

## Article

# A Knowledge-Guided Approach for Landslide Susceptibility Mapping Using Convolutional Neural Network and Graph Contrastive Learning

Huimin Liu <sup>1,2</sup> , Qixuan Ding <sup>1</sup>, Xuexi Yang <sup>1,2</sup> , Qinghao Liu <sup>1</sup> , Min Deng <sup>1,2,3</sup> and Rong Gui <sup>1,\*</sup> 

<sup>1</sup> School of Geosciences and Info-Physics, Central South University, Changsha 410083, China; lhmgis@csu.edu.cn (H.L.); 215011041@csu.edu.cn (Q.D.); yangxuexi@csu.edu.cn (X.Y.); 235001015@csu.edu.cn (Q.L.); dengmin@csu.edu.cn (M.D.)

<sup>2</sup> Hunan Geospatial Information Engineering and Technology Research Center, Changsha 410018, China

<sup>3</sup> School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China

\* Correspondence: ronggui@csu.edu.cn

**Abstract:** Landslide susceptibility mapping (LSM) constitutes a valuable analytical instrument for estimating the likelihood of landslide occurrence, thereby furnishing a scientific foundation for the prevention of natural hazards, land-use planning, and economic development in landslide-prone areas. Existing LSM methods are predominantly data-driven, allowing for significantly enhanced monitoring accuracy. However, these methods often overlook the consideration of landslide mechanisms and uncertainties associated with non-landslide samples, resulting in lower model reliability. To effectively address this issue, a knowledge-guided landslide susceptibility assessment framework is proposed in this study to enhance the interpretability and monitoring accuracy of LSM. First, a landslide knowledge graph is constructed to model the relationships between landslide entities and summarize landslide susceptibility rules. Next, combining the obtained landslide rules with geographic similarity principles, high-confidence non-landslide samples are selected to optimize the quality of the samples. Subsequently, a Landslide Knowledge Fusion Cell (LKF-Cell) is utilized to couple landslide data with landslide knowledge, resulting in the acquisition of informative and semantically rich landslide event features. Finally, a precise and credible landslide susceptibility assessment model is built based on a convolutional neural network (CNN), and landslide susceptibility spatial distribution levels are mapped. The research findings indicate that the CNN-based model outperforms traditional machine learning algorithms in predicting landslide probability; in particular, the Area Under the Curve (AUC) of the model was improved by 3–6% after sample optimization, and the AUC value of the LKF-Cell method was 6–11% higher than the baseline method.

**Keywords:** landslide susceptibility mapping; knowledge graph; convolutional neural network



**Citation:** Liu, H.; Ding, Q.; Yang, X.; Liu, Q.; Deng, M.; Gui, R. A Knowledge-Guided Approach for Landslide Susceptibility Mapping Using Convolutional Neural Network and Graph Contrastive Learning. *Sustainability* **2024**, *16*, 4547. <https://doi.org/10.3390/su16114547>

Academic Editor: Lucio Di Matteo

Received: 9 April 2024

Revised: 17 May 2024

Accepted: 21 May 2024

Published: 27 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Landslides, as geological disasters with cascade effects, pose significant threats to human life, property, and environmental sustainability, thereby severely impeding sustainable development [1]. Due to complex geological conditions, intense tectonic activity, and extensive engineering construction, landslides are widespread in China, causing substantial losses through the frequent occurrence of landslide disasters. According to data from the China Geological Survey Institute, tens of thousands of landslides occurred in China from 2010 to 2020, resulting in 8276 casualties and economic losses of USD 6.92 billion. Therefore, establishing accurate assessment and dynamic early warning models for landslide risks, allowing for the detection and identification of potential landslide hazards in advance, is an effective approach to reduce or even prevent major landslide accidents.

The key to effective prevention and control of landslide disasters lies in accurate LSM. With the development of big data and artificial intelligence technology, multi-modal

monitoring data and automated algorithms have been provided for landslide disaster early warning. Based on this, existing research has mainly adopted data-driven approaches to establish the relationship between landslide influencing factors and landslide events, as well as inferring the regional landslide risk level. Existing data-driven LSM methods are mainly divided into two types: mathematical statistical methods and machine learning methods. Mathematical statistical methods mainly rely on statistical information from large amounts of monitoring data to analyze the correlations between landslide influencing factors and landslide events, determine factors and their weights based on these correlations, and conduct LSM. Techniques such as multiple regression, frequency ratio, statistical index models, and evidence weight have been widely used for landslide susceptibility assessment [2–5]. However, traditional statistical models have weak modeling capabilities for the complex non-linear relationships between landslides and conditional factors [6], limiting the ability of LSM. With the continuous development of computer technology, various machine learning methods have gradually been used for regional landslide spatial prediction, such as random forest (RF) [7,8], logistic regression (LR) [9–11], artificial neural networks (ANNs) [12], and support vector machines (SVMs) [13,14]. Compared to mathematical statistical methods, machine learning models can better model the mixed correlations between landslide influencing factors and events, effectively improving the landslide prediction accuracy. However, machine learning methods directly classify input data, they are unable to fully explore the significant features of these data, and they have limited ability to represent inter-feature relationships [15]. In recent years, CNNs have been used for LSM, which have demonstrated higher predictive capabilities than conventional machine learning algorithms [16,17]. Liu et al. [18] compared the overall performance of CNN models with traditional machine learning models, such as random forest, logistic regression, and support vector machines, for LSM, and found that the CNN achieved the highest predictive performance, while also significantly reducing the salt and pepper effect compared to traditional machine learning approaches. The above research contributes to significantly improving the classification accuracy of LSM methods from the perspectives of data mining and model optimization.

Existing data-driven methods focus on fitting models using a large amount of data, and their classification accuracy highly depends on the scale and reliability of monitoring data. Although increasing the amount of sample data can improve model performance to some extent, there is still a “ceiling” effect on the improvement and breaking through this “ceiling” requires guidance from landslide mechanism knowledge. Therefore, using landslide mechanism knowledge as semantic supplementation to guide models in exploring the potential patterns, trends, and correlations in landslide data at a microscopic level is crucial [19,20].

The theory of landslide disaster systems points out that landslide influencing factors mainly include triggering factors, susceptible environments, and vulnerable bodies [21]. These factors are numerous and complex, posing severe challenges for the extraction of landslide mechanism knowledge. How to effectively organize multi-source heterogeneous landslide data is one of the most important problems to be solved. In recent years, using knowledge graphs to process multi-source heterogeneous information has become a research hotspot. Knowledge graphs express entities, concepts, and their relationships in the form of nodes and edges and have been widely used for the storage, management, querying, and analysis of multi-source heterogeneous data. Their powerful knowledge reasoning ability provides new solutions for the acquisition, use, and display of potential relationships and chain propagation of geological disaster information, facilitating the rapid collection, integration, and correlation of geological disaster-related data, as well as providing key support for comprehensive and high-precision knowledge management and decision support for landslide disasters. However, existing research on landslide knowledge graphs has mainly focused on constructing the graphs themselves [22–24], and landslide knowledge graphs have not yet been applied as prior knowledge to guide LSM.

At the same time, regarding the limitations of data-driven landslide monitoring samples, existing studies have mainly balanced the positive and negative ratios of the data set through generating pseudo-negative samples to replace non-landslide samples [25]. However, there is considerable uncertainty in the selection of non-landslide samples. Due to the randomness of model training, the selection of non-landslide samples has a significant impact on the performance of the model. Chen et al. [11] proposed limiting negative samples based on spatial distance; however, this method has strong data randomness and low reliability in selecting negative samples. Achour et al. [26] proposed a method for measuring the reliability of negative samples based on geographical environmental similarity; however, this method overly relies on the feature distribution of positive samples, easily leading to model over-fitting. In summary, existing non-landslide sample selection methods have strong subjectivity and randomness, resulting in insufficient representativeness of the obtained non-landslide samples, thereby reducing the quality of the test set and affecting the subsequent modeling performance.

To address the above issues, in this study, a landslide knowledge graph is first constructed to describe landslide mechanisms and summarize and generalize landslide susceptibility rules. Then, combining the landslide rules with the geographical similarity of samples, high-confidence non-landslide samples are filtered out. Subsequently, the LKF-Cell method is employed to couple landslide data with landslide knowledge, resulting in landslide event characteristics that are informative and semantically rich. Finally, an accurate and reliable landslide susceptibility assessment model is constructed based on a convolutional neural network, and the spatial distribution level of landslide susceptibility is mapped.

The three main contributions of this study are as follows:

- (1) Addressing the problem that existing knowledge representation models are not suitable for expressing the complex mechanistic knowledge in the landslide domain, the landslide knowledge system centered on “triggering factors–pregnancy environments–vulnerable bodies” is first constructed in this study, and guided by this, a landslide disaster knowledge graph is constructed, thus achieving the structured expression of landslide disaster monitoring data and the fusion of landslide knowledge.
- (2) Addressing the problem of poor directional features of non-landslide samples, a rule-constrained non-landslide sample selection method is first proposed to ensure the discrimination between positive and negative samples and to extract more comprehensive negative sample features as much as possible, effectively improving the LSM accuracy.
- (3) An innovative landslide knowledge fusion method based on a contrastive learning framework is proposed in this study, achieving the organic fusion of a knowledge graph representing landslide mechanisms and a deep neural network representing data features. Landslide event features that are informative and semantically rich are obtained, and the LSM accuracy is effectively improved through feature interpretation using convolutional networks.

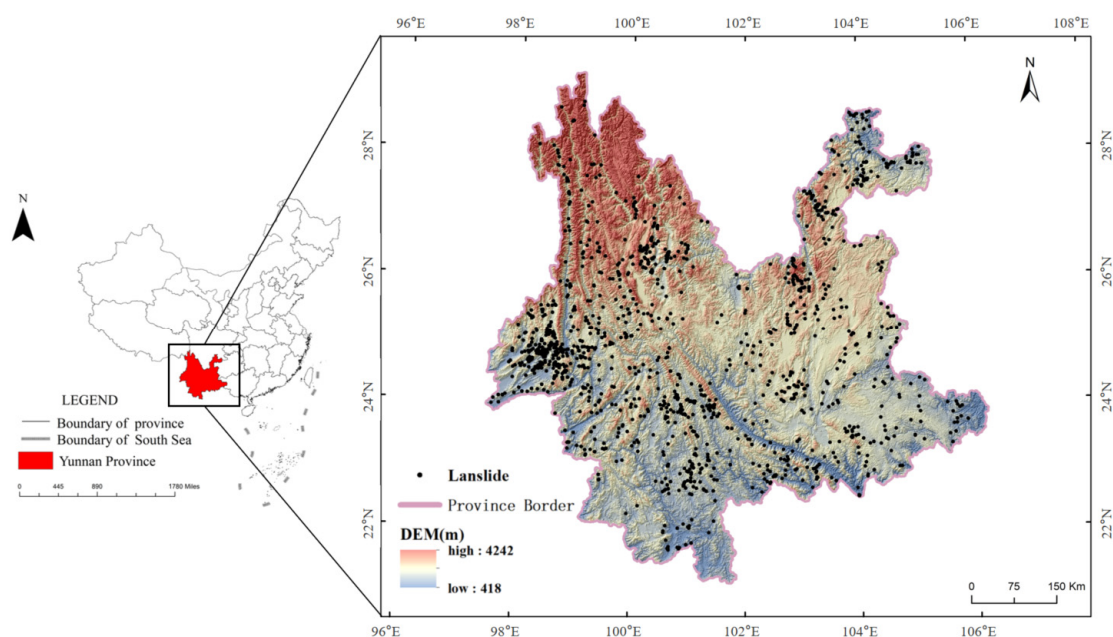
The remainder of this paper is organized as follows. Section 2 introduces the study area and data sources. Section 3 describes the research methods used. Section 4 details the experimental results and analysis. Section 5 summarizes the content of this paper and proposes future prospects.

## 2. Study Area and Available Data

### 2.1. Study Area

Yunnan Province is situated in the southwest of China, covering an area of approximately 390,000 km<sup>2</sup> and extending between 97°53′ E and 106°20′ E in longitude and 21°14′ N and 29°25′ N in latitude (Figure 1). The study area is located at the boundary between the Indian Ocean Plate and the Eurasian Plate, characterized by active tectonic movements and complex lithological formations. Bounded by the Red River Fault, the eastern part comprises karst plateaus, while the western part consists of high mountain

valleys. Under the influence of internal and external geological forces, near-surface rocks and soils are fragmented, resulting in high instability. The topography of Yunnan Province is predominantly mountainous, with some hills and plains. The elevations in the study area vary significantly, ranging from a maximum of 6740 m to a minimum of 76.4 m above sea level. The climate in the study area is classified as tropical monsoon, with an average annual maximum temperature of approximately 23 °C, minimum temperature around 7 °C, and average annual rainfall of about 1086 mm. Rainfall is concentrated mainly from May to October, accounting for 85% of the total annual precipitation. With rapid urbanization and infrastructure development, human activities are increasingly impacting the geological environment. Meanwhile, frequent extreme weather events and increased seismic activity contribute to the persistent trend of multiple and frequent landslide disasters.



**Figure 1.** Location of the study area.

## 2.2. Landslide Conditioning Factors

The landslide inventory data used in this study were sourced from the Yunnan Geological Environmental Monitoring Institute, covering 1712 landslide events from 2015 to 2022. Each landslide inventory was divided into two groups, designated for constructing the model's training data set (70%) and testing data set (30%).

Landslide occurrence is closely related to the combined effects of various factors, including topography, geological structure, and environmental factors. A proper combination of susceptibility factors can enhance the competitiveness of the model [27]. Different topographic features influence rock weathering, vegetation coverage, and soil moisture, thereby affecting slope stability [28,29]. Geological structures affect the stability of strata and rock structures, thereby increasing the risk of landslides [30]. Environmental factors directly impact soil stability and hydrological processes, consequently affecting slope stability [29,31]. Yong C et al. [32] summarized the importance of factors for landslide susceptibility of different causes. Taking into account the evaluation factors in the aforementioned study and the uniqueness of the research area, the following 16 regulating factors were selected: elevation, slope angle, slope aspect, plane curvature, profile curvature, soil types, distance from fault, fault density, lithology, land use, normalized difference vegetation index (NDVI), distance from river, river density, average annual precipitation, distance from road, and road density (see Figure 2). As the 16 evaluation indicators are represented in different forms or scales, all evaluation indicators were standardized into 30 m × 30 m raster data using the ArcGIS10.2. Multilinearity tests were performed on the selected feature factors



using stepwise regression [33]. The tolerance and variance inflation factor (VIF) results indicated that all feature factors had tolerance values greater than 0.1 and VIF values less than 10, indicating low collinearity among the factors and good independence.

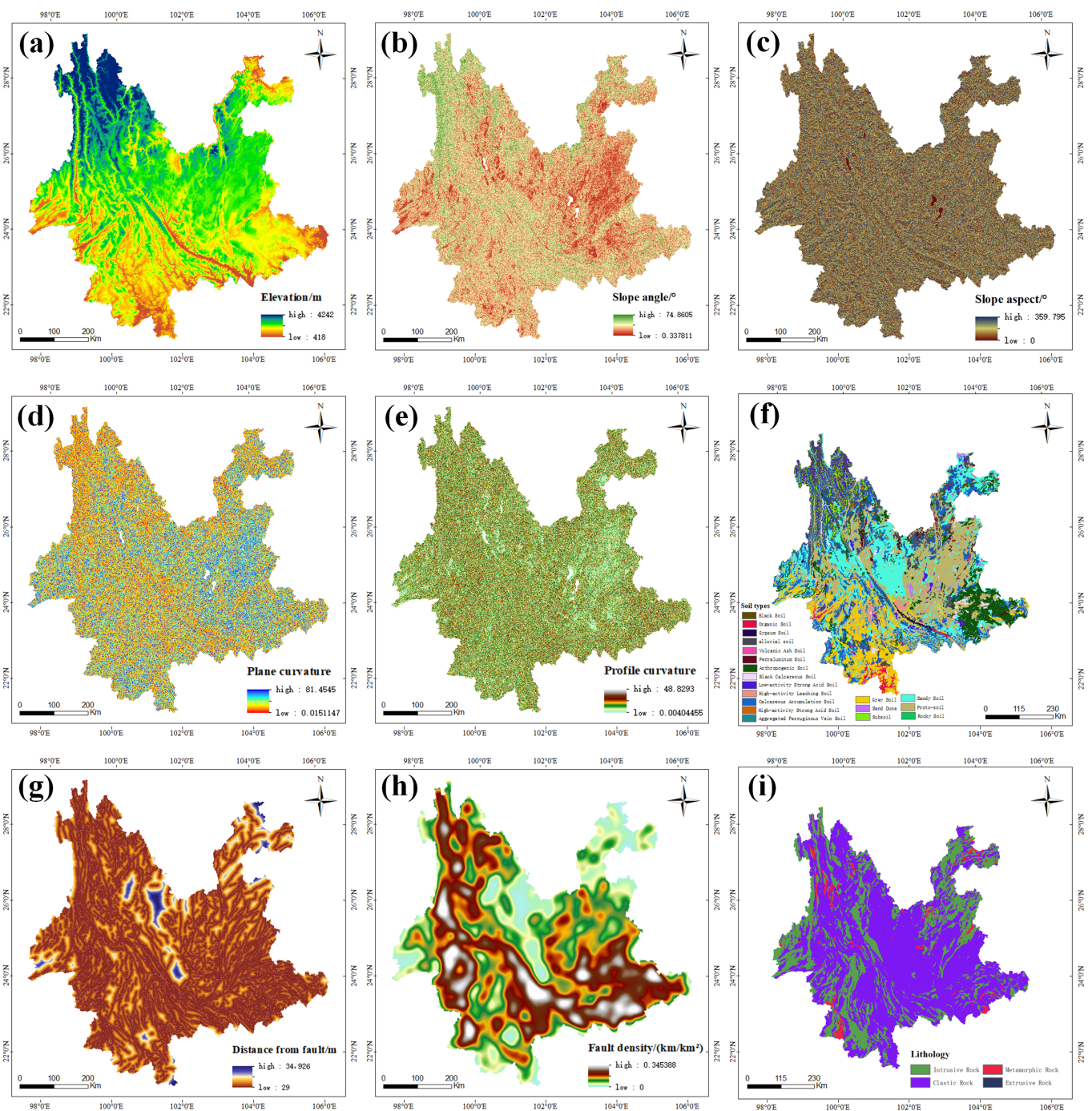
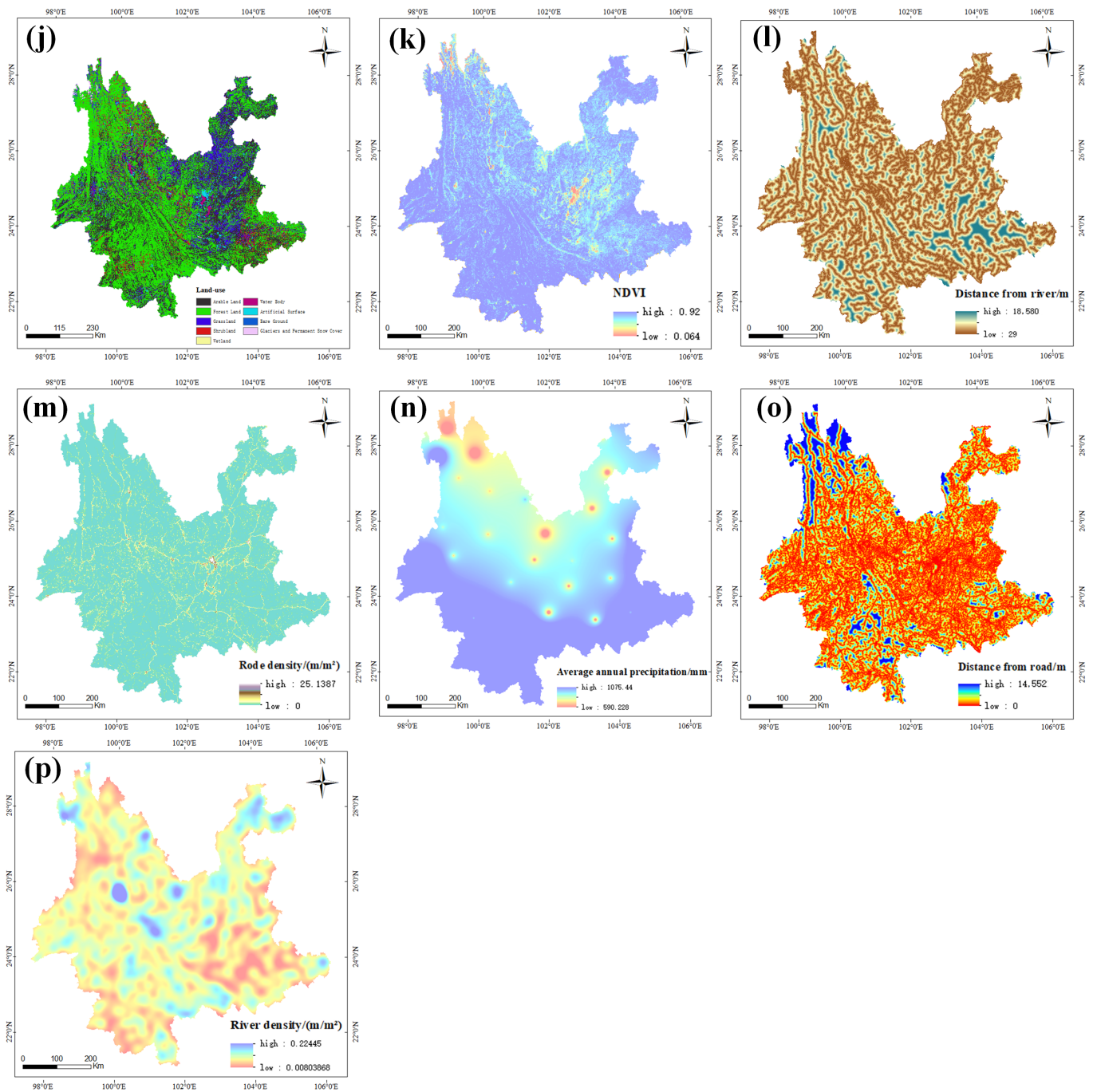


Figure 2. Cont.

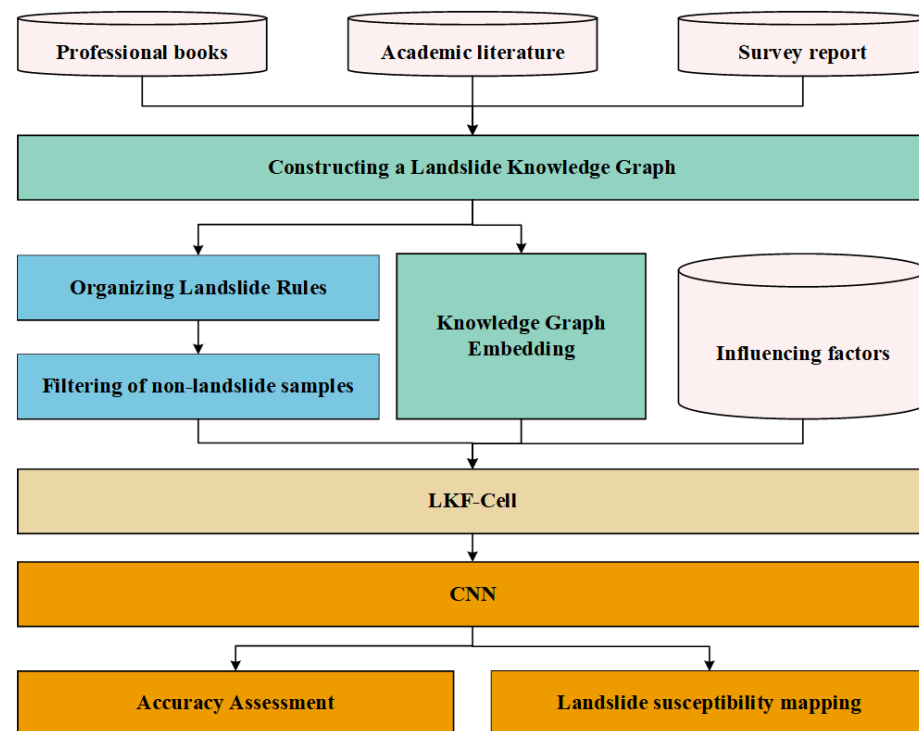


**Figure 2.** Illustration of conditioning factors: (a) elevation; (b) slope angle; (c) slope aspect; (d) plane curvature; (e) profile curvature; (f) soil type; (g) distance from fault; (h) fault density; (i) lithology; (j) land use; (k) NDVI; (l) distance from river; (m) road density; (n) average annual precipitation; (o) distance from road; and (p) river density.

### 3. Methodology

To implement knowledge-guided LSM, the CNN optimized with landslide knowledge fusion (LKF-CNN) model was introduced in this study, whose workflow is depicted in Figure 3. Initially, a landslide knowledge graph was constructed, and an encoded embedding of the landslide knowledge was derived using a knowledge graph embedding model. Subsequently, through combining the landslide rules and the geographical similarity of the samples, high-confidence non-landslide samples were selected to construct landslide event subgraphs. Then, the Landslide Knowledge Fusion Cell (LKF-Cell) method was

used to embed the knowledge with the landslide event subgraph to produce landslide event features integrated with external knowledge. Finally, the CNN model was employed to evaluate landslide sensitivity, generate landslide susceptibility maps, and evaluate the performance of the landslide model on the experimental data set using statistical methods and Receiver Operating Characteristic (ROC) curves.



**Figure 3.** Schematic flowchart of this study.

### 3.1. Prior Knowledge Extraction Method Based on Landslide Knowledge Graphs

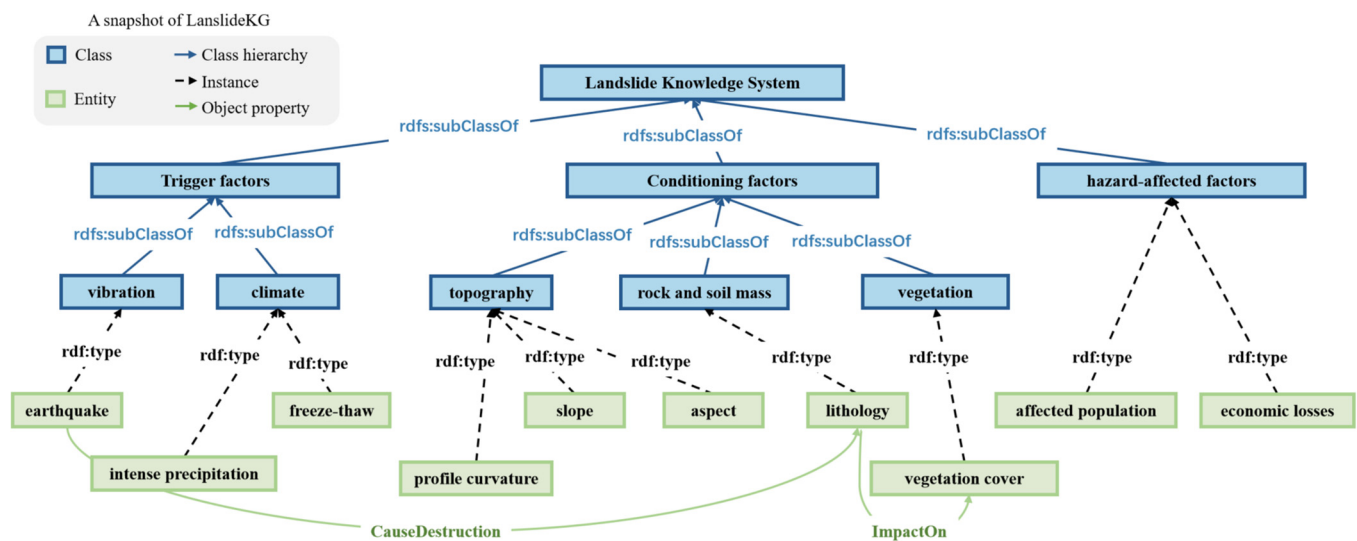
#### 3.1.1. The Construction and Embedding of Knowledge Graphs

In order to provide a comprehensive view of the landslide knowledge system, a landslide knowledge graph was constructed, outlining the hierarchical structure of landslide influencing factors and their object properties. Recognizing the close relationships among these factors, relevant knowledge regarding the mechanisms of landslide occurrence was gathered from hundreds of academic papers to enrich the content of landslide knowledge graph (LandslideKG). Figure 4 shows a snapshot of LandslideKG, which consists of two levels: the instance level and class level (represented in green and blue, respectively). At the instance level, landslide influencing factors are depicted as entities within LandslideKG (denoted by green blocks). Relations between entities, such as destructive and influential relationships among influence factors, are established according to object properties (indicated by green arrows). Subsequently, all entities are classified based on their commonalities to derive the class level of LandslideKG. Entities are assigned to corresponding classes via `rdf:type` (depicted by dashed black arrows). Blue blocks represent different classes, while blue arrows reflect their inclusion relationships (`rdfs:subClassOf`), forming the class hierarchy of LandslideKG, which serves as its backbone.

To fully exploit the structural and semantic information of all entities, relationships, and other components in LandslideKG to obtain meaningful representations, a KG embedding approach based on OWL2Vec\* [34] was employed. This method consists of two steps: first, extracting a corpus from LandslideKG, including structural documents, vocabulary documents, and composite documents and, second, training language models on the corpus to acquire high-quality knowledge graph embeddings [35]. Finally, embeddings for each



entity and relation in LandslideKG were obtained and used to initialize the input features of the molecular graph.



**Figure 4.** Illustration of LandslideKG.

### 3.1.2. Rules Derived from LandslideKG

In this study, the landslide knowledge graph was stored in the Neo4j graph database, and operations were conducted using the Cypher query language. Utilizing syntax such as pattern matching, precise matching of nodes and relationships, property queries, and filtering, event nodes, factor nodes, attribute nodes, and relationships can be queried and matched within the landslide knowledge graph. This enables the extraction of rules and knowledge which are implicit in landslide texts, thus generating rules for landslide susceptibility.

## 3.2. Selection of Non-Landslide Samples Based on Landslide Rules and Geographical Similarity

Non-landslide samples play a crucial role in LSM, as they can mitigate the over-estimation of landslide susceptibility by statistical models and enhance the monitoring accuracy of LSM. Existing LSM methods primarily employ the random generation of pseudo-samples to fill negative samples, which significantly increases the likelihood of misidentifying potential landslide risk points as negative samples. These potential landslide risk points can degrade the overall quality of the sample set, thus impacting the effectiveness of mapping. Therefore, a landslide rule-constrained geographic similarity sample optimization method is proposed in this study. This method evaluates the geographic similarity of landslide occurrences and the relationship between landslide influencing factors and occurrence frequency under the constraint of the susceptibility rules, in order to select high-confidence non-landslide samples.

### 3.2.1. Measurement of Geographic Environmental Similarity

Landslides occur more frequently under certain environmental conditions, indicating that these conditions can be considered as typical geographic environments conducive to landslide formation. To quantify this concept, the frequency ratio method and kernel density estimation were utilized to calculate the relationship between discrete and continuous landslide influencing factors and landslide occurrence frequency, respectively. This was carried out to determine the degree of similarity between individual landslide influencing factors and their typical values for triggering landslides in specific environments.

The frequency ratio method [36] is a single-factor quantitative analysis model that calculates frequencies for sample categories to roughly determine which categories of a discrete influencing factor have a significant impact on landslide occurrence. In Equation (1),



$S'_{i,j}$  represents the frequency of landslides occurring in category  $j$  of influencing factor  $i$ ,  $p_{i,j}$  represents the number of landslides occurring in category  $j$  of influencing factor  $i$ ,  $A_{i,j}$  represents the area of category  $j$  under influencing factor  $i$ ,  $m$  represents the number of categories of influencing factor  $i$ , and  $A$  represents the total area of the study area.

$$S_{i,j}' = \frac{p_{i,j} | A_{i,j}}{\sum_{j=1}^m p_{i,j} | A} \quad (1)$$

Normalizing the frequency of landslides occurring in category  $j$  of influencing factor  $i$  yields the similarity between category  $j$  of influencing factor  $i$  and the typical category under which landslides occur for influencing factor  $i$ . The expression is as follows:

$$S_{i,j} = \frac{S'_{i,j}}{\max(S'_{i,j})} \quad (2)$$

In Equation (2),  $S_{i,j}$  represents the similarity between category  $j$  of influencing factor  $i$  and the typical category under which landslides occur for influencing factor  $i$ .

Kernel density estimation [37] is a method for estimating the probability density function of a population from a sample. Let  $x_1, x_2, \dots, x_i, \dots, x_n$  be the values of an influencing factor  $x$  at  $n$  landslide points. The basic expression for kernel density estimation of the influencing factor  $x$  is as follows:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right) \quad (3)$$

In Equation (3),  $f(x)$  represents the probability density function of the relationship between the influencing factor  $x$  and the frequency of landslide occurrences;  $k(\cdot)$  denotes the kernel function;  $h$  represents the bandwidth, whose value affects the shape and smoothness of the kernel density estimation curve; and  $x - x_i$  represents the difference between the value of the influencing factor  $x$  and the value of the influencing factor at landslide points  $x_i$ . In this study, a Gaussian kernel function was employed to estimate the kernel density curve, and the bandwidth  $h$  was computed using the “rule of thumb” method.

After computing the probability density function, normalization was performed to obtain the degree of similarity between individual influencing factors and their typical values under landslide occurrences. The normalization expression is shown in Equation (4), where  $S_x$  represents the similarity between the influencing factor and its typical value under landslide occurrences, and  $f_{\max}(x)$  represents the maximum value of  $x$ .

$$S_x = \frac{f(x)}{f_{\max}(x)} \quad (4)$$

### 3.2.2. Calculation of Confidence for Non-Landslide Samples

The landslide geographic similarity measurement method highly relies on the quality of positive samples. However, the uncertainties in collecting positive samples result in biased descriptions of landslide features, thus affecting the reliability of the aforementioned measurement method. Therefore, in this study, non-landslide sample selection was constrained by landslide susceptibility rules to reduce the impact of erroneous features caused by sample quality on the overall model.

Specifically, the distribution patterns of individual landslide influencing factors were first assessed to determine whether they adhered to the landslide susceptibility rules. If they complied with these rules, they were deemed as Reliable factors; conversely, if they did not conform, they were classified as Unreliable factors. To assess the guiding effect of different factors on the collection of non-landslide samples, the comprehensive similarity

between each factor and typical geographical environments where landslides occur was calculated using a linear weighting method. The specific formula is as follows:

$$S = \sum_{n=1}^N W_n * S_n \quad (5)$$

$$W_n = \begin{cases} 1 & n \in \text{Reliable factor} \\ 0.5 & n \in \text{Unreliable factor} \end{cases} \quad (6)$$

In the equation,  $n$  denotes the sequence of feature influencing factors for the sample at that location,  $N$  represents the total number of influencing factors,  $W_n$  signifies the weight of the influencing factor,  $S_n$  denotes the geographical similarity of factor  $n$ , and  $S$  represents the comprehensive similarity.

Based on the fundamental principles of geography that “the more similar the geographical environment, the more similar the geographical features,” this study utilized Equation (7) to measure the reliability of negative samples. The reliability of negative samples is within the range of  $[0, 1]$ , where higher values indicate higher reliability.

$$\text{Reliability}_{i,j} = 1 - S_{i,j} \quad (7)$$

Here,  $S_{i,j}$  represents the geographical environment similarity between the raster point at location  $(i, j)$  and landslide points, and  $\text{Reliability}_{i,j}$  represents the reliability of this raster point being selected as a negative sample.

### 3.3. Landslide Knowledge Fusion Cell (LKF-Cell)

To enhance the model’s understanding of landslide domain knowledge, an LKF-Cell was designed with the aim of integrating external knowledge with landslide subgraphs, as a basis for subsequent knowledge-guided landslide susceptibility prediction. Detailed information of the LKF-Cell is illustrated in Figure 5. A contrastive learning framework was employed to learn representations of landslide event subgraphs. Notably, a knowledge-guided graph augmentation method was proposed to construct positive pairs in the contrastive learning stage, which enabled the model to capture not only the deep features of landslide event subgraphs but also the knowledge structure and semantic relationships of landslides. The input to the LKF-Cell consists of knowledge embeddings from LandslideKG and true landslide event subgraphs, with the landslide event subgraph  $G$  being composed of selected influencing factors. The output is the updated features of landslide subgraphs fused with external knowledge, denoted as  $x \in X$ .

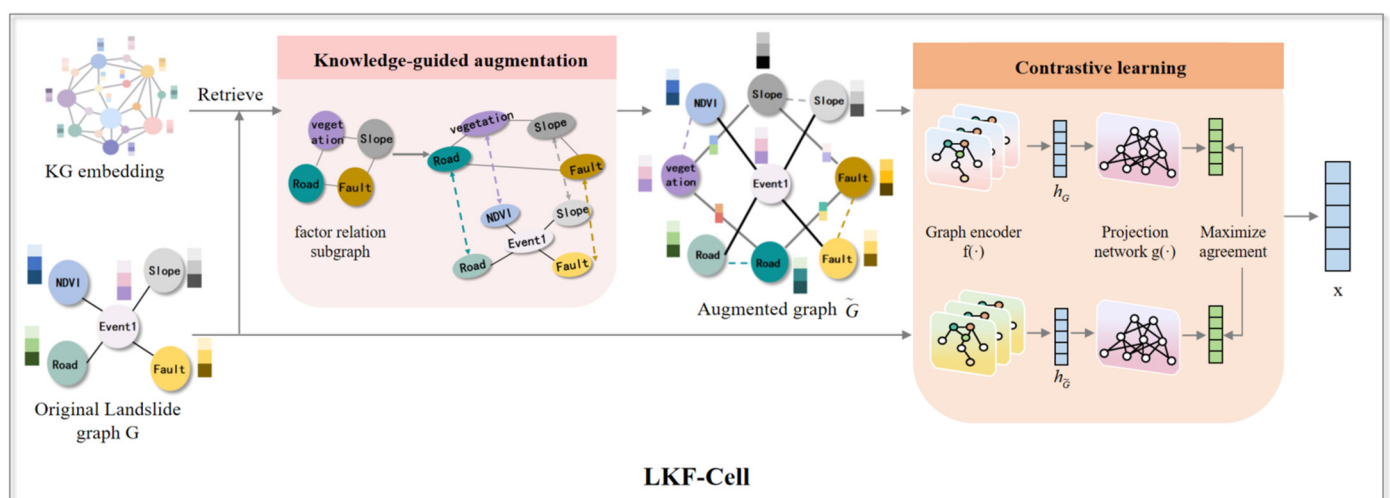


Figure 5. LKF-Cell structure.

As shown in Figure 5, the types of influencing factors present in the original landslide graph  $G$  (e.g., NDVI, slope angle, and road density) were first identified, and their corresponding entities and relationships were retrieved from the KG embedding (e.g., slope, influences, NDVI). This formed a subgraph of landslide influencing factors. The influencing factor entity nodes in the subgraph were then connected to the corresponding influencing factor nodes in the original landslide event graph, in order to create an augmented landslide event graph that integrates fundamental domain knowledge.

Based on this foundation, a contrastive learning framework was employed to train the graph encoder through maximizing the consistency between the original landslide graph  $G$  and the augmented landslide graph  $\tilde{G}$ . Indiscriminate embedding of the knowledge of the influence factors into the augmented graph was avoided throughout this process. Initially,  $N$  randomly sampled batches of landslide event graphs were given, and knowledge-guided graph augmentation techniques were utilized to transform the original landslide graph  $\{G_i\}_{i=1}^N$  into the augmented graph  $\{\tilde{G}_i\}_{i=1}^N$ , resulting in  $2N$  graphs. Among these, apart from the positive pairs consisting of the original landslide graph  $\{G_i\}$  and its augmented counterpart  $\{\tilde{G}_i\}$ ,  $2(N - 1)$  graphs in the same batch were considered as negative pairs. Subsequently, a graph encoder  $f(\cdot)$  was employed to extract graph embeddings  $\{h_{G_i}\}_{i=1}^N$  and  $\{h_{\tilde{G}_i}\}_{i=1}^N$  from the two views and, through a non-linear projection network  $g(\cdot)$ , these embeddings were mapped into a space where contrastive loss is applied, yielding two new representations:  $\{z_{G_i}\}_{i=1}^N$  and  $\{z_{\tilde{G}_i}\}_{i=1}^N$ . Finally, the contrastive loss was leveraged to maximize the consistency between positive pairs while simultaneously minimizing the consistency between negative pairs.

### 3.4. Knowledge-Guided LSM

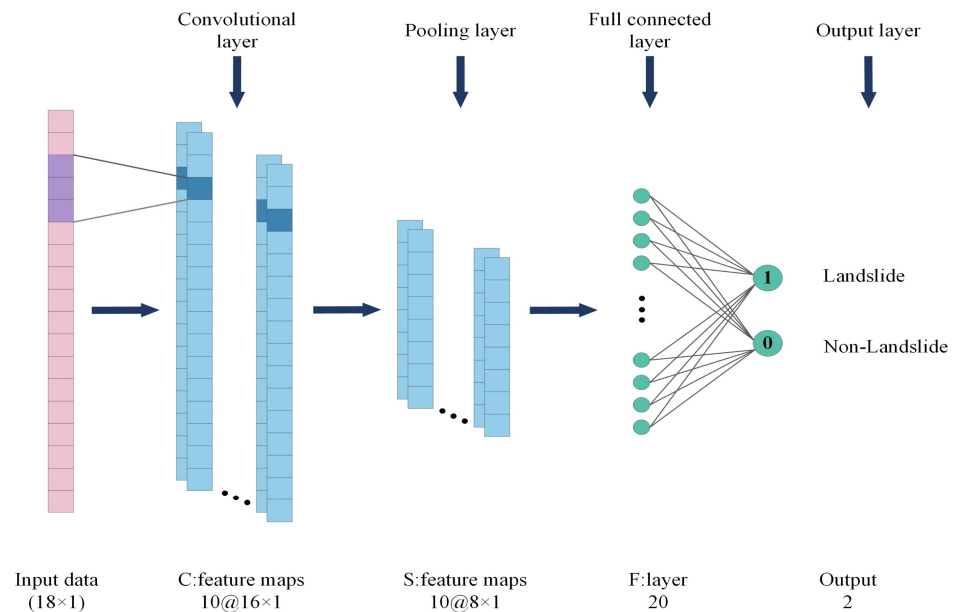
Addressing the insufficient consideration of landslide mechanisms in existing LSMs, this study incorporated landslide event features fused with external knowledge as inputs to a CNN model. This approach facilitates quantitative prediction of landslide susceptibility in a manner constrained by landslide knowledge. The landslide models were evaluated using Receiver Operating Characteristic (ROC) curves and statistical methods. Finally, a landslide susceptibility zoning map was generated.

#### 3.4.1. CNN Architecture

CNNs are among the most widely used deep learning algorithms. In this section, the CNN model serves directly as a classifier for LSM. First, a one-dimensional factor vector was input and, after undergoing convolution and pooling operations in the hidden layers, the extracted high-dimensional landslide information was mapped to a low-dimensional feature space through fully connected layers. Finally, through applying a non-linear activation function, the information was mapped to the sample label space, outputting landslide and non-landslide labels along with their corresponding probability values. With filters = 10, kernel\_size = 3, learning\_rate = 0.001, pool\_size = 2, and activation = "tanh", Figure 6 illustrates the architecture of the used 1D-CNN.

#### 3.4.2. Evaluation of Model Performance

In this study, the performance of the model on the test data set was evaluated using statistical methods and ROC curves. Accuracy (ACC) is defined as the percentage of correctly classified samples relative to the total number of samples selected for validation. The Kappa index primarily reflects the proportion of reduced errors compared to random classification [38,39]. The mathematical expressions for these metrics are provided below.



**Figure 6.** CNN architecture.

The Receiver Operating Characteristic (ROC) curve is considered an effective metric for evaluating the performance of predictive models [30]. The ROC curve plots “1-specificity” against “sensitivity” and illustrates the classifier’s performance at different thresholds. The specificity represents the proportion of non-landslide pixels that are correctly classified as non-landslide pixels, while the sensitivity represents the proportion of landslide pixels that are correctly classified as landslide occurrences. The Area Under the Curve (AUC) value represents the area under the ROC curve, with values typically ranging from 0.5 to 1. A higher AUC value generally indicates better performance of the corresponding model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (10)$$

$$\text{Kappa} = \frac{P_{obs} - P_{exp}}{1 - P_{exp}} \quad (11)$$

$$P_{obs} = \frac{TP + TN}{TS} \quad (12)$$

$$P_{exp} = \frac{(TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN)}{TS} \quad (13)$$

### 3.4.3. Landslide Susceptibility Maps

To construct the landslide susceptibility map, two main steps were followed: (1) all pixels within the study area were input into the trained model to generate landslide susceptibility indices (LSIs) using landslide models, and (2) the LSIs were re-classified [40].

However, the obstacle of transforming continuous data into some classes remains uncertain in the context of landslide susceptibility mapping, as susceptible zones are often determined according to expert knowledge and opinion [41]. For the current study, the natural breaks (Jenks) classification method [42], equal intervals method [43], and standard deviations method [41] were primarily employed. The natural breaks method categorizes data through analyzing distribution patterns and clustering tendencies, thereby dividing



the data into several categories. It does not rely on a pre-defined number of classes but instead utilizes inherent distribution characteristics to determine categories. The equal intervals method is a straightforward approach that evenly divides the data range into several equally sized intervals. However, this method emphasizes the sensitivity of one class, which is not applicable. The standard deviations method determines classification boundaries through analyzing the standard deviation of the data. It considers data variability and determines classification positions based on data dispersion. However, this method requires the prior specification of multiples of the standard deviation to determine classification boundaries, rendering it unsuitable for susceptibility studies.

Among these methods, the natural breaks method is the most widely applied. The natural breaks method was utilized to divide the LSI into five sensitivity levels, including very low (VLS), low (LS), medium (MS), high (HS), and very high (VHS) [44].

#### 4. Results and Discussion

##### 4.1. Optimal Selection of Non-Landslide Samples

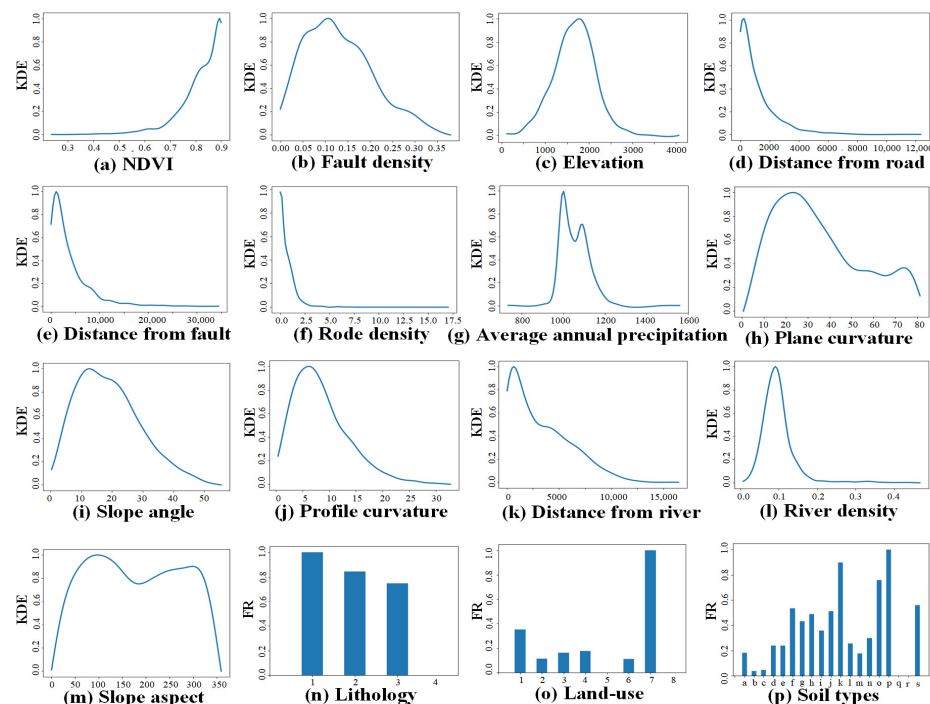
The similarity between the calculated values for the landslide influencing factors from Section 3.2.1 and their typical values associated with landslide occurrences is illustrated in Figure 7. The landslide susceptibility rules obtained from Section 3.1.2 are presented in Table 1. As per Rule 1, a high NDVI index indicates rock damage, thus affecting slope stability. However, this rule contradicts the findings shown in Figure 7a. According to Rule 10, land-use types such as cropland, forest land, grassland, and shrub land are prone to landslides. However, Figure 7o indicates that landslides mainly occur on bare land, which contradicts this rule. Therefore, given the insignificance of NDVI and land-use types as factors, and to reduce model over-fitting caused by limited training data, their weights should be reduced when calculating the confidence of non-landslide samples.

**Table 1.** Landslide susceptibility rules derived from LandslideKG.

#	Landslide Susceptibility Rules
1	Dense vegetation cover can reduce the negative impact of rainwater on slopes; however, the growth of vegetation exerts pressure on rocks, leading to their damage and increasing water infiltration, thereby affecting the stability of the slope.
2	Faults cause damage to the surrounding rock masses, thereby affecting the stability of slopes. Typically, landslide concentration areas are within a range of 1 km from the fault.
3	In areas with relatively low relative altitudes, landslides are more prone to occur due to frequent human engineering activities, especially within the range below 2000 m above sea level.
4	The development and usage of roads can alter the geological structure and hydrogeological characteristics of slopes, potentially leading to soil erosion, increased surface water runoff, vegetation destruction, and, consequently, increased landslide risks.
5	Minor rainfall can infiltrate underground, increasing groundwater content, thereby altering the stress state of the slope, affecting its stability. Intense rainfall can heavily erode the slope surface, directly leading to landslides. Typically, landslide-prone areas are concentrated in regions with annual rainfall between 500 mm and 2000 mm.
6	The degree of distortion and deformation on the slope surface directly affects the stress distribution within the slope, thereby influencing landslide occurrences to varying extents.
7	In environments of heavy rainfall, gently sloping surfaces are susceptible to strong surface erosion, resulting in deeper water infiltration and, consequently, structural damage within the slope, ultimately triggering landslides. Generally, landslides occur more frequently in regions with slope angles between 10° and 40°.
8	Different slope aspects receive varying intensities of solar radiation, leading to differences in vegetation cover and surface moisture content. Typically, landslide concentration areas lie within slope aspects ranging from 100° to 200° and 250° to 330°.
9	The physical and mechanical properties of rock masses, as well as their interlayer structures, directly influence stress distribution within the rock–soil mass, with predominantly clastic and metamorphic rocks being prone to landslides.

Table 1. Cont.

#	Landslide Susceptibility Rules
10	Land-use types most susceptible to landslides include cultivated land, forest land, grassland, and shrubland.
11	Different soil types have varying densities, pore structures, and moisture contents, thus exhibiting different responses to external forces. Soil types prone to landslides mainly include loess and black soil.
12	River infiltration softens the slope's weathering layer, reducing its shear strength. The closer the proximity to rivers, the higher the risk of landslides occurring.



(n)Lithology: 1-Intrusive rock; 2-Sedimentary rock; 3-Metamorphic rock; 4-Extrusive rock.

(o)Land-use: 1-Arable land; 2-Forest land; 3-Grassland; 4-Shrubland; 5-Wetland; 6-Water body; 7-Artificial surface; 8-Bare land.

(p)Soil types: a-Alluvial soil; b-Submerged soil; c-Low-activity acid soil; d-High-activity acid soil; e-Volcanic ash soil; f-Sandy soil; g-Artificial soil; h-Black calcareous soil; i-Calcareous soil; j-Proto-soil; k-Ferric-Aluminic soil; l-Gray soil; m-Gypsum soil; n-Organic soil; o-High-activity lime dissolution soil; p-Black soil; q-Polysulfide network soil; r-Rocky soil; s-Sand dunes.

**Figure 7.** The similarity between the values of conditioning factors and the typical values of landslide occurrence.

The spatial distribution of randomly selected negative samples (non-landslide disaster points) is shown in Figure 8a. The spatial distribution of negative samples guided by the determined rules is illustrated in Figure 8b. Combining the quantity of negative samples in various elevation categories, as shown in Figure 9, it can be observed that the number of samples obtained using the rule-guided selection method was significantly higher in the elevation range of 1500–2000 m, compared to the random selection method. As indicated by Figure 8b and Rule 3, the elevation range of 1500–2000 m is a high-risk area for landslides. Therefore, the selected samples demonstrate stronger robustness.

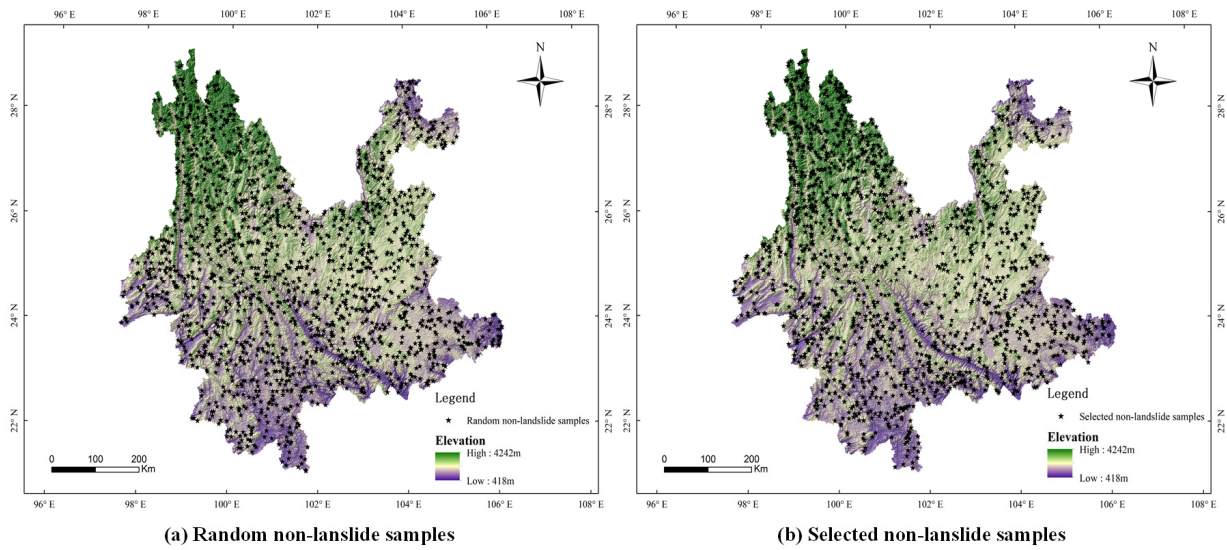


Figure 8. Distribution of sample sites.

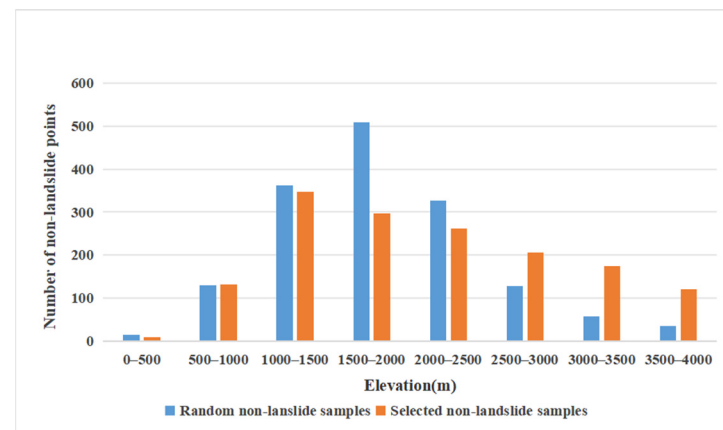
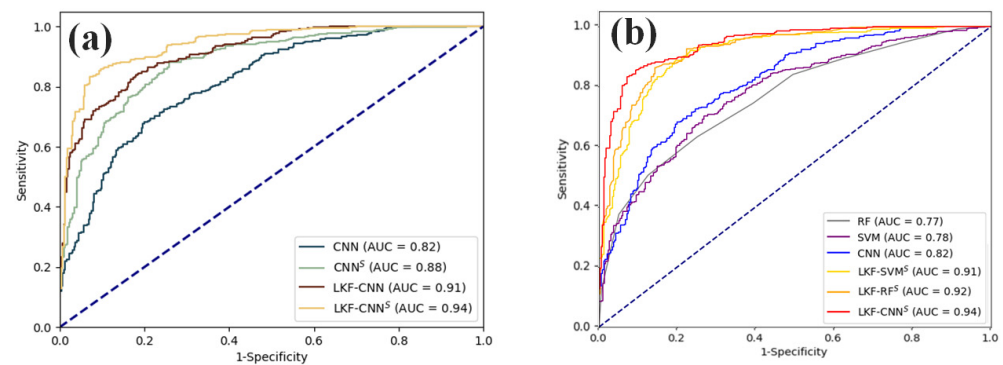


Figure 9. Elevation distribution of non-landslide samples.

#### 4.2. Model Comparison

To quantify the robustness of the sensitivity model, the performances of CNN, CNN optimized with sample enhancement (CNN<sup>S</sup>), CNN optimized with landslide knowledge fusion (LKF-CNN), and CNN optimized with landslide knowledge fusion and sample enhancement (LKF-CNN<sup>S</sup>) models on the landslide inventory were compared, based on the test data set, AUC, and statistical measures. The results are shown in Figure 10a. It can be observed that the AUC values for the Yunnan Province region ranged from 81% to 93%. Among them, the LKF-CNN<sup>S</sup> model had the highest AUC value (94%), followed by the LKF-CNN model (91%), CNN<sup>S</sup> model (88%), and CNN model (82%). It is worth noting that the AUC values of the sample-optimized models were 3–6% higher than those of the models without sample optimization, and the AUC values of the knowledge-guided models were 6–11% higher than those without knowledge guidance. As shown in Table 2, except for the CNN model, the ACC, sensitivity, and specificity of the other three models were all greater than 0.8, indicating a certain level of credibility in the model evaluation results. The Kappa coefficients of the four models ranged from 0.47 to 0.71, meeting the consistency strength. The precision evaluation indicators of CNN<sup>S</sup>, LKF-CNN, and LKF-CNN<sup>S</sup> were generally higher than those of the independent model CNN, with LKF-CNN<sup>S</sup> having the highest overall accuracy, sensitivity, specificity, and Kappa coefficient values. This suggests that the LKF-CNN<sup>S</sup> model has better accuracy and predictive power.



**Figure 10.** ROC curves of the (a) CNN, CNN<sup>S</sup>, LKF-CNN, and LKF-CNN<sup>S</sup> models and (b) RF, SVM, CNN, LKF-SVM<sup>S</sup>, LKF-RF<sup>S</sup>, and LKF-CNN<sup>S</sup> models.

**Table 2.** Model evaluation and comparison on the testing data sets.

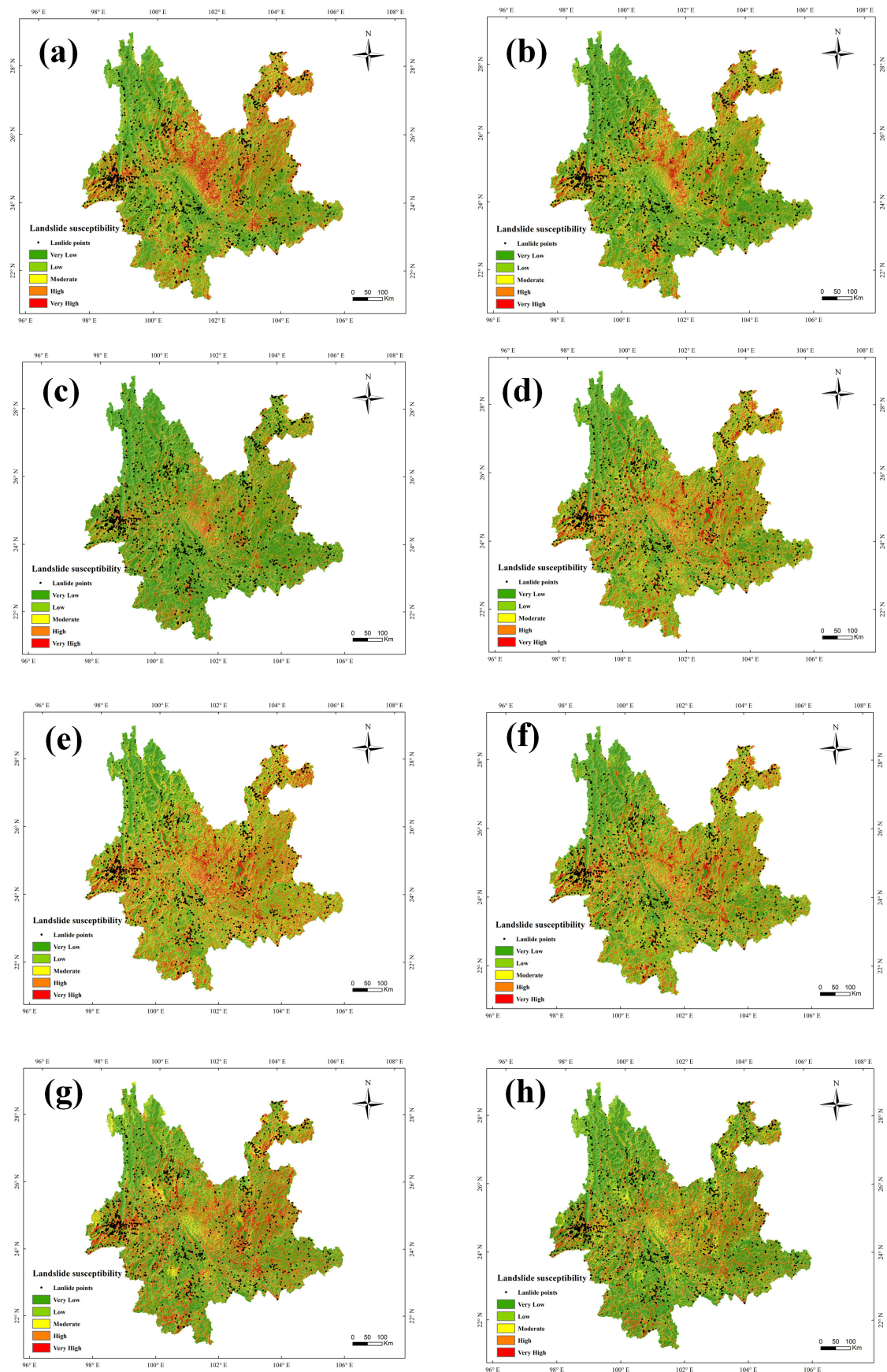
Model	ACC	Sensitivity	Specificity	Kappa
RF	0.68	0.63	0.75	0.37
SVM	0.65	0.65	0.76	0.4
CNN	0.73	0.71	0.76	0.47
CNN <sup>S</sup>	0.81	0.81	0.81	0.62
LKF-CNN	0.84	0.86	0.82	0.68
LKF-RF <sup>S</sup>	0.83	0.84	0.81	0.65
LKF-SVM <sup>S</sup>	0.79	0.79	0.84	0.62
LKF-CNN <sup>S</sup>	0.86	0.88	0.83	0.71

To compare the performance of different classification algorithms, this study also employed RF, RF optimized with landslide knowledge fusion and sample enhancement (LKF-RF<sup>S</sup>), SVM, and SVM optimized with landslide knowledge fusion and sample enhancement (LKF-SVM<sup>S</sup>) methods for slope stability classification. The model prediction results are presented in Figure 10b and Table 2. The AUC values for RF, SVM, and CNN were 0.77, 0.78, and 0.82, respectively. Compared to the RF and SVM models, CNN exhibited the highest overall accuracy, sensitivity, specificity, and Kappa coefficient values. It can be concluded that the CNN-based model demonstrated better performance in predicting landslide susceptibility, when compared to the two traditional ML models. The AUC values of the models optimized with sample enhancement and knowledge guidance were 12–15% higher than those of the original models. Among them, LKF-CNN<sup>S</sup> achieved the highest AUC value, indicating that the optimized structure possesses stronger predictive capabilities. Similar patterns were observed in accuracy, Kappa coefficient, sensitivity, and specificity, indicating a reasonable agreement between the predicted and actual landslides.

#### 4.3. Landslide Susceptibility Map

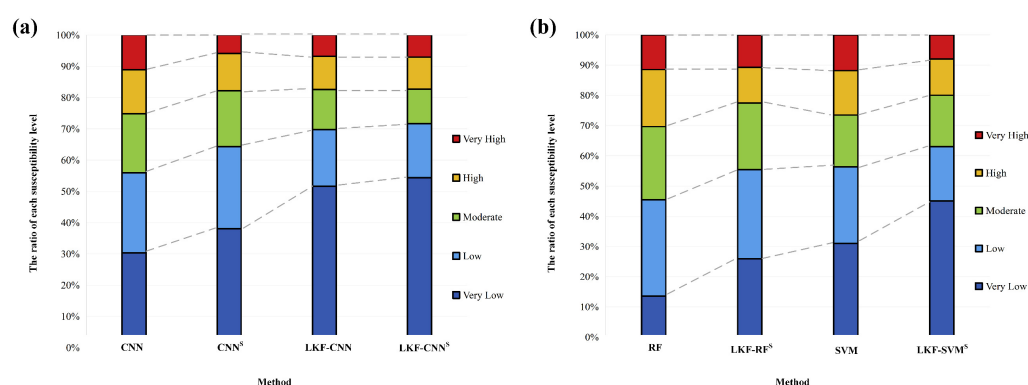
This study utilized CNN, CNN<sup>S</sup>, LKF-CNN, LKF-CNN<sup>S</sup>, RF, LKF-RF<sup>S</sup>, SVM, and LKF-SVM<sup>S</sup> models to generate landslide susceptibility maps for Yunnan Province (Figure 11). The study area was divided into five levels (VLS, LS, MS, HS, and VHS), based on the natural breakpoints' method. The susceptibility assessment results of the eight models indicated that VHS and HS areas are mainly located in the high mountain gorges of the Three Rivers Region in western Yunnan Province, the plateau gorges in the northeast of Yunnan Province, and the Ailao Mountains and Wuliang Mountains in central Yunnan Province. Meanwhile, the low hills and ridges in the eastern and southeastern parts of Yunnan Province exhibited extensive LS and VLS areas.





**Figure 11.** Landslide susceptibility maps generated using (a) CNN, (b)  $CNN^S$ , (c) LKF-CNN, (d)  $LKF-CNN^S$ , (e) RF, (f)  $LKF-RF^S$ , (g) SVM, and (h)  $LKF-SVM^S$ .

As shown in Figure 12a, the VLS area increased by 2–8% after sample optimization, reflecting the effective reduction in model uncertainty when using a high-confidence sample set. Models optimized with landslide knowledge guidance showed significant changes, mainly reflected in the reduction in the HS, MS, and LS areas and an increase in VLS areas. Guiding landslide susceptibility with landslide knowledge ensures correct classification, thus avoiding unnecessary expenditure of manpower, materials, and financial resources. Figure 12b illustrates that the spatial distribution of VHS, HS, and MS areas was similar with the two traditional ML methods, while RF had the most LS areas and the fewest VLS areas. The optimized models LKF-RF<sup>S</sup> and LKF-SVM<sup>S</sup> produced more reasonable distribution patterns for susceptibility zones. In flat terrain such as river valleys and plains, they predominantly exhibited very low and low susceptibility. Compared to LKF-CNN<sup>S</sup>, the VHS and HS regions were relatively smaller for both coupled models, at 11.47%, 18.89%, 8.78%, and 12.47%, respectively.

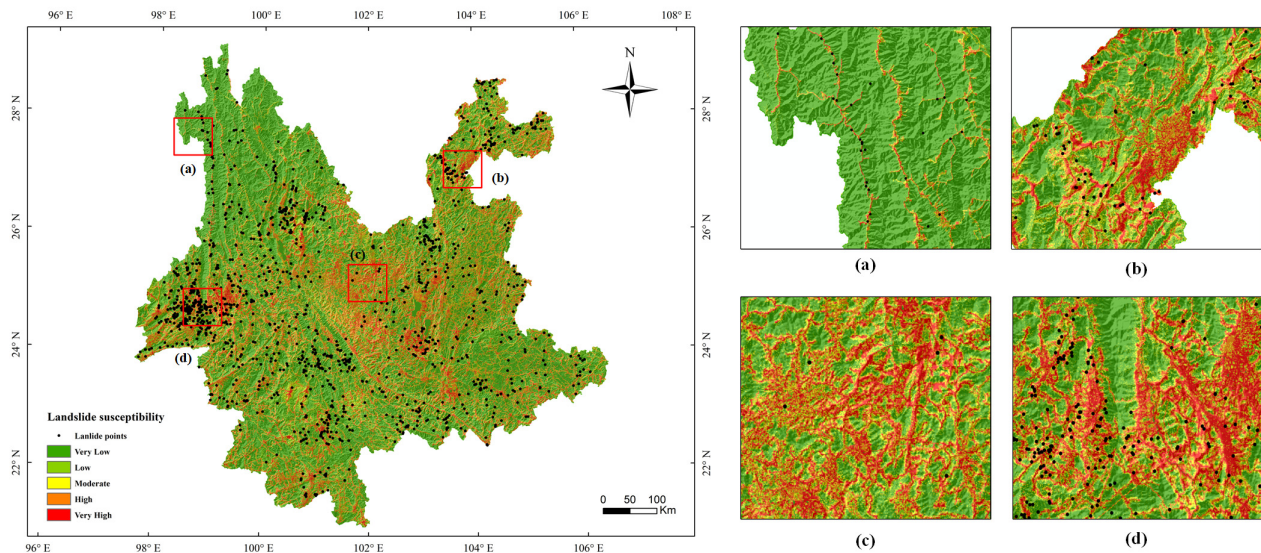


**Figure 12.** Percentages of different landslide susceptibility classes for (a) CNN, CNN<sup>S</sup>, LKF-CNN, and LKF-CNN<sup>S</sup>; and (b) RF, SVM, CNN, LKF-SVM<sup>S</sup>, LKF-RF<sup>S</sup>, and LKF-CNN<sup>S</sup>.

From Figure 13, it is evident that the locations of the existing landslide events matched the distribution results of the assessment well, with most of the locations falling within the VHS and HS ranges. Figure 13a is located in the Hengduan Mountains' transverse valley zone in the northwest of Yunnan Province. This area has complex geological structures with active north–south trending major faults and fragmented rock masses, leading to landslides distributed along the Hengduan Mountains. Figure 13b is situated in the northeast of Yunnan Province, belonging to the high and middle mountain gorge geomorphic region of the middle and lower reaches of the Jinsha River. This area experiences active tectonic movements and has a mixed composition of hard and soft rocks with numerous cliffs, making it prone to landslides. Figure 13c is in the central–northern part of Yunnan Province, within the high and middle mountain gorge region of the upper and middle reaches of the Jinsha River. The area has intense tectonic activities and frequent seismic events, indicating a high landslide risk zone. Figure 13d is located in the western high mountain gorge geomorphic region of Yunnan Province, characterized by complex geological structures, active fault development, widespread distribution of weak rock masses, intense human engineering activities, and frequent landslide occurrences.

Although Figure 13c displays extensive areas of high susceptibility zones, the number of positive landslide sample points in this area is relatively low, possibly due to incompleteness of the landslide inventory, necessitating further field investigations for supplementation. Additionally, while most positive landslide samples were located within high susceptibility zones, Figure 13d indicates that a small number of landslide points were situated in low susceptibility zones. After analysis, two possible scenarios emerged: (1) Although the geographical environment where the landslide event occurred may not typically lead to landslides, extreme weather conditions (e.g., short-term heavy rainfall, snowfall) and human engineering activities (e.g., excavation at the foot of the slope, reservoir impoundment) could trigger the occurrence of landslides. (2) Despite the favorable geological

environment for landslide formation in the area where the landslide event occurred, limited data and the atypical mechanisms of landslide formation, combined with the challenge of utilizing general landslide knowledge, may hinder the accurate identification of landslides by the model.



**Figure 13.** Landslide susceptibility maps generated using LKF-CNN<sup>S</sup> in (a) the northwest of Yunnan Province, (b) the northeast of Yunnan Province, (c) the central–northern part of Yunnan Province and (d) the western of Yunnan Province.

## 5. Conclusions

In response to the mechanisms and uncertainties surrounding landslides, this study proposed a knowledge-guided framework for LSM. The framework integrates existing landslide inventories with 16 landslide evaluation factors and was used to assess the susceptibility of landslides in Yunnan Province, China. The main conclusions are summarized as follows:

**Effective Handling of Landslide Information:** knowledge graphs were shown to be effective in managing diverse sources of landslide information and uncovering the causal mechanisms behind landslides, thereby providing comprehensive knowledge management and decision support for landslide disasters.

**Reliability of Derived Landslide Rules:** The landslide rules inferred from the knowledge graph exhibited high reliability and certainty. Through combining these rules with geographic similarity calculations between landslide and non-landslide samples, high-confidence non-landslide samples were obtained. This approach helps to mitigate the tendency of statistical models to over-estimate landslide hazard. The experimental results demonstrated that, after sample optimization, the model's AUC was improved by 3–6%.

**Enhanced Performance through LKF-Cell:** The LKF-Cell aids in capturing hidden semantic information and inherent knowledge connections in the data, thus providing landslide susceptibility models with richer and more informative landslide event characteristics. The experimental results indicated that the AUC of models based on knowledge guidance was 6–11% higher than that of those without knowledge guidance.

**Model Fusion for Improved Accuracy and Robustness:** The fusion of CNN, RF, and SVM models through sample optimization and LKF-Cell led to models with higher accuracy and robustness. The AUC of these models was improved by 12–15% compared to the base classifiers, and the ACC was increased by 13–15%. The CNN-based models demonstrated optimal predictive performance, in terms of both base classifiers and coupled models.

While this study provided a preliminary framework for integrating landslide knowledge and CNN models for landslide susceptibility mapping with promising results, there



are still areas for improvement. Given the vast and complex terrain of Yunnan Province, which encompasses diverse mechanisms for landslide formation, the utilization of generalized landslide knowledge in model training imposes certain limitations. Future research should aim to apply the proposed framework to smaller regions, employing more fine-grained knowledge guidance for landslide susceptibility prediction, thereby offering more scientifically guided recommendations for disaster prevention, land-use policies, and guidance for local governments and disaster management authorities.

**Author Contributions:** Conceptualization, Q.D., R.G., H.L., and X.Y.; data curation, Q.D. and Q.L.; methodology, H.L., Q.D., and R.G.; visualization, Q.D., H.L., Q.L., and X.Y.; writing—original draft, H.L. and Q.D.; writing—review and editing, H.L., Q.D., R.G., and M.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by grants from the National Key Research and Development Program of China (2022YFB3904203), the National Natural Science Foundation of China (42271485 and 42171459), the Scientific Research Project of the Education Department of Hunan Province (22B0015), the Frontier Cross Research Project of Central South University (2023QYJC002), the Hunan Province Natural Resources Science and Technology Project (20230121XX), the Open Research Fund Program of MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area and the Open Research Fund Program of Guangdong Key Laboratory of Urban Informatics (GEMLab-2023019) and the Jiangxi Province “Double Thousand Plan”, the third batch of short-term projects to introduce innovative leading talents (jxsq2020102062).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** This work was carried out in part using computing resources at the High-Performance Computing Platform of Central South University.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Fan, X.; Scaringi, G.; Korup, O.; West, A.J.; van Westen, C.J.; Tanyas, H.; Hovius, N.; Hales, T.C.; Jibson, R.W.; Allstadt, K.E.; et al. Earthquake-Induced Chains of Geologic Hazards: Patterns, Mechanisms, and Impacts. *Rev. Geophys.* **2019**, *57*, 421–503. [\[CrossRef\]](#)
2. Swain, J.B.; Singh, N.J.; Gupta, L.R. Landslide susceptibility zonation of a hilly region: A quantitative approach. *Nat. Hazards Res.* **2024**, *4*, 75–86. [\[CrossRef\]](#)
3. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [\[CrossRef\]](#)
4. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [\[CrossRef\]](#)
5. Aleotti, P.; Chowdhury, R. Landslide hazard assessment: Summary review and new perspectives. *Bull. Eng. Geol. Environ.* **1999**, *58*, 21–44. [\[CrossRef\]](#)
6. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, A.-X.; Pei, X.; Duan, Z. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci. Total Environ.* **2018**, *626*, 1121–1135. [\[CrossRef\]](#)
7. Melati, D.N.; Umbara, R.P.; Astisiasari, A.; Wisyanto, W.; Trisnafiah, S.; Trinugroho, T.; Prawiradisastra, F.; Arifianti, Y.; Ramdhani, T.I.; Arifin, S.; et al. A comparative evaluation of landslide susceptibility mapping using machine learning-based methods in Bogor area of Indonesia. *Environ. Earth Sci.* **2024**, *83*, 86. [\[CrossRef\]](#)
8. Wang, X.; Nie, W.; Xie, W.; Zhang, Y. Incremental learning-random forest model-based landslide susceptibility analysis: A case of Ganzhou City, China. *Earth Sci. Inform.* **2024**, *17*, 1645–1661. [\[CrossRef\]](#)
9. Mondal, S.; Mandal, S. RS & GIS-based landslide susceptibility mapping of the Balason River basin, Darjeeling Himalaya, using logistic regression (LR) model. *Georisk Assess. Manag. Risk Eng. Syst. Geohazards* **2018**, *12*, 29–44.
10. Das, G.; Lepcha, K. Application of logistic regression (LR) and frequency ratio (FR) models for landslide susceptibility mapping in Relli Khola river basin of Darjeeling Himalaya, India. *SN Appl. Sci.* **2019**, *1*, 1453. [\[CrossRef\]](#)
11. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.; Akgun, A.; Tian, Y.; Liu, J.; Zhu, A.X.; Li, S. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4397–4419. [\[CrossRef\]](#)



12. Shahri, A.A.; Spross, J.; Johansson, F.; Larsson, S. Landslide susceptibility hazard map in southwest Sweden using artificial neural network. *Catena* **2019**, *183*, 104225. [\[CrossRef\]](#)
13. Xu, Z.; Che, A.; Zhou, H. Seismic landslide susceptibility assessment using principal component analysis and support vector machine. *Sci. Rep.* **2024**, *14*, 3734. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Rai, S.C.; Pandey, V.K.; Sharma, K.K.; Sharma, S. Landslide susceptibility analysis in the Bhilangana Basin (India) using GIS-based machine learning methods. *Geosyst. Geoenviron.* **2024**, *3*, 100253. [\[CrossRef\]](#)
15. Ciurleo, M.; Cascini, L.; Calvello, M. A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. *Eng. Geol.* **2017**, *223*, 71–81. [\[CrossRef\]](#)
16. Hakim, W.L.; Rezaie, F.; Nur, A.S.; Panahi, M.; Khosravi, K.; Lee, C.-W.; Lee, S. Convolutional neural network (CNN) with metaheuristic optimization algorithms for landslide susceptibility mapping in Icheon, South Korea. *J. Environ. Manag.* **2022**, *305*, 114367. [\[CrossRef\]](#)
17. Meng, S.; Shi, Z.; Li, G.; Peng, M.; Liu, L.; Zheng, H.; Zhou, C. A novel deep learning framework for landslide susceptibility assessment using improved deep belief networks with the intelligent optimization algorithm. *Comput. Geotech.* **2024**, *167*, 106106. [\[CrossRef\]](#)
18. Liu, R.; Yang, X.; Xu, C.; Wei, L.; Zeng, X. Comparative study of convolutional neural network and conventional machine learning methods for landslide susceptibility mapping. *Remote Sens.* **2022**, *14*, 321. [\[CrossRef\]](#)
19. Zhu, A.X.; Miao, Y.; Wang, R.; Zhu, T.; Deng, Y.; Liu, J.; Yang, L.; Qin, C.; Hong, H. A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. *Catena* **2018**, *166*, 317–327. [\[CrossRef\]](#)
20. Roy, J.; Saha, S. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenviron. Disasters* **2019**, *6*, 11. [\[CrossRef\]](#)
21. Bui, D.T.; Bui, Q.-T.; Nguyen, Q.-P.; Pradhan, B.; Nampak, H.; Trinh, P.T. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agric. For. Meteorol.* **2017**, *233*, 32–44.
22. Feng, R.; Zheng, H.-J.; Gao, H.; Zhang, A.-R.; Huang, C.; Zhang, J.-X.; Luo, K.; Fan, J.-R. Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China. *J. Clean. Prod.* **2019**, *231*, 1005–1015. [\[CrossRef\]](#)
23. Pham, B.T.; Van Phong, T.; Nguyen-Thoi, T.; Trinh, P.T.; Tran, Q.C.; Ho, L.S.; Singh, S.K.; Duyen, T.T.T.; Nguyen, L.T.; Le, H.Q.; et al. GIS-based ensemble soft computing models for landslide susceptibility mapping. *Adv. Space Res.* **2020**, *66*, 1303–1320. [\[CrossRef\]](#)
24. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378.
25. Colkesen, I.; Sahin, E.K.; Kavzoglu, T. Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J. Afr. Earth Sci.* **2016**, *118*, 53–64. [\[CrossRef\]](#)
26. Achour, Y.; Pourghasemi, H.R. How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geosci. Front.* **2020**, *11*, 871–883. [\[CrossRef\]](#)
27. Miao, Y.; Zhu, A.; Yang, L.; Bai, S.; Liu, J. A Method for Measuring the Credibility of Landslide Negative Samples Based on Geographic Environment Similarity. *Prog. Geogr.* **2016**, *35*, 860–869.
28. Chen, X.; Chen, W. GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. *Catena* **2021**, *196*, 104833. [\[CrossRef\]](#)
29. Bui, D.T.; Tuan, T.A.; Hoang, N.-D.; Thanh, N.Q.; Nguyen, D.B.; Van Liem, N.; Pradhan, B. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* **2017**, *14*, 447–458.
30. Varnes, D.J. *Landslide Hazard Zonation: A Review of Principles and Practice*; Natural Hazards; Education, Scientific and Cultural Organization: Paris, France, 1984; Volume 3.
31. Kornejady, A.; Ownegh, M.; Bahremand, A. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena* **2017**, *152*, 144–162. [\[CrossRef\]](#)
32. Yong, C.; Jinlong, D.; Fei, G.; Bin, T.; Tao, Z.; Hao, F.; Li, W.; Qinghua, Z. Review of landslide susceptibility assessment based on knowledge mapping. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 2399–2417. [\[CrossRef\]](#)
33. Barber, J.; Thompson, S. Multiple regression of cost data: Use of generalised linear models. *J. Health Serv. Res. Policy* **2004**, *9*, 197–204. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Chen, J.; Hu, P.; Jimenez-Ruiz, E.; Holter, O.M.; Antonyrajah, D.; Horrocks, I. OWL2Vec\*: Embedding of OWL ontologies. *Mach. Learn.* **2021**, *110*, 1813–1845. [\[CrossRef\]](#)
35. Srinivasan, S. *Guide to Big Data Applications*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017.
36. Boulder Atomic Clock Optical Network (BACON) Collaboration\*. Frequency ratio measurements at 18-digit accuracy using an optical clock network. *Nature* **2021**, *591*, 564–569. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Gelb, J.; Apparicio, P. Temporal Network Kernel Density Estimation. *Geogr. Anal.* **2024**, *56*, 62–78. [\[CrossRef\]](#)
38. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [\[CrossRef\]](#)

39. Wu, Y.; Ke, Y.; Chen, Z.; Liang, S.; Zhao, H.; Hong, H. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena* **2020**, *187*, 104396. [[CrossRef](#)]
40. Pham, B.T.; Pradhan, B.; Bui, D.T.; Prakash, I.; Dholakia, M.B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [[CrossRef](#)]
41. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [[CrossRef](#)]
42. Chen, J.; Yang, S.T.; Li, H.W.; Zhang, B.; Lv, J.R. Research on Geographical Environment Unit Division Based on the Method of Natural Breaks (Jenks). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *40*, 47–50. [[CrossRef](#)]
43. Pradhan, B.; Lee, S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* **2010**, *25*, 747–759. [[CrossRef](#)]
44. He, S.; Pan, P.; Dai, L.; Wang, H.; Liu, J. Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. *Geomorphology* **2012**, *171–172*, 30–41. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.