) of the position. The TWI in this study area ranges from 2.28 to 29.86 and is divided into 5 classes. |
| 8 | Environmental | Land use (Fig. 2c) | Land use is a key factor in landslides and has an important influence on the stability of slopes because of vegetation cover (Pham et al. 2016a). | Land use data are from the research (Gong et al. 2019). |
| 9 | | Rainfall (Fig. 2k) | It is a key landslide inducing factor because it can affect the shear strength of the slope (Pham et al. 2016b). | The rainfall (mm/month) raster data in the study area, which are divided into 6 classes, is obtained by performing inverse distance weighting interpolation method (Bui and Hoang 2017) on 50 rain stations (http://data.cma.cn/) in the vicinity of the study area. |
| 10 | | NDVI (Fig. 2l) | It reflects the greenness of an area and may alter the distribution of soil and hydrological processes on slopes (Arabameri et al. 2020). | The NDVI data in this article are derived from the MODIS product at https://search.earthdata.nasa.gov/search. The obtained NDVI values range from −0.3 to 1 and are divided into 7 classes. |
| 11 | | Distance to rivers (Fig. 2f) | Rivers affect slope stability because they can cut and erode banks, and these actions reshape and sculpt the landscape. In addition, fluctuations in the water level greatly affect the groundwater level of the slope (Guo et al. 2021). | The data comes from National Geomatics Center of China (http://www.ngcc.cn/ngcc/), and the Euclidean distance tool of ArcGIS is used to obtain the river distance in the study area, and distance is divided into 5 classes. |
| 12 | | Distance to roads (Fig. 2h) | It is considered to be one of the most important human factors affecting the occurrence of landslides (Chen et al. 2017). | The method of obtaining road distance raster data is same as river distance. The continuous distance is divided into 5 classes. |
| 13 | Geological | Lithology (Fig. 2d) | The mechanical and hydrological properties of rock masses (such as permeability and friction angle) differ between lithological units, so this factor can greatly affect slope stability (Pham et al. 2016b). | Geological maps of the study area are obtained from the National 1:200,000 Digital Geological Map (Public Edition) of China (Li et al. 2019) and divided into 12 classes. |

**Table 2** (continued)

| No.# | Category | Factors | Reason for being selected | Data source |
|---|---|---|---|---|
| 14 | | Distance to faults (Fig. 2e) | It have an important influence on the distribution and scale of landslides in the study area (Pham et al. 2015). | The faults location data in the study area is obtained from the National 1:200,000 Digital Geological Map (Public Edition) Spatial Database of China (Li et al. 2019), and the Euclidean distance tool of ArcGIS is used to obtain the fault distance in the study area. The continuous distance is divided into 5 classes. |

use, lithology, normalized difference vegetation index (NDVI), plan curvature, profile curvature, rainfall (mm/month), surface roughness (standard deviation of the slope, SDS), slope (°), and topographic wetness index (TWI). Except the influencing factors of land use, lithology, and aspect, which are discrete variables, the rest are continuous variables. In order to standardize the input form, these landslide influencing factors need to be processed (Guo et al. 2021). Categorical variables are generally classified according to the heuristic classification of the related thematic information (Calvello and Ciurleo 2016). For numerical variables, it can be classified according to natural breaks method, which has been suggested for data with jumps (Calvello and Ciurleo 2016; Chen et al. 2017). Finally, the landslide influencing factor datasets are resampled to 30 m resolution raster (Fig. 2).

## Evaluation of influencing factors

The selection of features is important for the prediction of LSMs (Guo et al. 2021). Redundant features will interfere the recognition ability of a model, reduce the generalization ability, and increase the operation time (Zhou et al. 2021). In order to prove the validity of the selected landslide influencing factors or eliminate irrelevant factors to improve the predictive ability of the model, the gain ratio (GR) technique (Dash and Liu 1997) is adopted in this paper. When the GR of a factor is less than or equal to zero, it is considered irrelevant to the landslide and should not be used as the input of the model. The mathematical calculation of the GR is introduced in the following context.

Let $p_k$ be the probability of event occurring (either landslide or non-landslide) and the $n$ be the number of events, with the information entropy before splitting the factor into several classes is:

$$Ent(D) = - \sum_{k=1}^{n} p_k \log_2 p_k \qquad (1)$$

where $D$ is the total set of a factor (e.g., altitude, slope), and the $D_i$ ($i = 1, 2, \ldots, q$) is the subset of $D$ after splitting the factor. The $q$ is the number of classes (e.g., $q = 6$ in altitude, $q = 9$ in slope). Then, the information gain of each factor is:

$$Gain(D) = Ent(D) - \frac{1}{b} \sum_{i=1}^{q} a_i Ent(D_i) \qquad (2)$$

where $a_i$ is the number of samples (either landslide or non-landslide) in a class, and $b$ is the total number of samples. And the split information of a landslide influencing factor is computed as follows:

$$SplitInfo(D) = - \sum_{i=1}^{q} \frac{a_i}{b} \log_2 \frac{a_i}{b} \qquad (3)$$

Finally, the GR of the factor is calculated by the following formula:
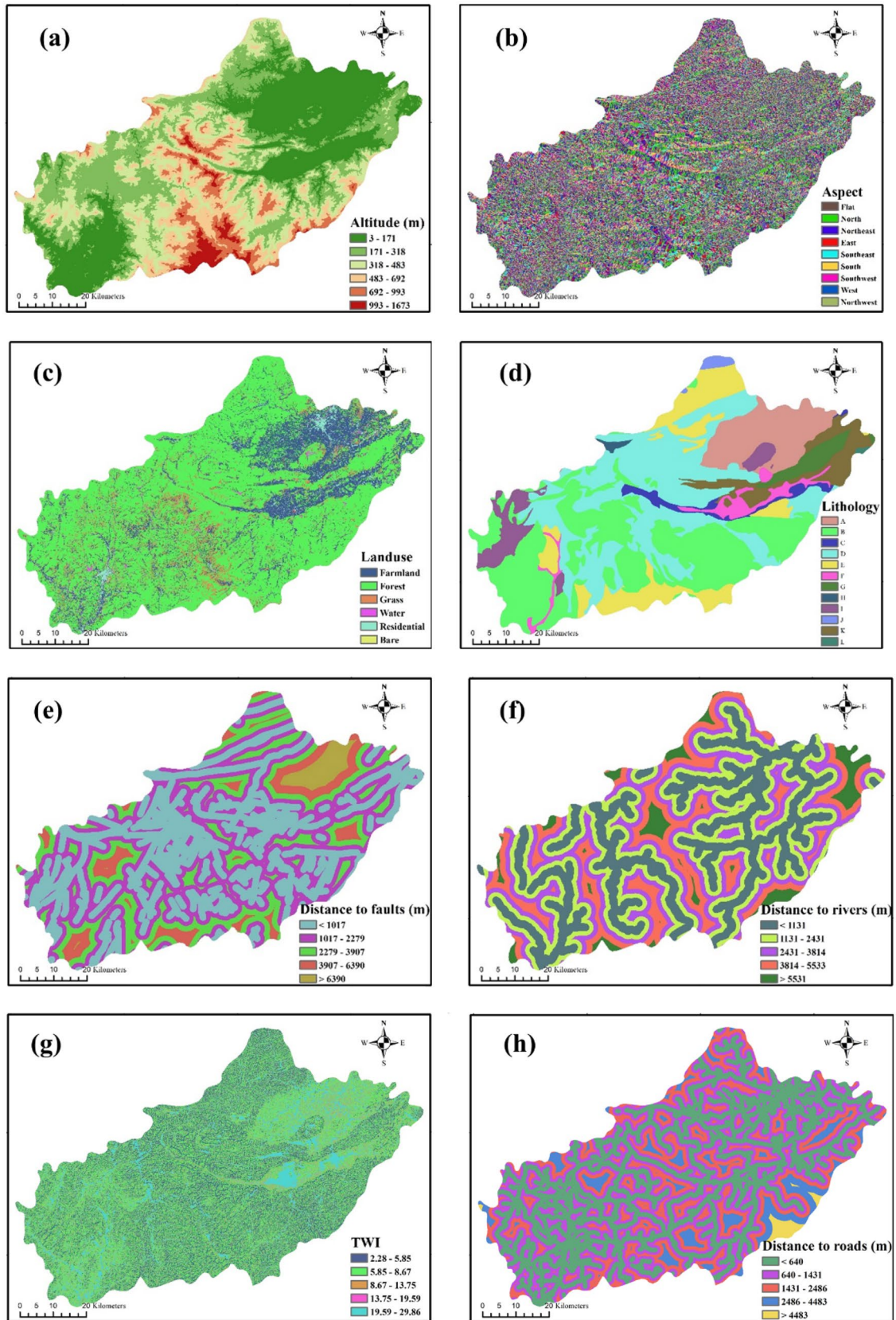
**Fig. 2** Thematic maps of the study area. **a** Altitude, **b** aspect, **c** land use, **d** lithology, **e** distance to faults, **f** distance to rivers, **g** TWI, **h** distance to roads, **i** plan curvature, **j** profile curvature, **k** rainfall, **l** NDVI, **m** SDS, **n** slope
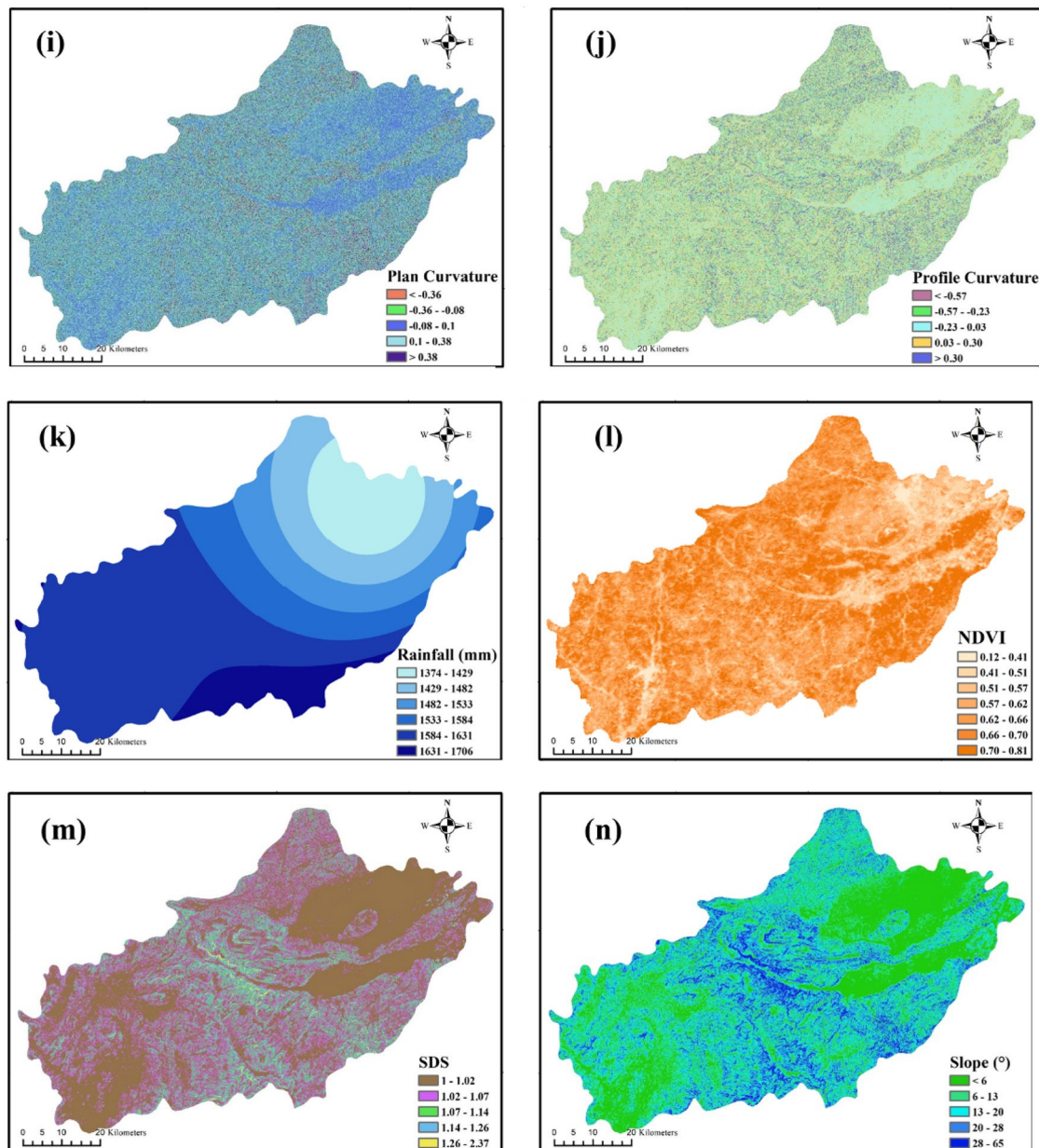
**Fig. 2** (continued)

$$GainRatio(D) = \frac{Gain(D)}{SplitInfo(D)} \tag{4}$$

## Methodology

### CNN

Convolutional neural networks (CNNs) are non-linear tools that can automatically extract key features from large amounts of data (Lin et al. 2022). They are often used as a prediction model in LSM tasks (Hakim et al. 2022; Thi-Ngo et al. 2021; Wang et al. 2019b). In LSP, the input of a CNN is the landslide influencing factors, and the output is the landslide (positive class) and non-landslide (negative class) labels.

### Optimization of CNN hyperparameters and training options

The hyperparameters of a CNN and training options (target variables) can significantly affect the prediction results, thus selecting the appropriate parameters is a challenge. In this paper, Bayesian optimization is chosen to obtain the

CNN hyperparameters and training parameters. When the objective function is not differentiable, Bayesian optimization algorithm has extraordinary performance in a short time. The optimization process mainly includes the following context.

The prerequisite is assumed that the relationship between the target variables and error rate of the validation data set belongs to the Gaussian process (Garrido-Merchán and Hernández-Lobato 2020), (a) then the original distribution is modified by the responses of the input (i.e., the target variable $X_i = \{x_{i1}, x_{i2}, \dots\}$) and output (i.e., the results $Y_i = \{y_{i1}, y_{i2}, \dots\}$). The mathematical expression can be represented by formula 6:

$$p(\theta|X, y) = \frac{p(y|X, \theta)p(\theta)}{p(y|X)} \tag{6}$$

where $p(\theta|X, y)$ is the unknown distribution (previous distribution), $p(\theta)$ is corrected distribution (posterior distribution), $p(y|X, \theta)$ is the likelihood estimate of the sample, and $p(y|X)$ is the probability distribution of the sample results.

(b) The points other than the previous sample points (i.e., $X_i$) will be selected as the input for the next iteration to obtain detailed distributions or explore unknown distribution areas. If the conditions (error rate of the validation set and the maximum number of optimization iterations) are not satisfied, the program will go back to step b, otherwise output the target variable $X_i$, which is the optimization solution.

### CNN architecture optimization

Furthermore, in order to explore the impact of increasing the number of convolutional layers and pooling layers in a CNN on the prediction performance, this paper uses a total of 5 convolution blocks (including a convolutional layer and relu layer) and pooling blocks (max pooling layer) to perform permutations (Fig. 3).

Table 3 summarizes these hyperparameters and their search spaces. The convolution blocks and the pooling blocks are arranged and combined, and a total of 59 cases are obtained. Meanwhile, in the comparison of learning rate strategies, the piecewise decay (PD) learning rate strategy will be optimized first. Then, five best CNN architectures selected among the optimization results will be trained with the constant learning rate strategy under the same other conditions (for example, the hyperparameters and their search space). Finally, the model with the maximum area under the curve (AUC) value is selected from the optimization results as the prediction model of the LSM in this paper.

### Accuracy measures

Receiver operating characteristic (ROC) curve is a standard technique for the performance evaluation of landslide prediction methods (Bradley 1997). It is produced by plotting the true positive (TP) rate against the false positive (FP) rate at various threshold values. Moreover, the AUC measure has been widely used to quantitatively evaluate the performance of LSM approaches (Reichenbach et al. 2018). Specifically, the AUC of a model is classified to 0.5–0.6, 0.6–0.7, 0.7–0.8, 0.8–0.9, and 0.9–1 to represent poor, average, good, very good, and excellent performance (Thi-Ngo et al. 2021).

In addition, the landslide frequency ratio (FR) can be used to evaluate the performance of the model even if the landslide susceptibility zone in a LSM are varied (Guo et al. 2021). The mathematical expression of the FR is shown as follows:

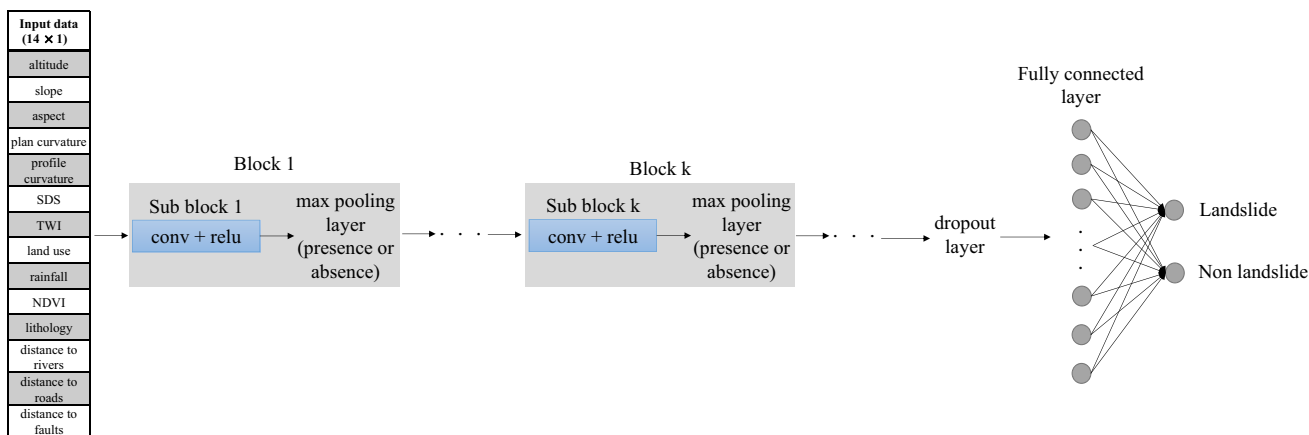$$FR = \frac{LA_i/TLA}{A_i/TA} \tag{7}$$



**Fig. 3** The CNN model before optimization, $k = [1, 5]$

**Table 3** Hyperparameters and their search space of CNN

| Method | Group | Parameters | Range | Type |
|---|---|---|---|---|
| Enumeration | CNN architecture | Block number | [1, 5] | Integer |
| | | Max pooling layer number | [0, 5] | Integer |
| Bayesian optimization | CNN hyperparameters | Convolution kernel size (height) | [1, 6] | Integer |
| | | Convolution kernel channel | [8, 16] | Integer |
| | | Max pool size | [2, 5] | Integer |
| | | Dropout rate | [0.1, 0.7] | Real |
| | Training options | Initial learning rate | [0.001, 1] | Real |
| | | Learn rate strategies | Piecewise decay, constant | String |
| | | Learn rate drop period | [1, 10] | Integer |
| | | Learn rate drop factor | [0.1 0.9] | Real |
| | | mini-batch size | [6, 30] | Integer |

The number of convolution kernel channels is the number of channels in the first convblock, and the number of convolution kernel channels in subsequent convblock is twice that of the previous one

*convblock* convolution block(s)

where $A_i$ and $LA_i$ are the area and landslide area in each susceptibility zone, while *TA* and *TLA* are the total area and total landslide area in the study area. The FR index also considers the relationship between the landslide area and the susceptibility zone in different grades, which indicates the reasonableness of a model to predict the susceptibility zone.

## The procedures of generating LSMs

The procedures of generating LSMs mainly contain following steps:

(a) The landslide inventory in the study area is produced, including the query of historical landslide data and the use of satellite remote sensing images for location proofreading.

(b) Landslide influencing factors of 424 landslide samples and 424 non-landslide samples are extracted by using ArcGIS (Environmental Systems Research Institute, Inc., Redlands, CA, USA). Then, 80% and 20% of samples are randomly assigned as the training set and testing set, respectively (Asadi et al. 2022; Nhu et al. 2020; Xing et al. 2021).

(c) Bayesian algorithm is used to optimize the CNN hyperparameters and training options. Meanwhile, the best-performing parameters will be selected from 59 cases, which are obtained by combining different numbers of convolutional layers and pooling layers. Then, the robustness of the best-performing CNN model is verified by fivefold cross-validation.

(d) The landslide influencing factors of the study area are extracted by using ArcGIS, then input into the best-performing CNN to obtain the landslide susceptibility

indices (probability). This step is performed in MATLAB (MathWorks Inc., Natick, MA, USA).

(e) Then, the LSM is generated according to the landslide susceptibility indices. And the natural breaks method (Chen et al. 2017) is used to classify indices as very low, low, moderate, high, and very high.

(f) Finally, the performances of the CNN models are evaluated by accuracy measures.

The overall flowchart is shown in Fig. 4.

## Results and discussion

### Relationships between landslides and influencing factors

The influencing degree of factors can be reflected by the evaluators, FR and GR. The FR reflects the landslide frequency in a class of each factor (Table 4). For instance, as shown in Table 4, the FR of slope angles from 28° to 65° is 4.19, indicating that most landslides in the study area occur on steep slopes. And FRs in SDS can also reflect similar conclusions. In addition, FRs roughly increase with a decreasing distance to faults, rivers and roads, and most landslides occur within distances less than 1017 m, 1131 m, and 640 m, respectively. In lithology, the H class (Xiaoyunwushan Formation, Triassic; the main components are conglomerate, glutenite intercalated with black siltstone, and lime shale) has the highest FR (3.42), indicating that this characteristic have a large impact on landslides. Then followed by class C (Liujiang Formation and Kuofeng Formation, Devonian; the main components are thin siliceous rock, siliceous mudstone,
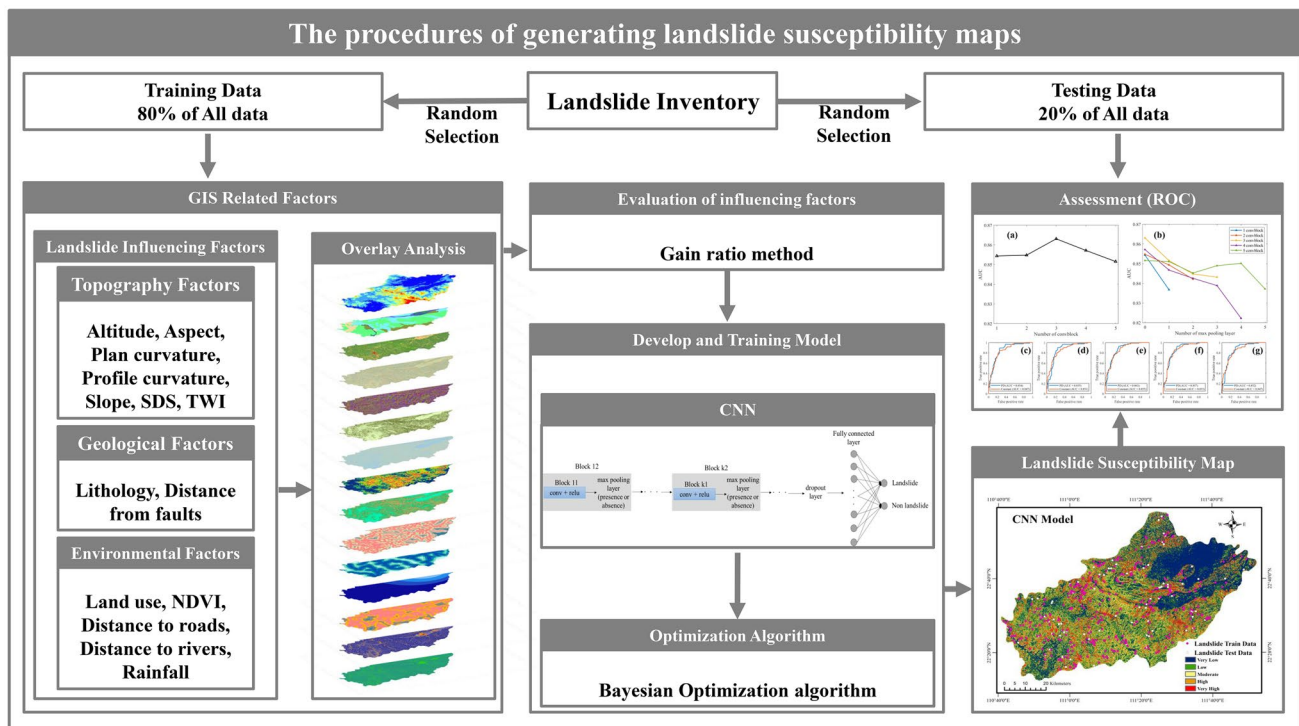
**Fig. 4** Illustration of the methodology in this study

dolomite), and class D (Pingzheng granulite formation, Fengtongkou formation, Lankeng formation and Shapingwan formation, Jixian System-Qingbaikou System; the main components are silicic biotite schist, plagioclase amphibole), and their FRs are 1.29 and 1.25, respectively.

GR reflects the relationship between influencing factors and landslides. Figure 5 shows that SDS and slope contribute the most to the landslide, and their GR are 0.113 and 0.103, respectively. The following factors such as distance to faults, TWI, land use, lithology, NDVI, distance to roads, altitude, and rainfall have GR values between 0.011 and 0.065. Meanwhile, GR values of some influencing factors such as distance to rivers, plan curvature, profile curvature, and aspect are less than 0.01 but positive, which means they do not decrease the model performance. An additional calculation also supports this point: the LSM without considering aspect, profile curve, or plan curve had an AUC value of 0.8446, 0.8602 and 0.8511, respectively, which is less than that of the map considering these factors (0.8612). In general, the topography factors (e.g., SDS and slope) are more important for landslides in this study area than the geological (e.g., lithology, distance from faults), hydrological (e.g., TWI) and environmental ones (e.g., NDVI, distance from rivers, rainfall, and land use). Additionally, GRs of all landslide influencing factors are greater than zeros, thus they all are considered to have positive impact on landslides.

## Optimization results and accuracy analysis of CNN architectures

The AUC results for the 59 cases obtained by combining different numbers of convolution blocks and pooling blocks are shown in Table 10. The range of AUC is between 0.8096 and 0.8631. Figure 6a shows that, in tasks (such as LSP) with few input features, the AUC first increases and then decreases with the increasing number of the convolution blocks for the case with no pooling layer in the CNN architecture. The AUC reaches the maximum value of 0.8631 when the number of convolution blocks is 3 (index 6).

In addition, the model performance decreases when pooling layers are used. It can be concluded from the Fig. 6b that the AUC decreases when the number of pooling layers is increased. It is worth noting that, the best AUC results appear in the CNN model without pooling blocks. The function of the pooling blocks is to extract key features from a large amount of information. However, for LSP, when the input dimension is small, adding pooling layers may lose key features, leading to the opposite of what is expected.

Figure 6d shows the comparison results of different learning rate strategies. AUC drops when the learning rate strategy is switched from PD to constant. This indicates that at the later stage of the model iteration, a small step size is beneficial to search for the smaller value in the loss function.

**Table 4** FRs of landslide influencing factors

| Factor | Class | Number of landslides | Percentage of landslides (%)[a] | Number of pixels in domain | Percentage of domain (%)[b] | FR[a/b] |
|---|---|---|---|---|---|---|
| Altitude | 3–171 | 94 | 22.17 | 1,822,781 | 30.21 | 0.73 |
| | 171–318 | 166 | 39.15 | 1,688,419 | 27.98 | 1.40 |
| | 318–483 | 113 | 26.65 | 1,277,917 | 21.18 | 1.26 |
| | 483–692 | 35 | 8.25 | 738,901 | 12.24 | 0.67 |
| | 692–993 | 14 | 3.30 | 377,711 | 6.26 | 0.53 |
| | 993–1673 | 2 | 0.47 | 128,623 | 2.13 | 0.22 |
| Aspect | Flat | 0 | 0.00 | 11,415 | 0.19 | 0.00 |
| | North | 65 | 15.33 | 817,750 | 13.55 | 1.13 |
| | Northeast | 42 | 9.91 | 734,758 | 12.18 | 0.81 |
| | East | 33 | 7.78 | 700,112 | 11.60 | 0.67 |
| | Southeast | 55 | 12.97 | 776,349 | 12.87 | 1.01 |
| | South | 63 | 14.86 | 783,090 | 12.98 | 1.14 |
| | Southwest | 43 | 10.14 | 721,850 | 11.96 | 0.85 |
| | West | 48 | 11.32 | 688,458 | 11.41 | 0.99 |
| | Northwest | 75 | 17.69 | 788,663 | 13.07 | 1.35 |
| Distance to faults (m) | <1017 | 233 | 54.95 | 2,542,194 | 42.13 | 1.30 |
| | 1017–2279 | 108 | 25.47 | 1,842,298 | 30.53 | 0.83 |
| | 2279–3907 | 61 | 14.39 | 1,042,670 | 17.28 | 0.83 |
| | 3907–6390 | 21 | 4.95 | 438,299 | 7.26 | 0.68 |
| | >6390 | 1 | 0.24 | 168,892 | 2.80 | 0.08 |
| Distance to rivers (m) | <1131 | 176 | 41.51 | 1,786,165 | 29.60 | 1.40 |
| | 1131–2431 | 108 | 25.47 | 1,637,921 | 27.14 | 0.94 |
| | 2431–3814 | 83 | 19.58 | 1,371,317 | 22.73 | 0.86 |
| | 3814–5533 | 50 | 11.79 | 925,341 | 15.33 | 0.77 |
| | >5533 | 7 | 1.65 | 313,609 | 5.20 | 0.32 |
| Distance to roads (m) | <640 | 239 | 56.37 | 2,616,770 | 43.36 | 1.30 |
| | 640–1431 | 131 | 30.90 | 1,906,915 | 31.60 | 0.98 |
| | 1431–2486 | 30 | 7.08 | 1,062,480 | 17.61 | 0.40 |
| | 2486–4483 | 21 | 4.95 | 386,407 | 6.40 | 0.77 |
| | >4483 | 3 | 0.71 | 61,781 | 1.02 | 0.69 |
| Land use | Farmland | 31 | 7.31 | 1,153,806 | 19.12 | 0.38 |
| | Forest | 342 | 80.66 | 4,131,636 | 68.47 | 1.18 |
| | Grass | 51 | 12.03 | 576,631 | 9.56 | 1.26 |
| | Water | 0 | 0.00 | 32,905 | 0.55 | 0.00 |
| | Residential | 0 | 0.00 | 138,151 | 2.29 | 0.00 |
| | Bare | 0 | 0.00 | 1223 | 0.02 | 0.00 |
| Lithology | A | 1 | 0.24 | 579,271 | 9.60 | 0.02 |
| | B | 148 | 34.91 | 1,903,079 | 31.54 | 1.11 |
| | C | 11 | 2.59 | 121,074 | 2.01 | 1.29 |
| | D | 154 | 36.32 | 1,754,678 | 29.08 | 1.25 |
| | E | 45 | 10.61 | 687,096 | 11.39 | 0.93 |
| | F | 1 | 0.24 | 184,184 | 3.05 | 0.08 |
| | G | 2 | 0.47 | 141,153 | 2.34 | 0.20 |
| | H | 5 | 1.18 | 20,832 | 0.35 | 3.42 |
| | I | 24 | 5.66 | 270,788 | 4.49 | 1.26 |
| | J | 3 | 0.71 | 36,772 | 0.61 | 1.16 |
| | K | 30 | 7.08 | 328,768 | 5.45 | 1.30 |
| | L | 0 | 0.00 | 6657 | 0.11 | 0.00 |

**Table 4** (continued)

| Factor | Class | Number of landslides | Percentage of landslides (%)[a] | Number of pixels in domain | Percentage of domain (%)[b] | FR[a/b] |
|---|---|---|---|---|---|---|
| NDVI | −0.38 | 56 | 13.21 | 926,757 | 15.36 | 0.86 |
| | 0.08–0.20 | 80 | 18.87 | 1,399,058 | 23.18 | 0.81 |
| | 0.20–0.33 | 65 | 15.33 | 910,094 | 15.08 | 1.02 |
| | 0.33–0.46 | 55 | 12.97 | 796,833 | 13.20 | 0.98 |
| | 0.46–0.59 | 60 | 14.15 | 791,611 | 13.12 | 1.08 |
| | 0.59–0.72 | 75 | 17.69 | 783,162 | 12.98 | 1.36 |
| | 0.72–1 | 33 | 7.78 | 426,837 | 7.07 | 1.10 |
| Plan curvature | <-0.36 | 20 | 4.72 | 313,704 | 5.20 | 0.91 |
| | −0.28--0.18 | 95 | 22.41 | 1,613,797 | 26.74 | 0.84 |
| | −0.18–0.10 | 129 | 30.42 | 2,201,743 | 36.49 | 0.83 |
| | 0.10–0.38 | 148 | 34.91 | 1,483,495 | 24.58 | 1.42 |
| | >0.38 | 32 | 7.55 | 421,613 | 6.99 | 1.08 |
| Profile curvature | <-0.57 | 4 | 0.94 | 134,496 | 2.23 | 0.42 |
| | −0.34--0.26 | 55 | 12.97 | 665,568 | 11.03 | 1.18 |
| | −0.26 – 0.03 | 135 | 31.84 | 2,336,350 | 38.72 | 0.82 |
| | 0.03–0.30 | 157 | 37.03 | 2,225,845 | 36.89 | 1.00 |
| | >0.30 | 73 | 17.22 | 672,093 | 11.14 | 1.55 |
| Rainfall (mm) | 1374–1429 | 18 | 4.25 | 652,107 | 10.81 | 0.39 |
| | 1429–1482 | 48 | 11.32 | 792,195 | 13.13 | 0.86 |
| | 1482–1533 | 49 | 11.56 | 753,551 | 12.49 | 0.93 |
| | 1533–1584 | 64 | 15.09 | 767,712 | 12.72 | 1.19 |
| | 1584–1631 | 216 | 50.94 | 2,623,185 | 43.47 | 1.17 |
| | 1631–1706 | 29 | 6.84 | 445,602 | 7.38 | 0.93 |
| SDS | 1–1.02 | 43 | 10.14 | 3,264,387 | 54.10 | 0.19 |
| | 1.02–1.07 | 123 | 29.01 | 1,675,033 | 27.76 | 1.04 |
| | 1.07–1.14 | 170 | 40.09 | 788,115 | 13.06 | 3.06 |
| | 1.14–1.26 | 82 | 19.34 | 251,362 | 4.17 | 4.63 |
| | 1.26–2.37 | 6 | 1.42 | 43,548 | 0.72 | 1.96 |
| Slope (°) | <6 | 15 | 3.54 | 1,708,702 | 28.32 | 0.12 |
| | 6–13 | 29 | 6.84 | 1,583,568 | 26.24 | 0.26 |
| | 13–20 | 89 | 20.99 | 1,413,807 | 23.43 | 0.89 |
| | 20–28 | 188 | 44.34 | 967,067 | 16.03 | 2.76 |
| | 28–65 | 103 | 24.29 | 349,301 | 5.79 | 4.19 |
| TWI | 2.28–5.85 | 201 | 47.41 | 1,523,078 | 25.24 | 1.87 |
| | 5.85–8.67 | 212 | 50.00 | 3,268,794 | 54.17 | 0.92 |
| | 8.67–13.75 | 3 | 0.71 | 418,028 | 6.93 | 0.10 |
| | 13.75–19.59 | 3 | 0.71 | 298,119 | 4.94 | 0.14 |
| | 19.59–29.86 | 5 | 1.18 | 514,426 | 8.52 | 0.14 |

After comparing the 59 cases, the optimal CNN model has 3 convolution blocks with no max pooling layer and a piecewise decay learn rate in Table 4. The optimal CNN architecture is shown in Fig. 7, and the hyperparameters and training options are shown in Table 5.

To elucidate the differences between optimization methods, the gravitational search algorithm is also applied to LSP.

Similarly, the results of the 59 cases are shown in Table 11. The AUC range is between 0.500 and 0.861, which is wider than that obtained by Bayesian optimization algorithm. It can be found that, in Fig. 6a, the AUCs obtained by the two algorithms are basically similar. For example, the AUCs increase first and then decreases with increasing the number of convolution blocks (Fig. 6a), and the maximum (0.8612) appears when

**Fig. 5** Gain ratio of each land-slide influencing factor
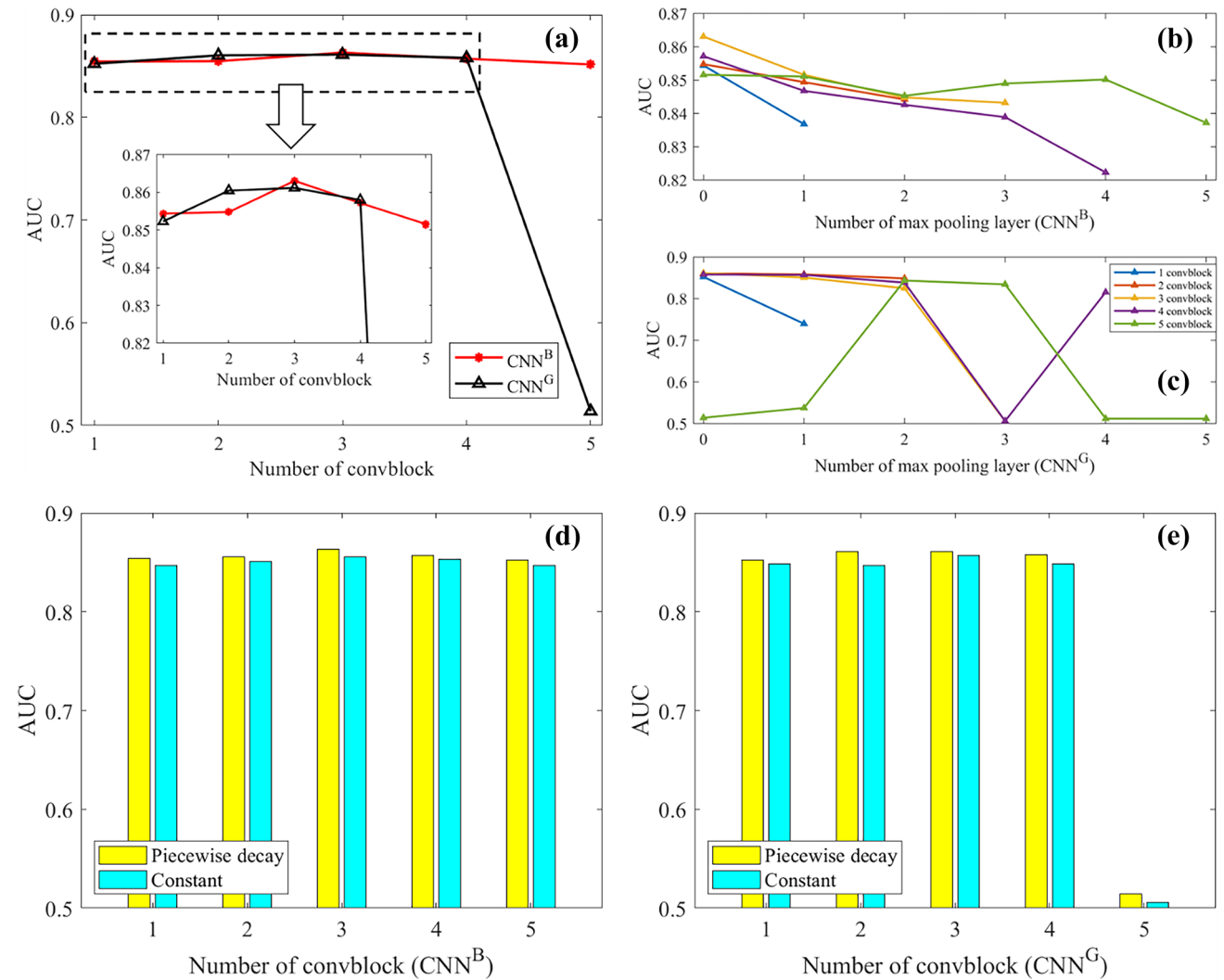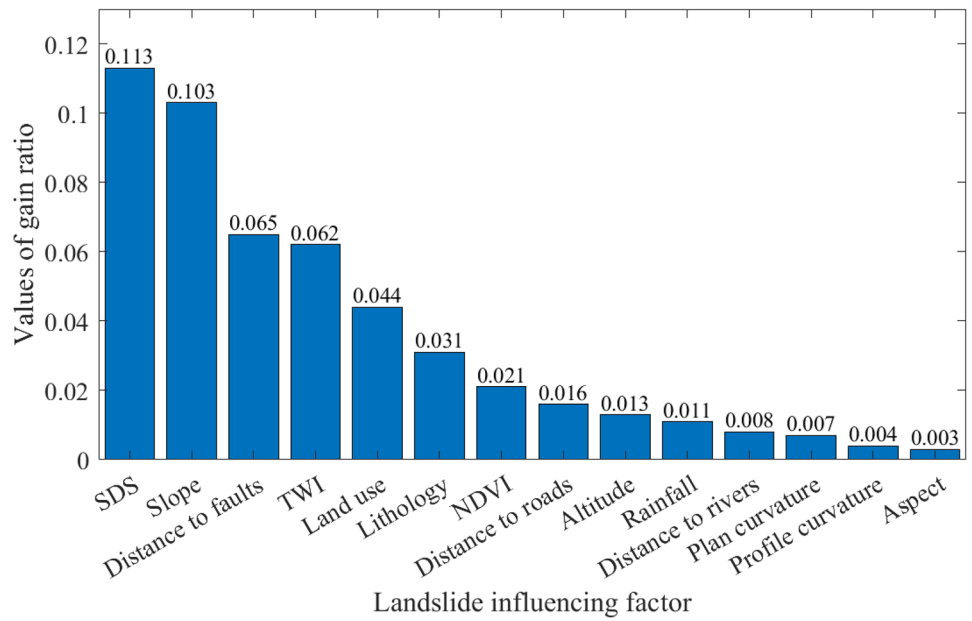




**Fig. 6** AUC Results from Bayesian optimization and gravitational search optimization algorithm

**Fig. 7** Selected CNN model by Bayesian optimization
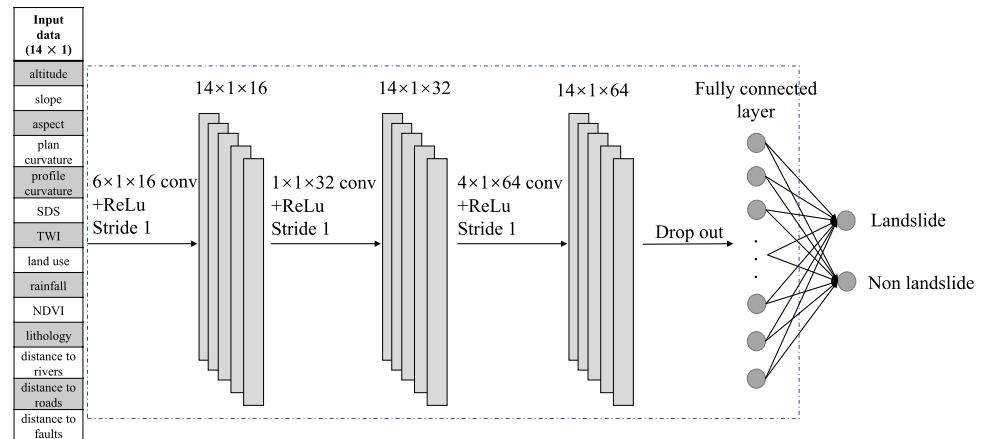


there are 3 convolution blocks; the AUCs decrease roughly as the number of pooling blocks increases (Fig. 6c); and they decrease when the learning rate strategy is switched from PD to constant (Fig. 6e). It is worth noting that, as shown in Table 11, when increasing the number of convblocks or pooling blocks, the AUC obtained by the gravitational search algorithm changes drastically. In addition, the CNN model is trained 10 times with 5 convolution blocks, no pooling blocks, and a PD learning rate. The results of the two optimization algorithms are shown in Fig. 8, which demonstrates that, compared with Bayesian optimization, the gravitational search algorithm makes results unstable when the number of target variables increases.

Comparisons show that Bayesian optimization algorithm is more stable than the gravitational search algorithm, and the corresponding AUC is higher. To demonstrate the robustness of the proposed CNN model, a fivefold cross-validation on the model of Fig. 7 is further carried out. Table 6 shows that, the average accuracy is 0.753 and the average AUC is 0.827 in testing dataset. AUC (Fig. 10) represents the division of positive (landslide)

and negative (non-landslide) probability thresholds for all cases and the average value of the fivefold cross-validation between 0.8 and 0.9 indicates that the model is very good. Besides, the classification accuracy rate (CAR) and AUC change in a small range and their distributions is relatively concentrated, which indirectly proves that the model is stable and not prone to be affected by the classification of dataset sets. The CNN model established by Bayesian optimization has excellent stability and reliability, which can be used for the analysis of landslide susceptibility in the research area.

## The application of CNN-based landslide susceptibility mapping methods

The LSMs of the study area are generated by using 14 landslide influencing factors, and the results are shown Fig. 11a

**Table 5** The best-selected parameters

| Group | Parameters | Result |
|---|---|---|
| CNN architecture | Block number | 3 |
| | Max pooling layer number | 0 |
| CNN hyperparameters | Convolution kernel size (height) | 6, 1, 4 |
| | Convolution kernel channel | 16, 32, 64 |
| | Max pool size | - |
| | Dropout rate | 0.2249 |
| Training options | Initial learning rate | 0.0059 |
| | Learn rate strategy | Piecewise decay |
| | Learn rate drop period | 2 |
| | Learn rate drop factor | 0.1462 |
| | Mini-batch size | 30 |
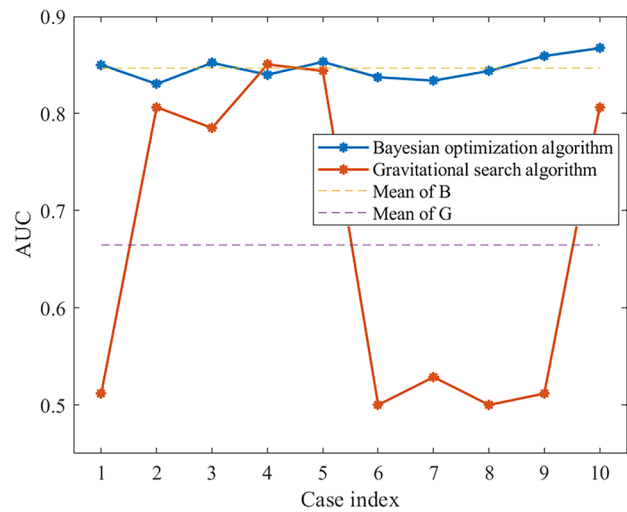


**Fig. 8** Performance of Bayesian optimization and the gravitational search optimization algorithm. Mean of B = mean AUC of Bayesian optimization, Mean of G = mean AUC of gravitational search optimization algorithm

**Table 6** The results of 5-fold cross-validation

| Order | TP | TN | FP | FN | CAR | AUC |
|---|---|---|---|---|---|---|
| 1 | 73 | 59 | 13 | 25 | 0.777 | 0.824 |
| 2 | 67 | 64 | 18 | 21 | 0.771 | 0.863 |
| 3 | 71 | 55 | 16 | 28 | 0.741 | 0.802 |
| 4 | 65 | 65 | 10 | 30 | 0.765 | 0.860 |
| 5 | 72 | 49 | 20 | 29 | 0.712 | 0.787 |
| Mean | 70 | 58 | 15 | 27 | 0.753 | 0.827 |

TP and TN are the correctly classified positive and negative classes, respectively. And FN and FP are the misclassified positive and negative classes, respectively. CAR = (TP + TN) / (TP + TN + FP + FN)

and b for the two different algorithms. The landslide susceptibility indices are divided into five classes through classification of each pixel using natural breaks method in ArcGIS to create five susceptibility classes: very low, low, moderate, high, and very high. Comparing the LSMs obtained by the two algorithms, the high and very high susceptibility areas are basically the same for the two algorithms. The obvious difference is that the low susceptibility landslide areas of the CNN optimized by the gravitational search algorithm (CNN$^G$) is larger than that of the CNN optimized by Bayesian optimization (CNN$^B$), and the very low susceptibility landslide areas of CNN$^G$ is small. From the spatial distribution of landslide susceptibility classes, the landslide influencing factors of slope and distance to faults are critical. On the one hand, the high and very high susceptibility landslide areas are mainly distributed in the central and western study area with steep slopes and close to the faults. On the other hand, low susceptibility landslide areas are mainly located in the central, northeast, and southwest study area with small slope angles and far from the fault. It is worth noting that, the importance of lithology for landslides is illustrated in Guo et al. (2021), which is consistent with the current results. As mentioned in Al-Najjar and Pradhan (2021), slope is an important factor in landslides, which is also consistent with the current results.

In order to better explain the rationality and capability of the optimized CNN model, two other machine learning models (back propagation neural network (BPNN) and support vector machine (SVM)) and a statistical model (entropy data mining model – Entropy model) are added for comparisons. The neuron number of the BPNN is determined by the formula $n = \sqrt{i + j} + k$ (where $i$ and $j$ are the number of input and output of the model, $k$ is an integer from 1 to 10) (Geng et al.

**Table 7** Hyperparameters search space and the best results of BPNN and SVM

| Model | Parameters | Range | Type | Best result |
|---|---|---|---|---|
| BPNN | Neuron number | [4, 24] | Integer | 12 |
| SVM | Suitable box constraint | $[10^{-3}, 10^3]$ | Real | 972.83 |
| | Kernel scale | $[10^{-3}, 10^3]$ | Real | 98.822 |

2018). Finally, the maximum AUC is found when $n = 12$. The SVM model uses Gaussian kernel function (Aktas and San 2019). And the search spaces of suitable box constraint and kernel scale are from $10^{-3}$ to $10^3$. Finally, the maximum AUC is optimized by Bayesian algorithm when box constraint = 972.83 and kernel scale = 98.822, which are shown in Table 7. And the LSMs for different models are shown in Fig. 11.
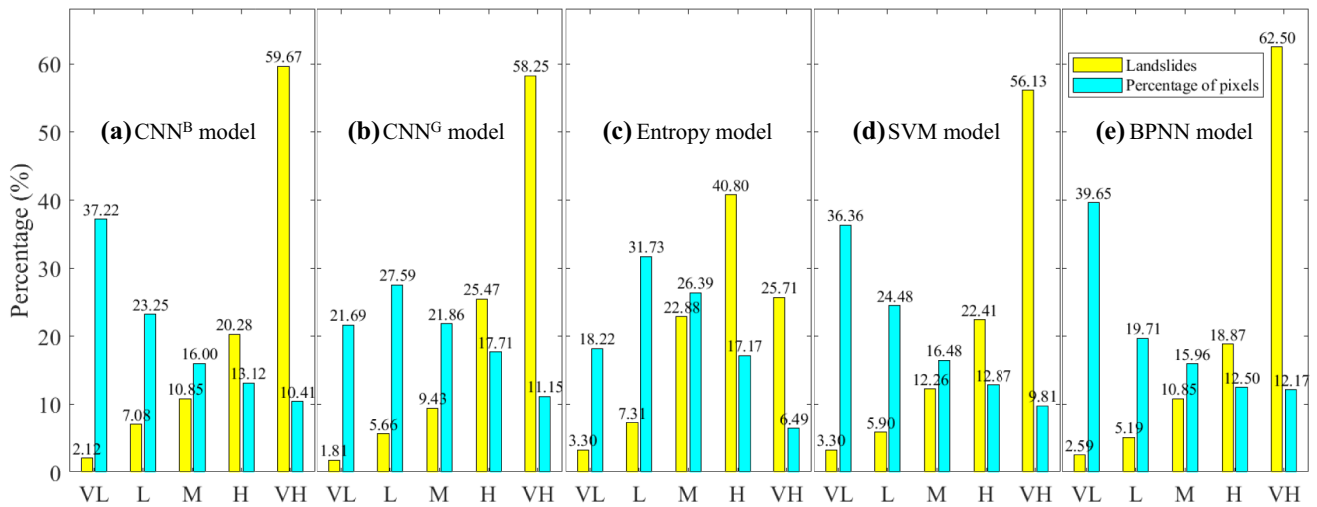
The LSMs obtained by the five models (CNN$^B$, CNN$^G$, BPNN, SVM, and Entropy model) are similar, but also different. For example, the high and very high landslide susceptibility areas of LSMs are all distributed in the middle of the study area, especially along the canyon in the middle area. Meanwhile, low and very low susceptibility areas are in the northeast and southwest of the study area. The LSM of the Entropy model is different from those of other models. It predicts the higher-elevation central and southern regions as high susceptibility areas, and the west as low susceptibility areas.

The distribution of susceptibility zones and the corresponding percentage of landslides in each level are computed and are shown in Fig. 9. The CNN models exhibit superior capabilities. In general, an ideal LSM for transmission tower foundations should contain fewer high and very high susceptibility areas. From this viewpoint, the machine learning methods have better performance than the statistical model. However, in the LSM generated by the Entropy model, less landslides occur in the very high susceptibility areas. Moreover, there are more than 80% of historical landslides occurred in the high and very high landslide susceptibility zones that generated by the CNN-based model, indicating that the CNN outperforms the other models. This conclusion is also confirmed in the FR results (Table 8). In the optimized CNN models by different optimization methods, the FR of CNN$^B$ in the high and very high susceptibility areas are higher than that of CNN$^G$. This discrepancy indicates that CNN$^B$ is more efficient than CNN$^G$.

In addition, Table 8 shows that the FRs of the five models in the very low and low susceptibility areas are all less than 1. The FRs increase in moderate susceptibility areas. And the FR is greater than 1 in the high and very high susceptibility areas. Especially in the very high susceptibility area, the FR of all five models reaches the maximum value. It

**Fig. 9** The classification results considering different methods. $CNN^B$ = obtained by Bayesian optimization, $CNN^G$ = obtained by the gravitational search optimization

is 5.730 ($CNN^B$), 5.225 ($CNN^G$), 3.961 (Entropy model), 5.723 (SVM), and 5.134 (BPNN), respectively. This reveals that the landslide distribution became gradually denser with increasing landslide susceptibility levels.

The $CNN^B$ model exhibit superior capabilities. It has the largest FR (5.730) in very high susceptibility areas and smallest FR (0.057) in the very low susceptibility areas. Meanwhile, in the ROC curve (Fig. 10), $CNN^B$ achieves the highest AUC (0.863), while the AUCs of the other models are 0.861 ($CNN^G$), 0.798 (Entropy model), 0.844 (SVM) and 0.861 (BPNN), respectively. In general, among the abovementioned machine learning and statistical models, the machine learning model is better than the statistical model. At the same time, the $CNN^B$ model is best in various evaluation indicators, thus it can be a powerful tool to mapping landslide susceptibility at a regional scale.

For the distribution of transmission towers, as shown in Fig. 11, comparatively most of towers (about 72%) are distributed in the very low and low landslide susceptibility areas, while about 14% towers are located in the high

and very high susceptibility areas. Most of unstable tower foundation are distributed in the canyon in the south of study area, mainly because of the heavy rainfall (> 1500 mm/year), steep slope (> 30°) and close to the road. The suggestions for the transmission tower construction are listed in Table 9. Avoiding these unsuitable combinations of factors that are accountable for landslides is conducive to ensuring the safety of the power supply system.

## Findings and limitations of this study

There are some researches generating the LSMs by using CNN models (Hakim et al. 2022; Sameen et al. 2020; Wang

**Table 8** FR values in each susceptibility level

| Model | Susceptibility level | | | | |
|---|---|---|---|---|---|
| | Very low | Low | Moderate | High | Very high |
| $CNN^B$ | 0.057 | 0.304 | 0.678 | 1.546 | 5.730 |
| $CNN^G$ | 0.054 | 0.205 | 0.432 | 1.438 | 5.225 |
| Entropy | 0.181 | 0.230 | 0.867 | 2.376 | 3.961 |
| SVM | 0.091 | 0.241 | 0.744 | 1.741 | 5.723 |
| BPNN | 0.065 | 0.263 | 0.680 | 1.509 | 5.134 |

$CNN^B$ obtained by Bayesian optimization, $CNN^G$ obtained by Gravitational search optimization
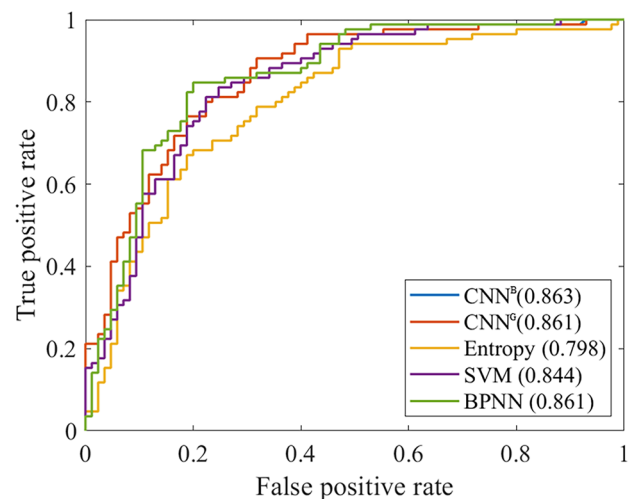


**Fig. 10** ROC curves of different models
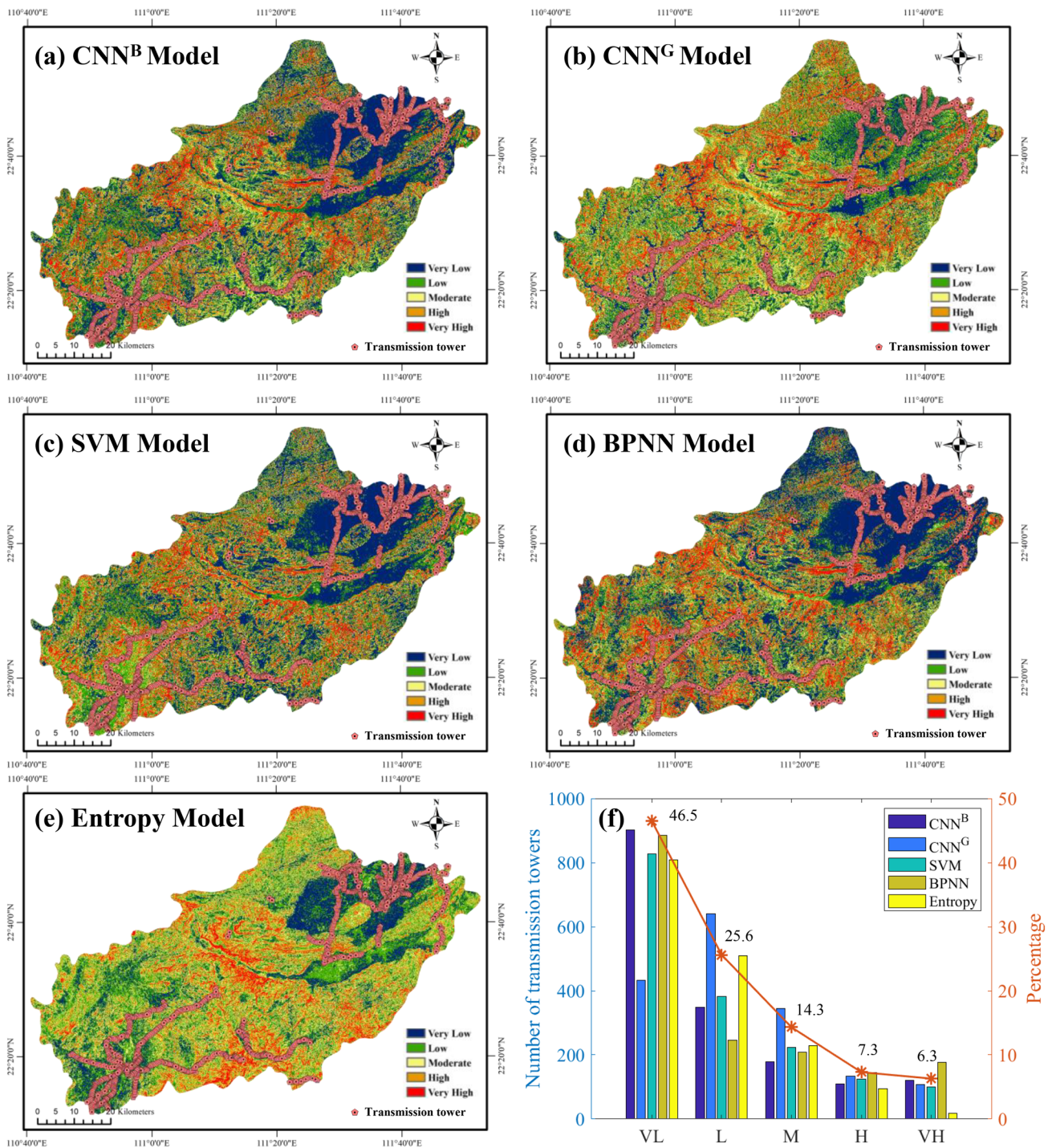
**Fig. 11** **a–e** Distribution of transmission tower in LSM predicted by each model and **f** the percentage of transmission towers in each susceptibility zone.

et al. 2019b; Yi et al. 2020). Although, the remarkable conclusions and practical application examples are summarized by these researches, however, there are no further investigations on framework regularity of model performance, which

should be much beneficial to understand the CNN architectures. The method based on the LSMs obtained by the CNNs with Bayesian optimization provides a new idea for the evaluation of transmission tower foundation stability.

**Table 9** Suggestions for the transmission tower construction

| #No. | Category | Factors | Suitable combinations | Unsuitable combinations |
|---|---|---|---|---|
| 1 | Topography | Altitude (m) | 171–318 | 3–171 |
| 2 | | Slope (°) | 6–20 | 20–28 |
| 3 | | Aspect | Northeast | Northwest |
| 4 | | Plan curvature | −0.08–0.1 | 0.10–0.38 |
| 5 | | Profile curvature | −0.23–0.03 | 0.03–0.30 |
| 6 | | SDS | 1.26–2.37 | 1.07–1.14 |
| 7 | | TWI | 8.67–13.75 | 5.85–8.67 |
| 8 | Environmental | Land use | Frost | Farmland |
| 9 | | Rainfall (mm/y) | 1374–1429 | 1584–1631 |
| 10 | | NDVI | 0.08–0.20 | 0.08–0.20 |
| 11 | | Distance to rivers (m) | 3814–5533 | < 1131 |
| 12 | | Distance to roads (m) | 1431–2486 | < 640 |
| 13 | Geological | Lithology | A (JxQb, Jixian-Qingbaikou System (silica biotite schist)) | D ($K_1$, lower cretaceous (siltstone)) |
| 14 | | Distance to faults (m) | 3907–6390 | < 1017 |

One of the main limitations of this study is that the sample points are very located in the test site, which is common in most existing studies (Chang et al. 2022; Huang et al. 2022, 2020a). It is a challenge to collect these landslide inventory data no matter from an on-site field survey or a search from remote sensing images and historical data because a lot of professional knowledge will be needed (Zhu et al. 2020). Some studies (Aditian et al. 2018; Huang et al. 2019) converted the landslides to pixels (for example 30 m) of landslide cells to increase the number of samples. However, the above operation requires the landslide influencing factors with higher spatial resolution to provide richer spatial information around the landslide (Yi et al. 2020). And the determination of the optimal spatial resolution factors such as DEM and its derivatives requires a lot of experiments (Merghadi et al. 2020). Meanwhile, the generative adversarial networks (GANs) are also used to increase the number of landslide samples (Al-Najjar and Pradhan 2021), however, this method does not always improve the model performance due to overfitting. Overall, the impact of these sample generation methods on model performance has not been explored in the same region. Thus, their applicability for LSM deserves further investigation. The Bayesian optimization-based CNNs used in LSMs can also be applied to analyze the susceptibility of other geohazards, for example, floods (Panahi et al. 2021; Tien Bui et al. 2020), land subsidence (Hakim et al. 2020) and multi-hazard (Ullah et al. 2022). As the landslide prediction model in this study is only limited to a deep learning framework, future researches on hybrid deep learning methods (e.g., hybrid deep learning frameworks, hybrid deep learning-machine learning frameworks) may improve the reliability and accuracy of LSMs.

# Concluding remarks

Two well-known CNN hyperparameter optimization algorithms, namely Bayesian optimization and the gravitational search algorithm, are applied to generate the LSM of the transmission tower foundation slopes in Xinyi and Luoding counties. Overall, three regularities are found:

(a) The CNNs produce the best result with 3 convolution layers.
(b) The best performance is achieved without using pooling layers.
(c) A piecewise decay learning rate strategy yields better performance.

Additionally, Bayesian optimization algorithm is more efficient and stable than the gravitational search algorithm. The LSM results indicate that there are 72%, 14%, and 14% of transmission towers are located in low and very low, moderate, and high and very high landslide susceptibility zones, respectively, among which, those in high and very high landslide susceptibility zones need more attention.

# Appendix

**Table 10** The AUC of 59 cases obtained by Bayesian optimization

| Situation | Index | MPLN | AUC | Situation | Index | MPLN | AUC | Situation | Index | MPLN | AUC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 convblock | 1 | 0 | 0.8544 | 4 convblock | 11–3 | 1 | 0.8428 | 5 convblock | 17–2 | 2 | 0.8453 |
| | 1–1 | 0 | 0.8472 | | 11–4 | 1 | 0.8426 | | 17–3 | 2 | 0.8383 |
| | 2 | 1 | 0.8368 | | 12 | 2 | 0.8313 | | 17–4 | 2 | 0.8354 |
| 2 convblock | 3 | 0 | 0.8548 | | 12–2 | 2 | 0.8257 | | 17–5 | 2 | 0.8374 |
| | 3–1 | 0 | 0.8515 | | 12–3 | 2 | 0.8382 | | 17–6 | 2 | 0.8410 |
| | 4 | 1 | 0.8494 | | 12–4 | 2 | 0.8426 | | 17–7 | 2 | 0.8364 |
| | 4–2 | 1 | 0.8457 | | 12–5 | 2 | 0.8338 | | 17–8 | 2 | 0.8404 |
| | 5 | 2 | 0.8442 | | 12–6 | 2 | 0.8274 | | 17–9 | 2 | 0.8363 |
| 3 convblock | 6 | 0 | 0.8631 | | 13 | 3 | 0.8241 | | 17–10 | 2 | 0.8364 |
| | 6–1 | 0 | 0.8551 | | 13–2 | 3 | 0.8293 | | 18 | 3 | 0.8280 |
| | 7 | 1 | 0.8457 | | 13–3 | 3 | 0.8389 | | 18–2 | 3 | 0.8455 |
| | 7–2 | 1 | 0.8516 | | 14 | 4 | 0.8223 | | 18–3 | 3 | 0.8490 |
| | 7–3 | 1 | 0.8418 | 5 convblock | 15 | 0 | 0.8516 | | 18–4 | 3 | 0.8361 |
| | 8 | 2 | 0.8432 | | 15–1 | 0 | 0.8469 | | 18–5 | 3 | 0.8400 |
| | 8–2 | 2 | 0.8448 | | 16 | 1 | 0.8493 | | 18–6 | 3 | 0.8389 |
| | 9 | 3 | 0.8432 | | 16–2 | 1 | 0.8511 | | 19 | 4 | 0.8486 |
| 4 convblock | 10 | 0 | 0.8572 | | 16–3 | 1 | 0.8418 | | 19–2 | 4 | 0.8410 |
| | 10–1 | 0 | 0.8531 | | 16–4 | 1 | 0.8421 | | 19–3 | 4 | 0.8502 |
| | 11 | 1 | 0.8468 | | 16–5 | 1 | 0.8404 | | 20 | 5 | 0.8372 |
| | 11–2 | 1 | 0.8386 | | 17 | 2 | 0.8096 | | | | |

Bold are the best in the same number of convblock

*convblock* convolution block, *MPLN* max pooling layer number

**Table 11** The AUC of 59 cases obtained by the gravitational search optimization algorithm

| Situation | Index | MPLN | AUC | Situation | Index | MPLN | AUC | Situation | Index | MPLN | AUC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1 convblock** | 1 | 0 | **0.8523** | 4 convblock | 11–3 | 1 | 0.5059 | 5 convblock | 17–2 | 2 | 0.5059 |
| | 1–1 | 0 | 0.8484 | | 11–4 | 1 | 0.5059 | | 17–3 | 2 | 0.5291 |
| | 2 | 1 | 0.7397 | | 12 | 2 | 0.5059 | | 17–4 | 2 | 0.8331 |
| **2 convblock** | 3 | 0 | **0.8605** | | 12–2 | 2 | 0.6580 | | 17–5 | 2 | 0.5096 |
| | 3–1 | 0 | 0.8472 | | 12–3 | 2 | 0.8386 | | 17–6 | 2 | **0.8436** |
| | 4 | 1 | 0.5116 | | 12–4 | 2 | 0.5059 | | 17–7 | 2 | 0.5243 |
| | 4–2 | 1 | 0.8584 | | 12–5 | 2 | 0.8269 | | 17–8 | 2 | 0.5059 |
| | 5 | 2 | 0.8486 | | 12–6 | 2 | 0.8259 | | 17–9 | 2 | 0.5059 |
| **3 convblock** | 6 | 0 | **0.8612** | | 13 | 3 | 0.5000 | | 17–10 | 2 | 0.5059 |
| | 6–1 | 0 | 0.8567 | | 13–2 | 3 | 0.5060 | | 18 | 3 | 0.6473 |
| | 7 | 1 | 0.5000 | | 13–3 | 3 | 0.5000 | | 18–2 | 3 | 0.5000 |
| | 7–2 | 1 | 0.8507 | | 14 | 4 | 0.8156 | | 18–3 | 3 | 0.8343 |
| | 7–3 | 1 | 0.5000 | 5 convblock | 15 | 0 | 0.5138 | | 18–4 | 3 | 0.5120 |
| | 8 | 2 | 0.5059 | | 15–1 | 0 | 0.5059 | | 18–5 | 3 | 0.5235 |
| | 8–2 | 2 | 0.8249 | | 16 | 1 | 0.5374 | | 18–6 | 3 | 0.5000 |
| | 9 | 3 | 0.5059 | | 16–2 | 1 | 0.5118 | | 19 | 4 | 0.5000 |
| **4 convblock** | 10 | 0 | **0.8580** | | 16–3 | 1 | 0.5059 | | 19–2 | 4 | 0.5000 |
| | 10–1 | 0 | 0.8484 | | 16–4 | 1 | 0.5291 | | 19–3 | 4 | 0.5118 |
| | 11 | 1 | 0.8570 | | 16–5 | 1 | 0.5176 | | 20 | 5 | 0.5117 |
| | 11–2 | 1 | 0.8325 | | 17 | 2 | 0.5000 | | | | |

Bold are the best in the same number of convblocks

*convblock* convolution block, *MPLN* max pooling layer number

**Declarations**

**Conflict of interest** The authors declare no competing interests.

# References

Aditian A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology 318:101–111

Aktas H, San BT (2019) Landslide susceptibility mapping using an automatic sampling algorithm based on two level random sampling. Comput Geosci 133:104329

Al-Najjar HAH, Pradhan B (2021) Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. Geosci Front 12(2):625–637

Anwer RM, Khan FS, Weijer JVD, Molinier M, Laaksonen J (2018) Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. ISPRS J Photogramm Remote Sens 138:74–85

Arabameri A, Chen W, Loche M, Zhao X, Li Y, Lombardo L, Cerda A, Pradhan B, Bui DT (2020) Comparison of machine learning models for gully erosion susceptibility mapping. Geosci Front 11(5):1609–1620

Asadi M, Mokhtari LG, Shirzadi A, Shahabi H, Bahrami S (2022) A comparison study on the quantitative statistical methods for spatial prediction of shallow landslides (case study: Yozidar-Degaga Route in Kurdistan Province, Iran). Environ Earth Sci 81(2):51

Atkinson PM, Massari R (1998) Generalised linear modelling of susceptibility to landsliding in the central apennines, Italy. Comput Geosci 24:373–385

Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains Central, Japan. Geomorphology 65(1–2):15–31

Azarafza M, Azarafza M, Akgun H, Atkinson PM, Derakhshani R (2021) Deep learning-based landslide susceptibility mapping. Sci Rep 11(1):24112

Bradley P (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. P Bradley 30(7):1145–1159

Bui DT, Hoang ND (2017) A Bayesian framework based on a Gaussian mixture model and radial-basis-function Fisher discriminant analysis (BayGmmKda V1.1) for spatial prediction of floods. Geosci Model Dev 10(9):3391–3409

Calvello M, Ciurleo M (2016) Optimal use of thematic maps for landslide susceptibility assessment by means of statistical analyses: case study of shallow landslides in fine grained soils. Paper presented at the Proceedings of the 12th International Symposium on Landslides, Napoli, Italy

Chang Z, Catani F, Huang F, Liu G, Meena SR, Huang J, Zhou C (2022) Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. J Rock Mech Geotech Eng

Chang ZL, Du Z, Zhang F, Huang FM, Chen JW, Li WB, Guo ZZ (2020) Landslide susceptibility prediction based on remote sensing images and GIS: comparisons of supervised and unsupervised machine learning models. Remote Sens 12(3):502

Chen W, Pourghasemi HR, Naghibi SA (2017) A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. Bull Eng Geol Env 77(2):647–664

Chen W, Pourghasemi HR, Zhao Z (2016) A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural

network models for landslide susceptibility mapping. Geocarto Int 32(4):367–385

Dash M, Liu H (1997) Feature Selection for Classification. Intell Data Anal 1(3):131–156

Ezzat D, Hassanien AE, Ella HA (2021) An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization. Appl Soft Comput 98:106742

Garrido-Merchán EC, Hernández-Lobato D (2020) Dealing with categorical and integer-valued variables in Bayesian optimization with Gaussian processes. Neurocomputing 380:20–35

Geng XY, Lu SZ, Jiang MS, Sui QM, Lv SS, Xiao H, Jia YX, Jia L (2018) Research on FBG-Based CFRP structural damage identification using BP neural network. Photonic Sensors 8(2):168–175

Gong P, Liu H, Zhang MN, Li CC, Wang J, Huang HB, Clinton N, Ji LY, Li WY, Bai YQ, Chen B, Xu B, Zhu ZL, Yuan C, Ping Suen H, Guo J, Xu N, Li WJ, Zhao YY, Yang J, Yu CQ, Wang X, Fu HH, Yu L, Dronova I, Hui FM, Cheng X, Shi XL, Xiao FJ, Liu QF, Song LC (2019) Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Sci Bull 64(6):370–373

Guo ZZ, Shi Y, Huang FM, Fan XM, Huang JS (2021) Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. Geosci Front 12(6):101249

Hakim W, Achmad A, Lee C-W (2020) Land subsidence susceptibility mapping in Jakarta using functional and meta-ensemble machine learning algorithm based on time-series InSAR data. Remote Sens 12(21)

Hakim WL, Rezaie F, Nur AS, Panahi M, Khosravi K, Lee CW, Lee S (2022) Convolutional neural network (CNN) with metaheuristic optimization algorithms for landslide susceptibility mapping in Icheon, South Korea. J Environ Manag 305:114367

Huang F, Tao S, Li D, Lian Z, Catani F, Huang J, Li K, Zhang C (2022) Landslide Susceptibility prediction considering neighborhood characteristics of landslide spatial datasets and hydrological slope units using remote sensing and GIS technologies. Remote Sens 14(18)

Huang F, Zhang J, Zhou C, Wang Y, Huang J, Zhu L (2019) A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. Landslides 17(1):217–229

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

Huang FM, Cao ZS, Guo JF, Jiang SH, Li S, Guo ZZ (2020b) Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. Catena 191:104580

Jiang S-H, Huang J, Huang F, Yang J, Yao C, Zhou C-B (2018) Modelling of spatial variability of soil undrained shear strength by conditional random fields for slope reliability analysis. Appl Math Model 63:374–389

Kayastha P (2015) Landslide susceptibility mapping and factor effect analysis using frequency ratio in a catchment scale: a case study from Garuwa sub-basin East, Nepal. Arab J Geosci 8(10):8601–8613

Kim SH, Geem ZW, Han GT (2020) Hyperparameter optimization method based on harmony search algorithm to improve performance of 1D CNN human respiration pattern recognition system. Sensors (Basel) 20(13):3697

Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90

Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553): 436–444

Lee ML, Ng KY, Huang YF, Li WC (2013) Rainfall-induced landslides in Hulu Kelang area, Malaysia. Nat Hazards 70(1):353–375

Lee WY, Park SM, Sim KB (2018) Optimal hyperparameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm. Optik 172:359–367

Li CY, Wang XC, He CZ, Wu X, Kong ZY, Li XL (2019) National 1 : 200 000 Digital Geological Map (Public Edition) Spatial Database. Geol China 46:1–10

Li Y, Yang XD, Fang H, Yin CR, Qu XY (2012) Zoning Atlas of Geological Disaster Susceptibility Levels in Typical Counties (Cities) in China. Science Press, Beijing

Lin MS, Teng S, Chen GF, Lv JB, Hao ZY (2022) Optimal CNN-based semantic segmentation model of cutting slope images. Front Struct Civ Eng 16(4):414–433

Liu Y, Xu C, Huang B, Ren XW, Liu CQ, Hu BD, Chen Z (2020) Landslide displacement prediction based on multi-source data fusion and sensitivity states. Eng Geol 271:105608

Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, Avtar R, Abderrahmane B (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Sci Rev 207

Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5:3–30

Nhu VH, Mohammadi A, Shahabi H, Ahmad BB, Al-Ansari N, Shirzadi A, Geertsema M, R. Kress V, Karimzadeh S, Valizadeh Kamran K, Chen W (2020) Landslide detection and susceptibility modeling on Cameron Highlands (Malaysia): a comparison between random forest, logistic regression and logistic model tree algorithms. Forests 11(8):830

Nikoobakht S, Azarafza M, Akgün H, Derakhshani R (2022) Landslide susceptibility assessment by using convolutional neural network. Appl Sci 12(12):5992

Oh HJ, Pradhan B (2011) Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. Comput Geosci 37(9):1264–1276

Panahi M, Jaafari A, Shirzadi A, Shahabi H, Rahmati O, Omidvar E, Lee S, Bui DT (2021) Deep learning neural networks for spatially explicit prediction of flash flood probability. Geosci Front 12(3)

Park JY, Lee SR, Lee DH, Kim YT, Lee JS (2019) A regional-scale landslide early warning methodology applying statistical and physically based approaches in sequence. Eng Geol 260:105193

Pham BT, Bui DT, Indra P, Dholakia MB (2015) Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS – based statistical approach of frequency ratio method. Int J Eng Res Technol 4(11):338–344

Pham BT, Bui DT, Prakash I, Dholakia MB (2016a) Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. Nat Hazards 83(1):97–127

Pham BT, Bui DT, Dholakia MB, Prakash I, Pham HV (2016b) A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. Geotech Geol Eng 34(6):1807–1824

Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. Earth Sci Rev 180:60–91

Sameen MI, Pradhan B, Lee S (2020) Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. Catena 186

Sun DL, Wen HJ, Wang DZ, Xu JH (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology 362:107201

Thi-Ngo PT, Panahi M, Khosravi K, Ghorbanzadeh O, Kariminejad N, Cerda A, Lee S (2021) Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. Geosci Front 12(2):505–519

Tian YY, Xu C, Ma SY, Xu XW, Wang SY, Zhang H (2018a) Inventory and spatial distribution of landslides triggered by the 8th August 2017 MW 6.5 Jiuzhaigou Earthquake, China. J Earth Sci 30(1):206–217

Tian YY, Xu C, Hong HY, Zhou Q, Wang D (2018b) Mapping earthquake-triggered landslide susceptibility by use of artificial neural network (ANN) models: an example of the 2013 Minxian (China) Mw 5.9 event. Geomatics Nat Hazards Risk 10(1):1–25

Tien Bui D, Hoang ND, Martinez-Alvarez F, Ngo PT, Hoa PV, Pham TD, Samui P, Costache R (2020) A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. Sci Total Environ 701:134413

Ullah K, Wang Y, Fang Z, Wang L, Rahman M (2022) Multi-hazard susceptibility mapping based on Convolutional Neural Networks. Geosci Front 13(5)

Wang HJ, Xiao T, Li XY, Zhang LL, Zhang LM (2019a) A novel physically-based model for updating landslide susceptibility. Eng Geol 251:71–80

Wang Y, Fang Z, Hong H (2019b) Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Sci Total Environ 666:975–993

Wang YL, Zhang HX, Zhang GW (2019c) cPSO-CNN: an efficient PSO-based algorithm for fine-tuning hyper-parameters of convolutional neural networks. Swarm Evol Comput 49:114–123

Wu YL, Ke YT, Chen Z, Liang SY, Zhao HL, Hong HY (2020) Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. Catena 187:104396

Xie HL, Zhang L, Lim CP (2020) Evolving CNN-LSTM models for time series prediction using enhanced grey wolf optimizer. IEEE Access 8:161519–161541

Xing Y, Yue JP, Guo ZZ, Chen Y, Hu J, Travé A (2021) Large-scale landslide susceptibility mapping using an integrated machine learning model: a case study in the Lvliang Mountains of China. Front Earth Sci 9:722491

Yi J, Gao L, Li XY, Gao J (2015) An efficient modified harmony search algorithm with intersect mutation operator and cellular local search for continuous function optimization problems. Appl Intell 44(3):725–753

Yi YN, Zhang ZJ, Zhang WC, Jia HH, Zhang JQ (2020) Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. Catena 195:104851

Zhou QQ, Si-Tu ZX, Teng S, Chen GF (2021) Convolutional neural networks–based model for automated sewer defects detection and classification. J Water Resour Plan Manag 147(7):04021036

Zhou YQ, Sheng Q, Chen J, Li N, Fu XD, Zhou YB (2022) The failure mode of transmission tower foundation on the landslide under heavy rainfall: a case study on a 500-kV transmission tower foundation on the Yanzi landslide in Badong, China. Bull Eng Geol Env 81(3):125

Zhu Q, Chen L, Hu H, Pirasteh S, Li H, Xie X (2020) Unsupervised feature learning to improve transferability of landslide susceptibility representations. IEEE J Select Topics Appl Earth Observ Remote Sens 13:3917–3930