



Article

Improving Landslide Prediction: Innovative Modeling and Evaluation of Landslide Scenario with Knowledge Graph Embedding

Luanjie Chen ^{1,2} , Ling Peng ^{1,2} and Lina Yang ^{1,2,*}

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100049, China; chenluanjie20@mails.ucas.ac.cn (L.C.); pengling@aircas.ac.cn (L.P.)

² College of Resources and Environment (CRE), University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: yangln@aircas.ac.cn

Abstract: The increasing frequency and magnitude of landslides underscore the growing importance of landslide prediction in light of factors like climate change. Traditional methods, including physics-based methods and empirical methods, are beset by high costs and a reliance on expert knowledge. With the advancement of remote sensing and machine learning, data-driven methods have emerged as the mainstream in landslide prediction. Despite their strong generalization capabilities and efficiency, data-driven methods suffer from the loss of semantic information during training due to their reliance on a ‘sequence’ modeling method for landslide scenarios, which impacts their predictive accuracy. An innovative method for landslide prediction is proposed in this paper. In this paper, we propose an innovative landslide prediction method. This method designs the NADE ontology as the schema layer and constructs the data layer of the knowledge graph, utilizing tile lists, landslide inventory, and environmental data to enhance the representation of complex landslide scenarios. Furthermore, the transformation of the landslide prediction task into a link prediction task is carried out, and a knowledge graph embedding model is trained to achieve landslide predictions. Experimental results demonstrate that the method improves the F1 score by 5% in scenarios with complete datasets and 17% in scenarios with sparse datasets compared to data-driven methods. Additionally, the application of the knowledge graph embedding model is utilized to generate susceptibility maps, and an analysis of the effectiveness of entity embeddings is conducted, highlighting the potential of knowledge graph embeddings in disaster management.

Keywords: landslide prediction; knowledge graph; ontology; remote sensing application



Citation: Chen, L.; Peng, L.; Yang, L. Improving Landslide Prediction: Innovative Modeling and Evaluation of Landslide Scenario with Knowledge Graph Embedding. *Remote Sens.* **2024**, *16*, 145. <https://doi.org/10.3390/rs16010145>

Academic Editors: Constantinos Loupasakis, Ioannis Papoutsis and Konstantinos G. Nikolakopoulos

Received: 8 December 2023

Revised: 24 December 2023

Accepted: 25 December 2023

Published: 29 December 2023



Copyright: © 2022 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

In recent years, the frequency and magnitude of landslide disasters have been on the rise, attributable to factors including climate change, human activities, and geological processes. These developments pose a substantial threat to human lives and property safety [1–3]. Consequently, the timely and effective prediction of landslide occurrence times, locations, and scales is of paramount significance in mitigating the losses stemming from landslide disasters. Traditional landslide prediction methods encompass physics-based methods and empirical methods. However, these methods have various issues, including high demands for data accuracy and model builder expertise [4–6], expensive modeling and computational costs [7,8], and a significant reliance on expert experience [9–13]. These issues collectively result in lower efficiency in landslide prediction.

With the development of remote sensing and artificial intelligence technologies, data-driven methods have gradually become the mainstream for predicting landslides. Data-driven methods typically select landslide conditioning factors (LCFs) that affect landslide occurrence as input variables. They use known landslide occurrences or non-occurrences

as labels to train models capable of landslide prediction. These data-driven methods can adaptively adjust model parameters using substantial historical data, resulting in improved model generalization and robustness. In contrast to empirical methods, data-driven methods can efficiently handle data with multiple attributes, such as geology, landform, and climate. These datasets are often high-dimensional, but data-driven methods can process them more swiftly, thereby enhancing the landslide prediction efficiency. Commonly used data-driven methods today include support vector machines (SVMs), artificial neural networks (ANNs), and random forest (RF). These methods have been successfully applied in various landslide cases with promising results [14–16]. Nevertheless, data-driven landslide prediction methods also have their limitations.

An important challenge lies in the oversimplification of landslide scenarios by data-driven methods. The occurrence of landslides is a multifaceted process shaped by the interplay of various geographical elements. For instance, geological formations, soil compositions, and vegetation cover within distinct geographical settings may collectively affect landslide occurrence. These effects can manifest in explicit or implicit ways. Explicit relations are those that can be distinctly articulated, such as the greater likelihood of landslides in areas with steep slopes or the influence of topography on the flow and distribution of precipitation. Implicit relations, conversely, pertain to connections that defy precise definition. Take, for example, the impact of certain human activities on landslide risk. While both human activities and landslides can be documented, elucidating the precise mechanisms through which human activities influence landslides remains challenging.

We describe these interactions in the geographic environment, whether explicit or implicit, as ‘semantic information in the geographic environment’. Explicit relations are challenging for data-driven methods to express because, when predicting landslides, data-driven methods typically use grid cells as prediction units [4,17]. They rely on generating a landslide dataset with a ‘sequence’ structure for each grid cell to train the model and predict landslides, treating each piece of data as an independent sample. Modeling landslide scenarios based on this ‘sequence’ structure makes it difficult to capture the relation between different grid cells. On the other hand, data-driven methods have difficulty in capturing implicit relations in semantic information. For the same reason, this is because the sequence structure in the data-driven method cannot learn the relations between sequences during the training process. The data-driven method loses semantic information during modeling and training, which reduces the accuracy of landslide prediction. In short, this simplified modeling method cannot adequately represent the complexity of landslide scenarios, which in turn has an impact on the accurate prediction of landslides.

To address this issue, we propose the modeling of landslide scenarios based on a knowledge graph [18–21]. The “graph” structure within the knowledge graph can more directly represent the explicit relations between each LCF. It is also more conducive to uncovering the implicit relations between grid cells and LCFs, allowing us to discover spatial patterns in the landslide process. Figure 1 illustrates the contrast between the data-driven method and the knowledge graph method for modeling landslide scenarios. It is evident from the figure that modeling landslide scenarios using the “graph” structure outperforms the “sequence” structure, particularly in capturing semantic information.

Furthermore, we advocate performing landslide prediction based on knowledge graph embedding (KGE) [1,2,19]. KGE assigns semantic interpretations to the vectors of entities and relations within the knowledge graph by learning semantic associations in the vector space. This means that similar entities and relations are also similar in the vector space. In some applications, knowledge graph embedding (KGE) has demonstrated the ability to capture complex semantic relationships [22,23] and improve performance under conditions of data scarcity [24,25]. For landslide prediction, compared to data-driven methods, KGE can automatically learn the influence of local context factors (LCFs) in the vector space based on entities and relations. KGE effectively maps multi-source data into a vector space, enhancing its capability to capture the implicit relations between LCFs. Additionally, in cases of sparse datasets, KGE can infer relations and patterns within the limited data, thus

filling in gaps and improving the accuracy of landslide prediction methods. Consequently, it contributes to the enhancement of the precision and generalization ability of landslide prediction methods.

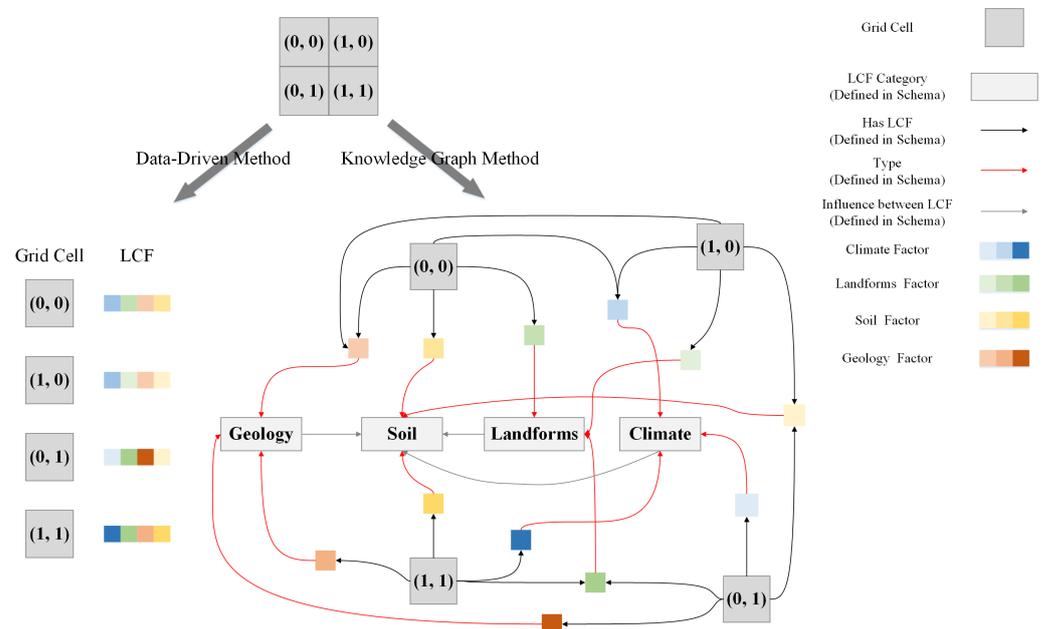


Figure 1. Contrasting the modeling of landslide scenarios with the data-driven method and the knowledge graph method. The data-driven method represents LCFs as sequences, simplifying the relation between grid cells and LCFs. In contrast, the knowledge graph method represents LCFs and grid cells using a graph structure, which is better suited for exploring the relation between grid cells and LCFs.

In this study, we present a comprehensive approach to construct a knowledge graph tailored for landslide prediction, effectively transforming this task into a graph-based link prediction problem using KGE techniques. The paper is structured as follows: Section 2 introduces the study area and data sources. Section 3 details the entire process of landslide prediction using knowledge graph techniques, encompassing data preprocessing, knowledge graph creation, and the application of KGE for prediction. Section 4 offers a comparative analysis with a generic landslide prediction model and showcases our prediction results. Section 5 delves into the strengths and limitations of employing KGE in landslide prediction. Finally, Section 6 concludes the paper, summarizing our findings and contributions.

2. Study Area and Data

Xiji County is situated in the southern part of Guyuan City, within the Ningxia Province, China. It is positioned between longitude 105°20′–106°04′E and latitude 35°35′–36°14′N and is geographically proximate to the western foothills of the Liupan Mountains. Xiji County is situated in the Loess Plateau, characterized by an arid hilly landscape. The terrain encompasses Hulu River plains, loess hills and gullies, and soil and rocky mountains. The elevation gradually increases from south to north, spanning from 1688 to 2633 meters. The susceptibility to loess landslides in Xiji County is attributed to its rugged terrain and narrow ridges. The combination of these geographical factors creates conditions conducive to landslide occurrences, making the region a suitable area for the validation of landslide prediction methods. To support our study, we gathered 741 landslide records from the Ningxia Remote Sensing Survey and Mapping Institute, forming the basis for constructing the knowledge graph. The study area's details are depicted in Figure 2.

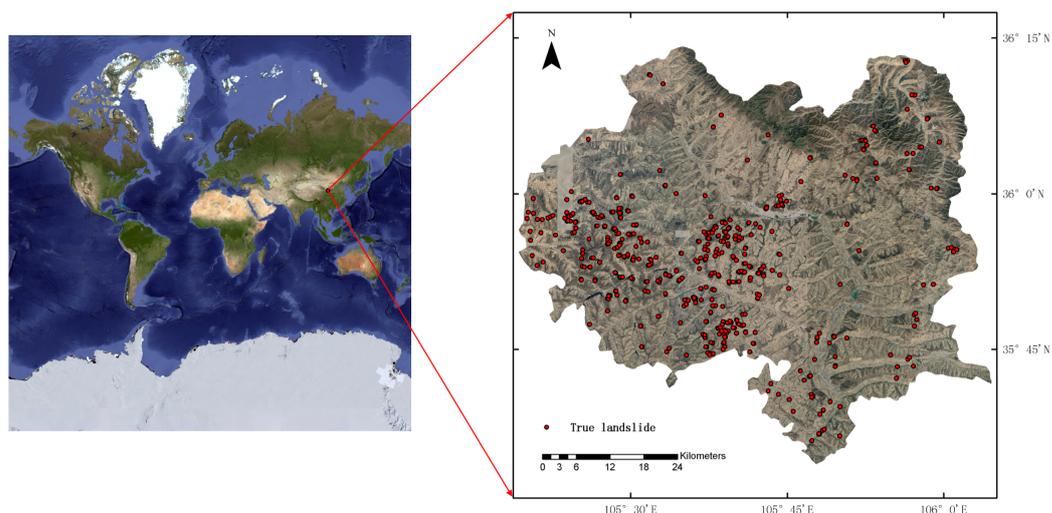


Figure 2. Overview of the study area located in Xiji County, Ningxia Province, China; 741 landslide records were selected to produce landslide inventory.

The environmental data obtained from the study area consist of seven categories: geology, landform, soil, climate, vegetation, transportation systems, and population. These data sources are multi-sourced. We supplemented the non-public data, provided by the Ningxia Remote Sensing Survey and Mapping Institute, with terrain, precipitation, and road data from various public datasets. Each data category comprises multiple LCFs, and all LCFs from the environmental data are recorded in the schema layer of the knowledge graph. Table 1 displays the LCFs within each environmental data category, and Table 2 provides detailed information about each category of environmental data. It is worth noting that the time span of these data coincides with the time range of the landslide records.

Table 1. LCF statistics for different environmental data categories.

Category	LCF
Geology	Sedimentary stratum thickness; sedimentary stratum unit symbol; sedimentary mineral species; sedimentary biota assemblage; sedimentary rock structure; metamorphic rock texture; metamorphic rock name; metamorphic rock color; protolith; geological contact; fault property; fault structural type; occurrence type; occurrence dip angle; occurrence strike; fossil-bearing lithostratigraphic horizon; fossil category; total drilling depth; drilling type
Landform	Aspect; elevation; slope; surface roughness; curvature; hydrology; body of water; land use and land cover (LULC)
Soil	Soil texture; soil type
Climate	Precipitation
Vegetation	Normalized difference vegetation index (NDVI)
Transportation	Road type
Population	Population distribution

Table 2. Sources and details of the environment data.

Category	Name	Format	Source
Geology	Basic geological data	Vector polygon and vector point	Ningxia Remote Sensing Survey and Mapping Institute
Landform	DEM terrain data	30 m raster data	Shuttle radar topography mission DEM [26,27]
	Hydrological data	Vector polygon and vector line	OSM hydrological data [28], Ningxia Remote Sensing Survey and Mapping Institute
	LULC data	Vector polygon	Ningxia Remote Sensing Survey and Mapping Institute
Soil	Soil data	1 km raster data	Resource and environment science and data center [29]
Climate	Normalized precipitation	30 m raster data	Ningxia Remote Sensing Survey and Mapping Institute
Vegetation	NDVI data	1 m raster data data	Ningxia Remote Sensing Survey and Mapping Institute
Transportation	Road data	Vector line	OSM highway data [30]
Population	Population distribution data	CSV	Ningxia Remote Sensing Survey and Mapping Institute

3. Methodology

The knowledge graph-based landslide prediction method comprises three stages, as shown in Figure 3. Initially, data from various sources are gathered and subjected to preprocessing to create a data collection consisting of a tile list, landslide inventory, and environmental data, all in a standardized format. The tile list is a record of coordinates covering the study area, where each tile functions as a grid cell for landslide prediction. The landslide inventory includes details of past landslide events in the study area, encompassing factors like location, magnitude, landslide type, and the resulting impact. This inventory forms the foundation for landslide susceptibility and risk assessment [4,31,32]. Next, the data in this uniform format are transformed into a collection of triples, an example of which is illustrated in Figure 4. A triple is the fundamental unit of the knowledge graph, denoted by (h, r, t) , where h represents the head entity, t represents the tail entity, and r signifies the relation between them. These triples are then used to construct the knowledge graph according to the designed schema. Finally, the knowledge graph undergoes embedded representation learning, while the landslide prediction task is redefined as a graph-based prediction task, enabling the assessment of susceptibility within the study area through link prediction.

3.1. Preprocessing

The purpose of preprocessing is to standardize the format of heterogeneous data from various sources and to create a collection of tile coordinates for the study area. Using tiles as the mapping unit for landslide prediction offers the advantage of adaptability to various geographic scales and efficient processing. During the preprocessing stage, a tile collection is created based on the specified tile level. Specifically, the study area's location is represented in tile coordinates following the Web Mercator rule [33]. The conversion rules between real latitude and longitude coordinates and tile coordinates are as follows:

$$x = \frac{lon + 180}{360} \times 2^z \quad (1)$$

$$y = \left(1 - \frac{\ln(\tan(lat \times \frac{\pi}{180}) + \frac{1}{\cos(lat \times \frac{\pi}{180})})}{\pi}\right) \times 2^{z-1} \quad (2)$$

where *lon* and *lat* denote the input longitude and latitude coordinates, respectively. The transformed horizontal and vertical coordinates of the tile are denoted by *x* and *y*, while *z* denotes the zoom level of the tile. Each tile corresponds to a specific set of longitude and latitude coordinates, and the quantity of coordinates varies depending on the zoom level. Notably, higher zoom levels lead to tiles with fewer coordinates, enhancing spatial accuracy at the expense of increased computational complexity. In this paper, we select zoom level 18, as illustrated in Figure 5, to strike a balance between spatial accuracy and computational complexity. With level 18, the study area comprises a total of 205,330 tiles.

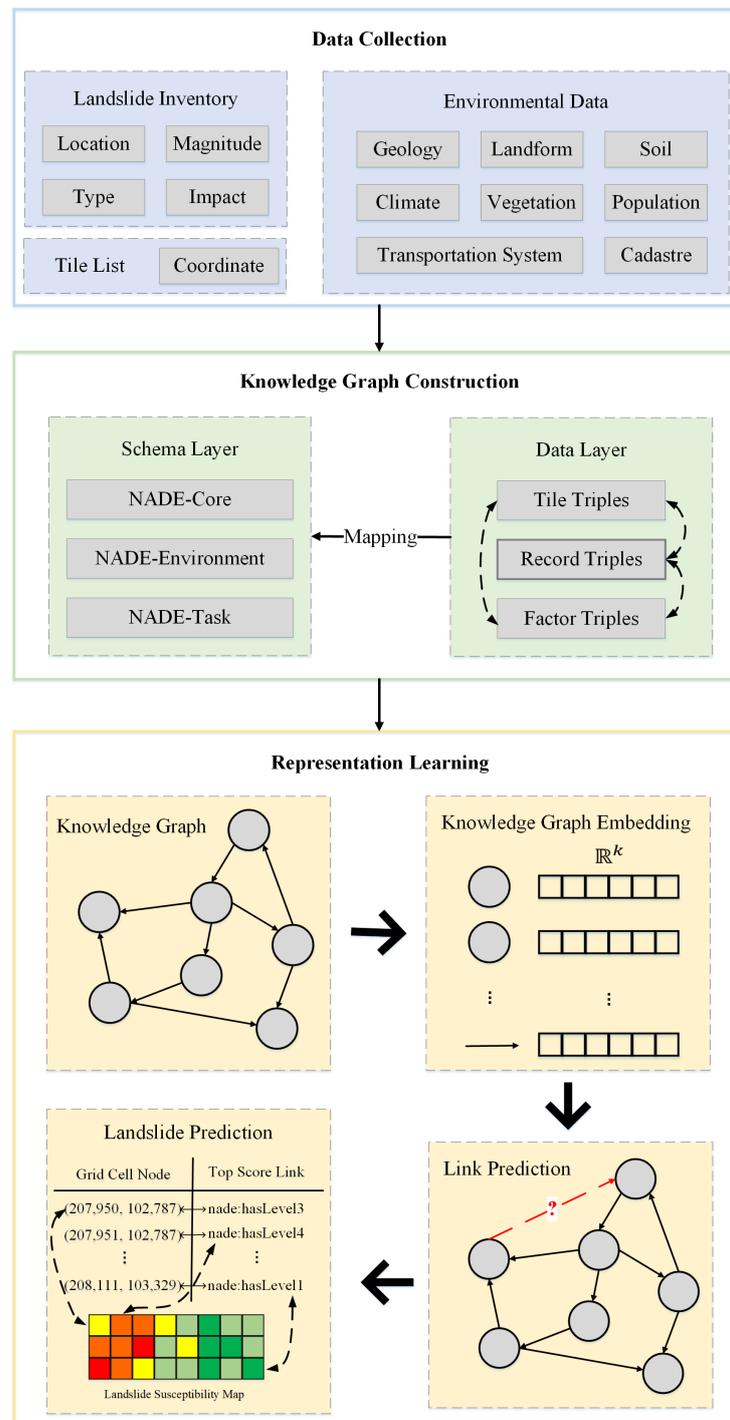


Figure 3. Workflow for landslide task based on knowledge graph representation learning. The “?” in the figure represents the attributes of the edges that need to be predicted in link prediction.

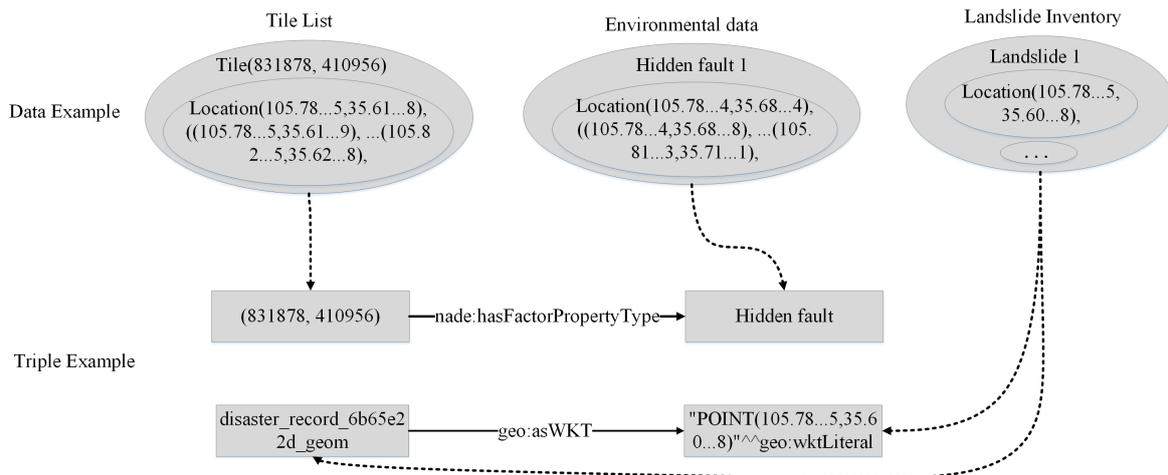


Figure 4. An example illustrating the transformation of tile list, environmental data, and landslide inventory into triples. In this example, a factor triple describing the location of the hidden fault is generated based on the latitude and longitude set contained in the environmental attribute “Hidden Fault” and the coordinates of the tile in which this set is located. A record triple is generated based on the latitude and longitude coordinates of the landslide location.

Then, the format of the heterogeneous data from multiple sources is harmonized, and an indexed list of data attribute values and tiles is generated. The data structure primarily comprises vector points (e.g., landslide records), vector lines (e.g., hydrological data), vector surfaces (e.g., geological data), raster data at various scales (e.g., terrain data), and CSV files (e.g., population distribution data). Initially, the geographic coordinate systems of these data are standardized. Subsequently, we extract the corresponding attribute values from each tile. For discrete attribute values (e.g., fault types with a limited number of values like normal fault, reverse fault, strike-slip fault, and hidden fault), it is relatively straightforward to create an index list for each tile based on these attribute values. However, for continuous attribute values (e.g., elevation, which is continuous), we first discretize the attribute values by assessing the data and selecting the appropriate scale, and then generate an indexed list for each tile using these discretized values of the attribute.

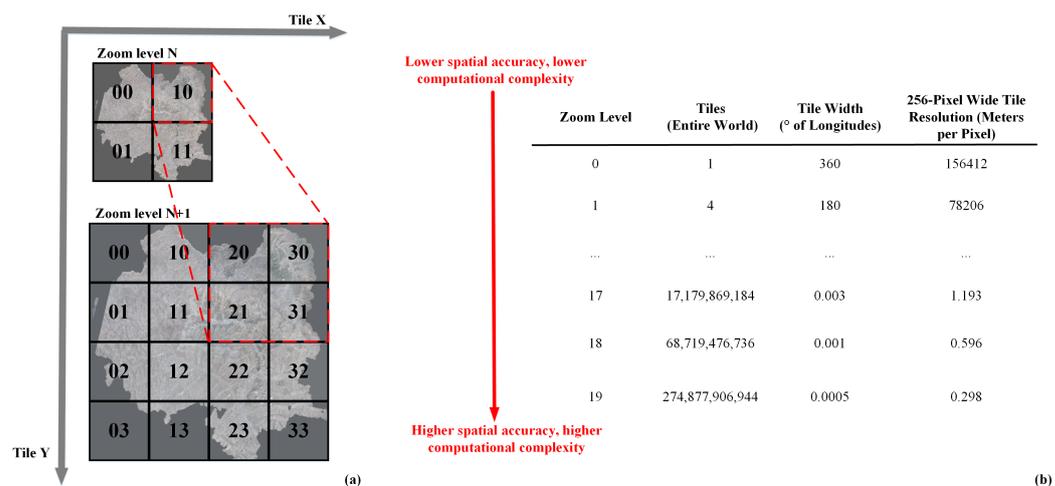


Figure 5. Differences in the tiles at various zoom levels: (a) the changes in tile distribution across zoom levels, with higher levels indicating increased spatial precision; (b) tile sizes at different zoom levels. Higher tile levels enhance spatial accuracy but also raise computational complexity.

3.2. Knowledge Graph Construction

3.2.1. Schema Layer

The schema layer defines the structure and specifications of the concepts within the knowledge graph. It serves as a metadata model for describing the relations between entities in the knowledge graph, including their attributes. The schema layer establishes a unified semantic framework for the data in the knowledge graph. This framework enhances the organization, query capabilities, interpretability, and aids in reasoning computations. In this paper, the schema layer for describing disaster scenarios is composed of a basic vocabulary and an ontology that defines concepts related to disasters, as shown in Figure 6.

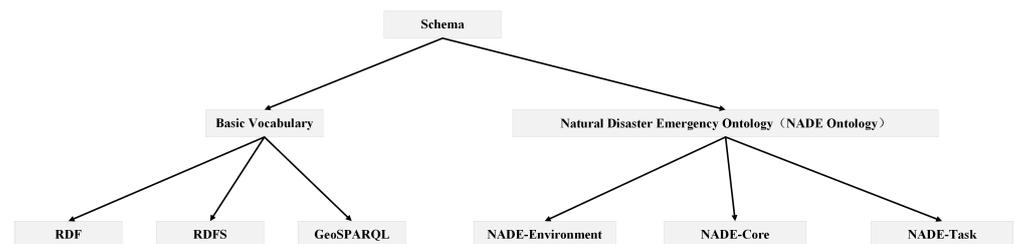


Figure 6. Components of the schema layer, including basic vocabulary and NADE ontology. Basic vocabulary includes vocabularies of RDF, RDFS, GeoSPARQL, and NADE consists of NADE-Core, NADE-Environment, and NADE-Task.

We utilize the resource description framework (RDF) [34] and resource description framework schema (RDFS) [35] vocabularies to establish the foundational terminology within the knowledge graph. RDF is a standard designed for representing relations among resources in the semantic web. RDFS, an extension of RDF, is responsible for defining more intricate hierarchies between resources. RDF/RDFS encapsulates knowledge within triples, each composed of subject, predicate, and object. In RDF/RDFS triples, the subject corresponds to the head entity within the knowledge graph triple, the object corresponds to the tail entity, and the predicate signifies the relation, as illustrated in Figure 7.

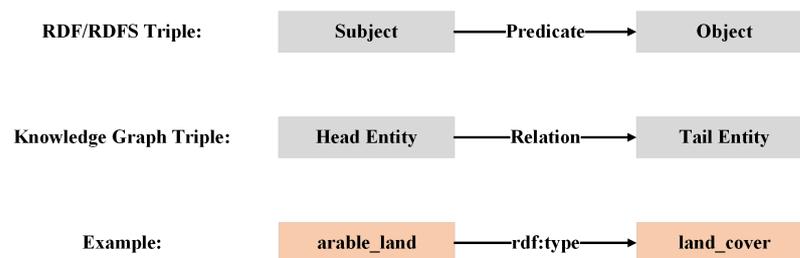


Figure 7. Correspondence between triple in RDF/RDFS and triple in knowledge graph.

The RDF/RDFS vocabularies employed in this paper primarily encompass: `rdf:type`, `rdf:subject`, `rdf:predicate`, `rdf:object`, `rdfs:Class`, and `rdfs:subClassOf`. To enhance the clarity of our proposed approach, we omitted the resource prefixes and retained only the relation prefixes in this paper. The utilization of RDF/RDFS vocabularies offers a fundamental and standardized method for data description. Additionally, the schema layer constructed using RDF/RDFS enables more effective comprehension and processing of heterogeneous disaster data from various sources.

Furthermore, we utilize the GeoSPARQL glossary [36,37] to define the fundamental spatial information vocabulary for disaster scenarios. The GeoSPARQL glossary encompasses a comprehensive set of terms designed for representing geospatial information. These terms facilitate the description of various spatial aspects, including geographical coordinates, types of geographic elements (e.g., points, lines, and polygon), and geospatial relations (e.g., intersects, overlaps, disjoint, etc.) within disaster records, disaster-related tasks, and LCFs in the study area. GeoSPARQL offers a standardized approach for integrat-

ing and sharing geospatial data. Utilizing the GeoSPARQL terminology ensures a consistent interpretation and utilization of disaster data from diverse heterogeneous sources.

Ontology [38] is a formal knowledge representation for describing concepts, entities, attributes, and relations within a domain. The goal of ontology is to capture consensus and semantics within a domain, enabling different systems to share and comprehend domain-specific knowledge. Leveraging the expressive advantages of ontology, we designed and implemented the natural disaster emergency ontology (NADE ontology). NADE is capable of representing the semantics among landslide data while also supporting extensions to other disaster domains. Given the characteristics of disaster-related data, we subdivided the NADE ontology into three parts: NADE-Core, which delineates the fundamental concepts of disasters; NADE-Environment, which describes the disaster environment; and NADE-Task, which outlines the disaster tasks. For each term in the NADE ontology, we referenced relevant disaster emergency standards and existing disaster ontology definitions when formulating their definitions. Additionally, we considered the practical aspects of disaster emergency task handling.

NADE-Core establishes the essential vocabulary for describing disasters and forms the foundation of the NADE ontology. For instance, at any stage of a disaster, fundamental attributes such as the disaster type, its current phase, the affected objects, and the resulting impacts must be described. NADE-Environment defines the concept of the environment within a disaster scenario. Environmental changes are often the root causes of disasters, and social environmental factors directly influence the extent of damage caused. Providing a unified description of the environment in a disaster scenario helps identify patterns in disaster occurrences within a specific region. NADE-Task specifies the indicators of a disaster task, encompassing elements such as risk, hazard, impact, severity, likelihood, susceptibility, exposure, vulnerability, and their relations. Figure 8 illustrates the core vocabulary and primary relations defined by the NADE ontology.

3.2.2. Data Layer

The data layer converts the data generated in the preprocessing stage into triples. These triples include tile triples, record triples, and factor triples. Tile triples describe the positional relation between tiles, for example: ((831,878, 410,956), nade:hasAdjacentTile, (831,878, 410,957)). Record triples are generated by combining tile entities and landslide record entities. Each landslide record entity represents information about landslide events in the study area, for example: (disaster_record_6b65e22d, nade:hasDisasterType, landslide). With the integration of the GeoSPARQL ontology from the schema layer, record triples can provide detailed spatial information about landslide events, for example: (disaster_record_6b65e22d_geom, rdf:type, point). Factor triples are created by combining tile entities and environment entities, offering descriptions of the environmental properties in each tile in the study area, for example, for discrete factor values, generate triple((831,878, 410,956), nade:hasFactorPropertyType, hidden_fault). For continuous factor values, scale them into discrete categories, and then generate triples, for example, ((831,878, 410,956), nade:hasFactorPropertyType, altitude_range_3). When combined with the NADE ontology in the schema layer, the tile triples, record triples, and factor triples illustrate the relations among multiple sources of heterogeneous data within the disaster scenario, which is essential for effective landslide hazard modeling in the study area.

In the prediction phase, the entities and relations to be forecasted within the knowledge graph are also represented as triples in the data layer. For instance, consider the triple (disaster_task_e3a9851a, nade:hasLevel2, susceptibility), where 'disaster_task_e3a9851a' denotes the landslide tasks to be predicted, 'susceptibility' denotes the susceptibility level, and 'nade:hasLevel2' denotes that the task with the ID 'e3a9851a' is associated with the susceptibility level 2. Each landslide record and task possesses a unique ID linked to an individual tile. Since tiles serve as the mapping units for landslide prediction, the core of knowledge graph-based landslide susceptibility prediction lies in forecasting the susceptibility level of each tile's associated landslide task.

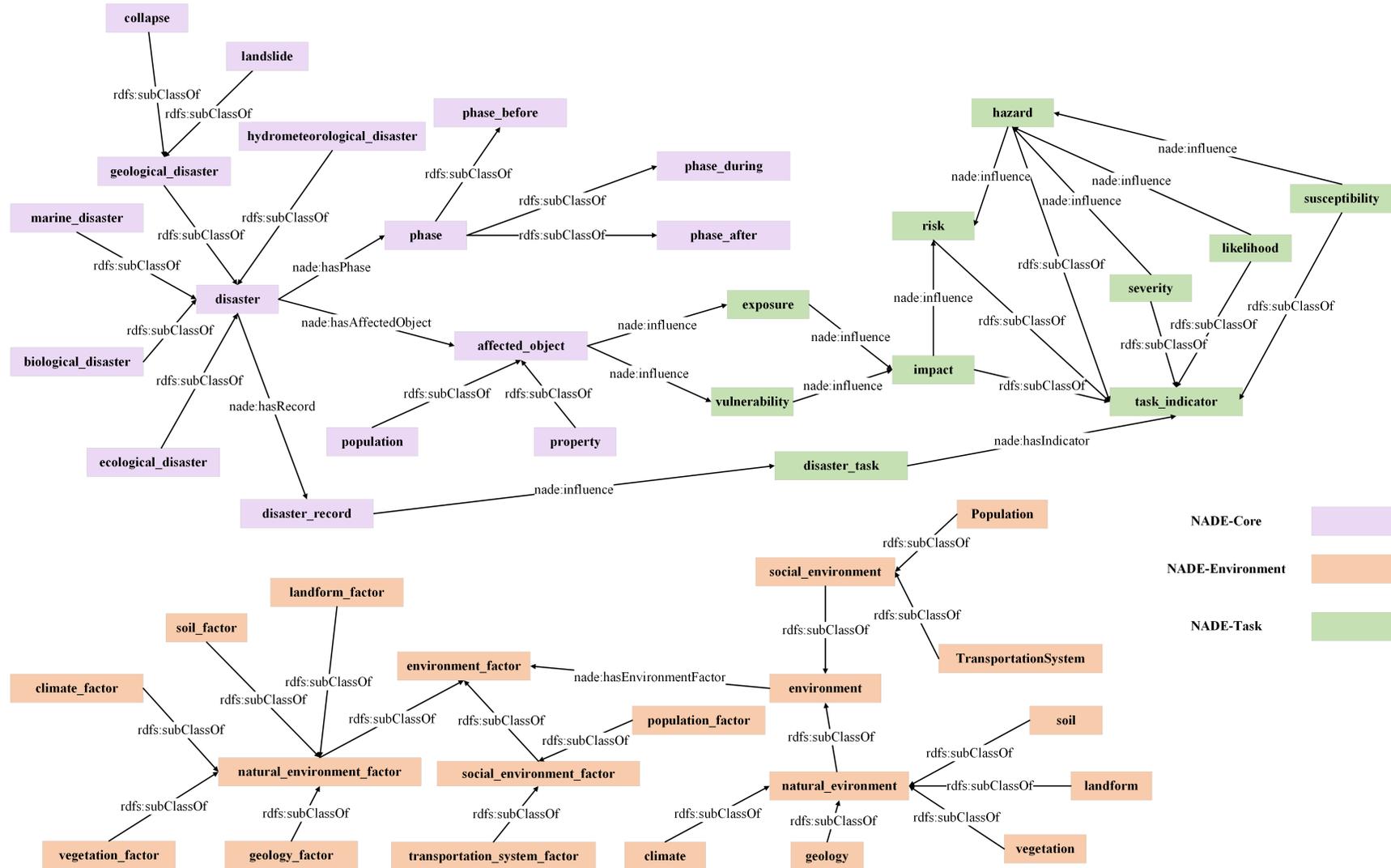


Figure 8. The schema layer definition based on NADE, including concepts and relations, consists of three parts: NADE-Core, NADE-Environment, and NADE-Task.

After creating the triples in the data layer, these triples are mapped to the schema layer to generate the knowledge graph. Figure 9 illustrates the connection between the data layer and schema layer. This connection primarily involves mapping tile triples to the NADE-Environment (i.e., factor triples), mapping record triples to NADE-Core, mapping record triples to GeoSPARQL, and particularly, establishing links between task triples and NADE-Task during the prediction phase.

3.3. Knowledge Graph Embedding

KGE is a representation learning technique that aims to map information about entities, relations, and attributes in a knowledge graph into a continuous vector space to better capture implicit relations between entities. KGE can measure the semantic similarity between entities and the semantic associations of relations in a vector space. For instance, if two entities share similar relations in the knowledge graph, their vector representations will be closer in the embedded vector space. KGE models typically learn the vector representations of entities and relations by minimizing or maximizing a loss function, preserving semantic relations between entities in the knowledge graph within the embedded vector space.

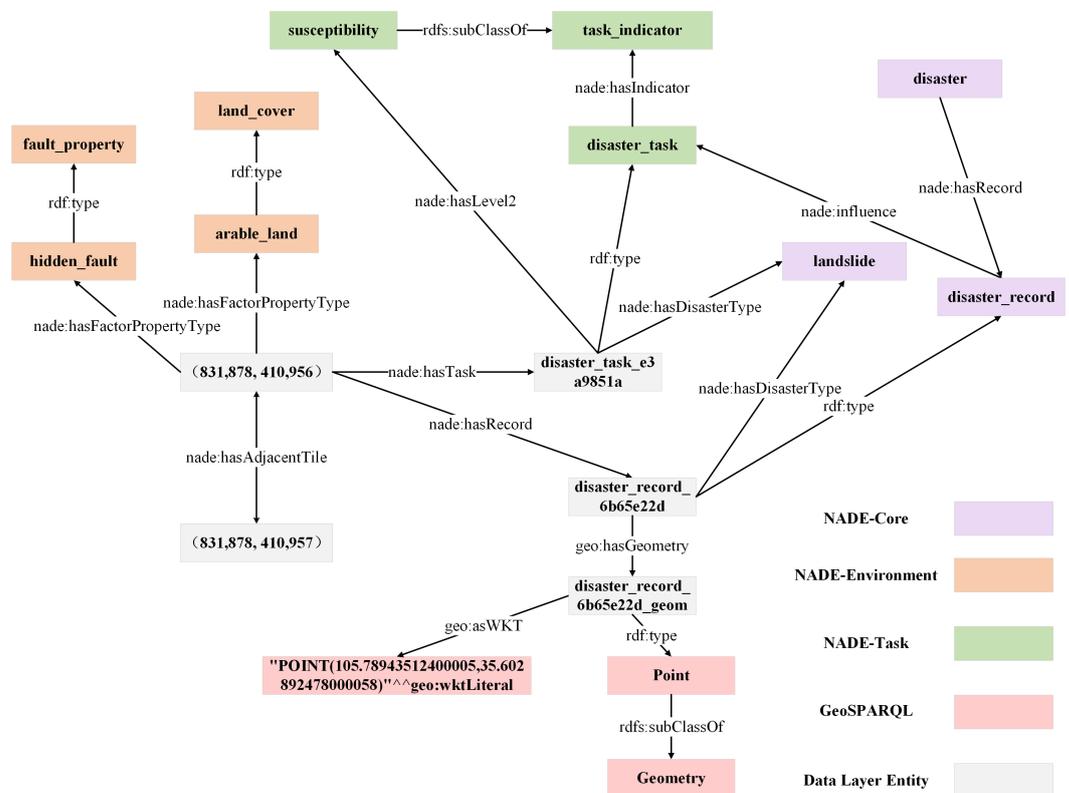


Figure 9. The connection between the data layer and schema layer, showcasing examples of a landslide record and a landslide prediction task under a tile. "" is the delimiter in GeoSPAEQL used to define coordinate properties.

3.3.1. Task Formalization

In this paper, we transform the landslide prediction task into a knowledge graph-based link prediction task. Link prediction involves predicting potential connections within a knowledge graph by analyzing the information associated with existing nodes and edges. It aims to determine the probable relation, denoted by r , between a given pair of entities, represented as (h, t) :

$$Score(h, r, t) = f(e_h, e_r, e_t) \tag{3}$$

where the vectors of entities h and t are denoted by e_h and e_t , respectively, and the vector of relations r is denoted by e_r . The $Score(h, r, t)$ denotes the score of the relation r between

the entity pairs (h, t) ; if the value of $Score(h, r, t)$ is larger, it means that it is more likely that the relation r exists between the entity pairs. The function f denotes a function that maps entity vectors and relation vectors to scores, the exact form of which can be chosen according to different KGE models, e.g., using the L_2 norm [39].

The aim of link prediction is to forecast new relations that may exist within the knowledge graph but have not been revealed yet, based on known entities and relations. In the context of landslide prediction, this entails forecasting the level of the landslide task indicator. For instance, when predicting landslide susceptibility for a grid cell with task number 'e3a9851a', the objective of link prediction is to anticipate which level of susceptibility is more likely to be associated with a pair of entities: a head entity of 'disaster_task_e3a9851a' and a tail entity of 'susceptibility'.

3.3.2. Model Training

The training of the KGE model first involves ternaries that are present in the knowledge graph, i.e., positive samples. The goal of the model is to embed these positive samples into the vector space in order to capture the semantic information between entities and relations through vector operations. On the other hand, negative samples need to be introduced to train the KGE model and enhance its ability to recognize triples that do not exist in the knowledge graph. For each positive sample (h, r, t) , a head entity h or a tail entity t is randomly selected and replaced by an irrelevant head entity h' or a tail entity t' , thus generating a negative sample (h', r, t) or (h, r, t') . The positive and negative samples are used together to train the KGE model, with the training objective being to minimize the embedding distance of the positive samples while maximizing the embedding distance of the negative samples. This process enhances the model's ability to understand and represent the knowledge graph.

In this paper, five typical KGE models are used for training and their score functions are shown in Table 3. The loss functions defined in the training are as follows:

$$Loss = \sum_{(h,r,t) \in \mathbb{D}^+} \sum_{(h',r',t') \in \mathbb{D}^-} \max(0, \gamma - Score(h, r, t) + Score(h', r', t')) \quad (4)$$

where $Loss$ denotes the loss function, which is the objective function we aim to minimize during training. The triple (h, r, t) denotes a positive triple, and (h', r', t') denotes a negative triple. \mathbb{D}^+ and \mathbb{D}^- denote the sets of positive and negative triples, respectively. γ denotes the margin used to ensure a minimum score difference between positive and negative triples.

Table 3. Typical KGE models and their score functions

Model	$Score(h, r, t)$
TransE [39]	$-\ h + r - t\ $
RESICAL [40]	$h^\top M_r t$
DistMult [41]	$h^\top \text{diag}(r) t$
ComplEx [42]	$h^\top \text{Re}(\text{diag}(r) t)$
RotateE [43]	$\ h \circ r - t\ $

3.3.3. Prediction

In this paper, landslide susceptibility assessment is employed as an illustrative example for landslide prediction. Initially, the KGE-based link prediction model generates scores for various landslide susceptibility levels on each grid cell, representing the likelihood of each grid cell belonging to different susceptibility classes. Subsequently, the highest scoring susceptibility class is selected for each grid cell, thereby determining the susceptibility result for that particular location. Finally, the susceptibility results for each grid cell are aggregated to produce a susceptibility map for the entire study area.

4. Results

4.1. Experimental Settings

4.1.1. Metrics

We use *Precision*, *Recall*, and F_1 to evaluate the effectiveness of the landslide prediction model with the following equations:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F_1 = \frac{2PrecisionRecall}{Precision + Recall} \quad (7)$$

where TP denotes true positives, representing instances correctly predicted as positive; FP denotes false positives, representing instances incorrectly predicted as positive when they are actually negative; TN denotes true negatives, representing instances correctly predicted as negative; FN denotes false negatives, representing instances incorrectly predicted as negative when they are actually positive.

Additionally, we utilize four widely adopted evaluation metrics in knowledge graph embedding (KGE) to rigorously evaluate the effectiveness of our link prediction model. Mean reciprocal rank (MRR) serves as a crucial metric, assessing the quality of the item ranking by taking into account the reciprocal of the rank at which the first relevant item is located. MRR provides insights into how efficiently the model prioritizes relevant items within the overall ranking. Hits@1, Hits@3, and Hits@10 offer further granularity in evaluating the model's performance. Hits@1 measures the model's ability to correctly identify relevant items at the very top of the ranked outcomes. Hits@3 and Hits@10 extend this assessment to the top-3 and top-10 ranked outcomes, respectively. These metrics provide a nuanced understanding of the model's capacity to accurately predict and include relevant items within different segments of the ranking. Collectively, these evaluation metrics contribute to a thorough and multifaceted evaluation of our model's proficiency in addressing landslide prediction tasks.

4.1.2. Implementation Detail

We used Dglke [44], an efficient tool for knowledge graph embedding, to conduct our experiments. The learning rate for ComplEx is set to 0.02, while the rest of the models are set to 0.01. The batch size is fixed at 2048, and the embedding dimensions are set to 400. For each positive sample in training, 256 negative samples are generated through adversarial sampling. In a single process, the maximum number of training steps is set at 3000, with the total number of training steps limited to 12,000. The intervals γ for TransE, RESCAL, DistMult, ComplEx, and ROTATE are set to 30, 12, 25, 1, and 3, respectively, with a regularization factor of 1.00×10^{-8} . We utilized four NVIDIA Titan Xp graphics cards to accelerate our training process in the experiments.

4.2. Prediction Result

To demonstrate the advantages of knowledge graph modeling over general data-driven methods, we selected commonly used machine learning models for landslide prediction to make comparisons, including SVM [45], RF [46], KNN [47], and GCF [48]. First, we generated positive samples using landslide records and environmental data, and negative samples were created with a positive-to-negative sample ratio (β) of 1.5. Next, we divided the dataset based on these generated samples and utilized the training set to train the machine learning model. Finally, we tested both the machine learning model and the trained knowledge graph embedding model on the test set, and the results of the landslide prediction are presented in Table 4. Furthermore, in our experiments, we paid attention

to the comparison of the efficiency between data-driven methods and knowledge graph embedding (KGE) methods. Generally, during the model training process, KGE methods tend to have a larger data volume and higher training costs compared to data-driven methods, as KGE methods involve more explicit relationships and need to learn deeper semantic information. We consider this to be a trade-off for achieving precision advantages. However, it is noteworthy that all tested models under the conditions of this experiment can complete training within 3 h, which we deem an acceptable range.

In the process of link prediction for knowledge graphs, we employ common evaluation metrics used in knowledge graph embedding (KGE) to assess the learning capabilities of KGE models. For each positive triple, 256 negative triples are randomly sampled. For example, replace the positive triple (disaster_record_6b65e22d, nade:hasDisasterType, landslide) with the negative triple (disaster_record_6b65e22d, nade:hasDisasterType, hidden_fault). The dataset is divided according to the generated triples for training. We calculate the MRR, Hits@1, Hits@3, and Hits@10 on the test set, and the results are presented in Table 5. Table 5 illustrates that the ComplEx model performs exceptionally well, achieving the highest MRR score, indicating its significant advantage in accurately predicting triple groups. Furthermore, it demonstrates a high likelihood of identifying the correct triple in the first, first three, and first ten positions. In contrast, the RotateE model performs poorly, particularly in terms of MRR and Hits@1 scores, which are relatively low. This suggests limitations in its average ranking quality and the likelihood of correctly identifying the triple in the first position. In practical applications, we predict the susceptibility of each grid cell based on the KGE model, generating a susceptibility map for the study area, as shown in Figure 10. From the figure, it is evident that TransE tends to categorize regions densely populated with landslide points as high-risk zones. However, concurrently, TransE is prone to misclassifying less densely populated landslide point areas as high-risk zones. In contrast, ROTATE and ComplEx exhibit a weaker ability to differentiate landslide-prone areas. RESCAL and DistMult demonstrate a stronger capability to distinguish landslide risk zones, with RESCAL leaning towards classifying non-landslide areas as “very-low” risk zones, and DistMult tending to categorize non-landslide areas as “low”-risk zones.

Table 4. Precision, Recall, and F_1 score results on the dataset (best values in bold).

Model	Precision	Recall	F_1
SVM	0.70	0.75	0.72
RF	0.83	0.83	0.83
KNN	0.82	0.78	0.80
GCF	0.87	0.90	0.88
TransE	0.85	0.89	0.87
RESCAL	0.90	0.93	0.91
DistMult	0.91	0.90	0.90
ComplEx	0.92	0.92	0.92
RotateE	0.89	0.86	0.87

Table 5. Learning results of different KGE models on the knowledge graph (best values in bold).

Model	MRR	Hits@1	Hits@3	Hits@10
TransE	0.555	0.489	0.567	0.695
RESCAL	0.612	0.529	0.639	0.783
DistMulti	0.580	0.502	0.596	0.745
ComplEx	0.653	0.569	0.684	0.836
RotateE	0.361	0.221	0.427	0.648

In addition, data scarcity poses a common challenge in landslide prediction [49], and we also conducted tests to evaluate the model’s performance under data-sparse conditions. Initially, we randomly selected half of the landslide records and created a dataset following the steps of the data-driven method. Subsequently, we trained and tested the model using machine learning, calculating the prediction accuracy. Simultaneously, we concealed the

unselected landslide records in the knowledge graph, training the KGE model using only the chosen landslide records, and calculated the prediction accuracy. The comparative results are presented in Table 6.

To comprehensively evaluate the performance of each KGE model in generating embedded representations, we selected several typical entities from the knowledge graph. These entities encompass the study area tiles at various distances, different types of LCFs, instances of disaster tasks, and instances of disaster records. We then calculated their similarity in the embedding space, and the resulting similarity matrix, with entities numbered for reference, is shown in Figure 11. We observed that the RESCAL (Figure 11b) and ComplEx (Figure 11d) models excel in entity differentiation, particularly in their ability to distinguish between entities of the same category and entities from different categories, surpassing the other models. However, RESCAL exhibits relative weaknesses in representing LCF entities, and ComplEx occasionally experiences errors in the embedded representation of specific entities. In contrast, TransE (Figure 11a) effectively discriminates between the entities of different categories but tends to struggle in distinguishing entities within the same category. RotateE (Figure 11e), on the other hand, places greater emphasis on capturing the unique semantic information of each entity but is somewhat less effective in distinguishing between the entities of different categories. It is speculated that RotateE tends to focus on denser nodes in the knowledge graph and may not pay sufficient attention to sparse graphs. Furthermore, DistMult (Figure 11c) outperforms TransE in representing entities within the same category, although it may exhibit some limitations in capturing the semantic information of certain entities. In general, for sparse categories in the knowledge graph, where the degree of a category node is low, KGE models typically exhibit a weaker learning performance. Conversely, for dense categories in the knowledge graph, KGE models tend to demonstrate stronger learning effects.

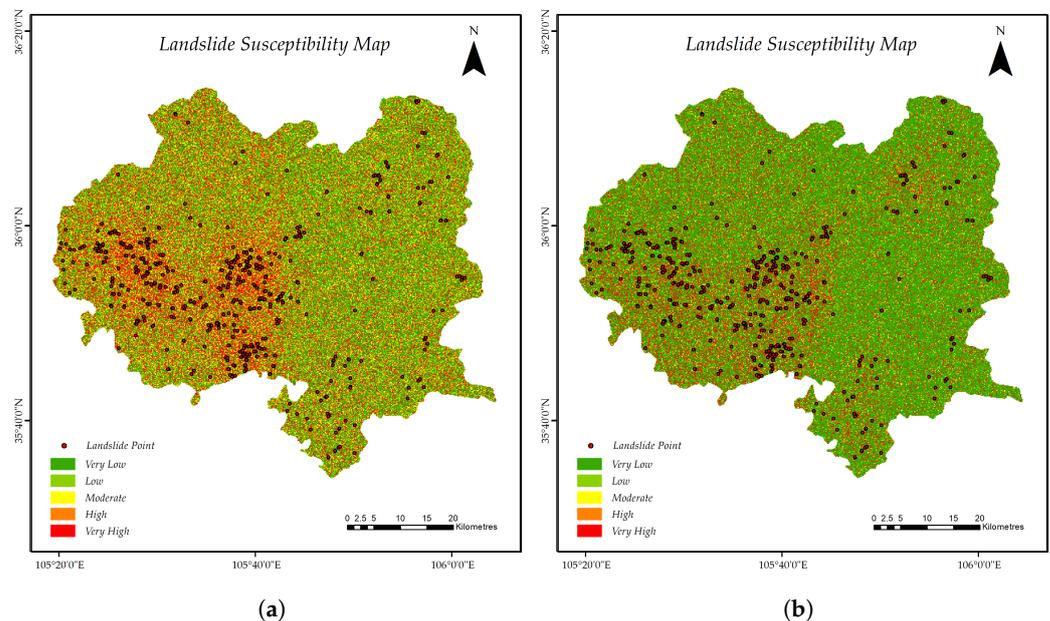


Figure 10. Cont.

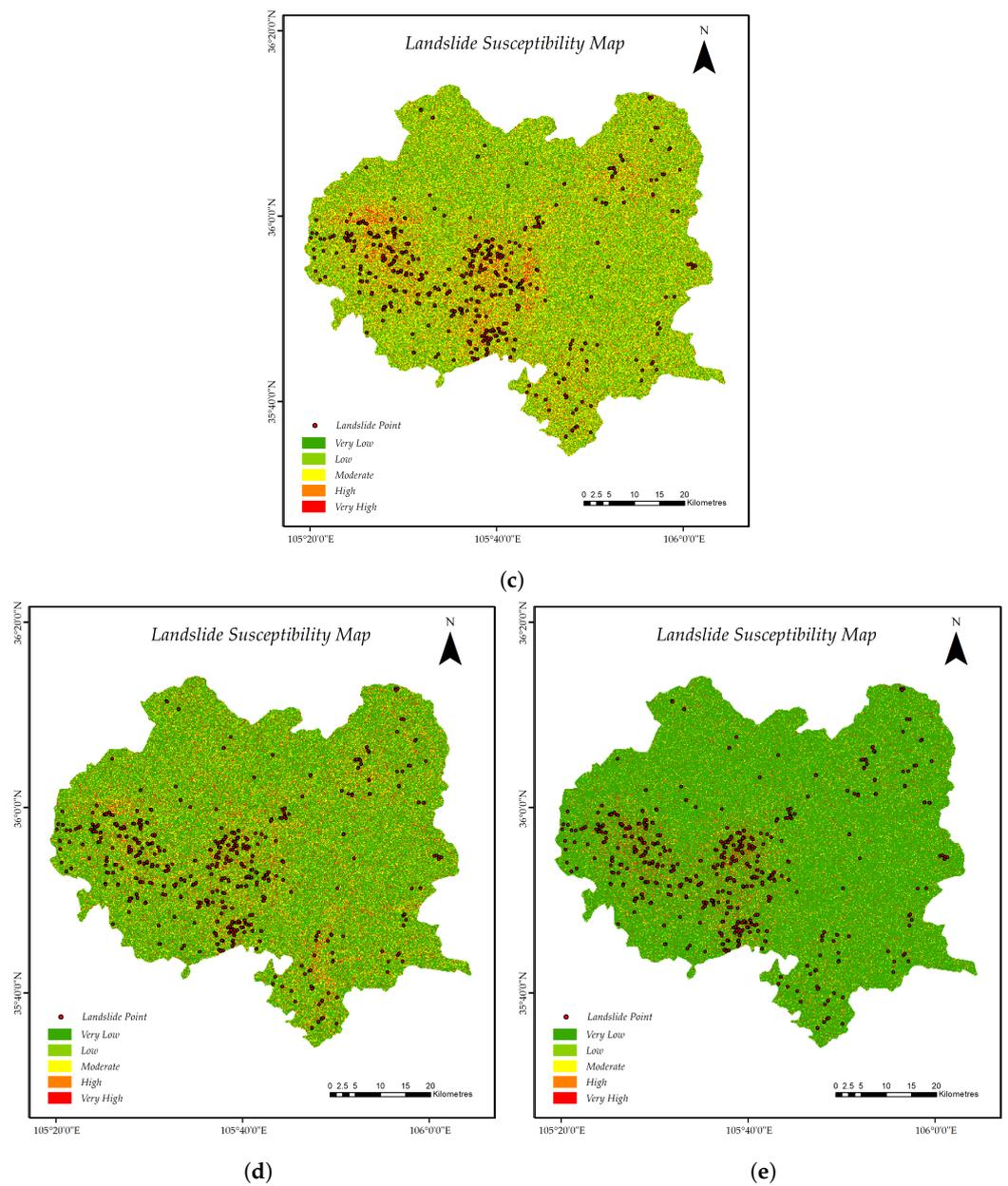


Figure 10. Landslide susceptibility map generated with the KGE method using (a) TransE; (b) RESCAL; (c) DistMult; (d) ComplEx; and (e) ROTATE.

Table 6. Results of *Precision*, *Recall*, and F_1 score on the sparse dataset (best values in bold).

Model	<i>Precision</i>	<i>Recall</i>	F_1
SVM	0.56	0.52	0.54
RF	0.69	0.64	0.66
KNN	0.64	0.65	0.64
GCF	0.71	0.68	0.69
TransE	0.71	0.76	0.73
RESCAL	0.84	0.79	0.81
DistMult	0.78	0.82	0.80
ComplEx	0.81	0.81	0.81
RotateE	0.73	0.69	0.71

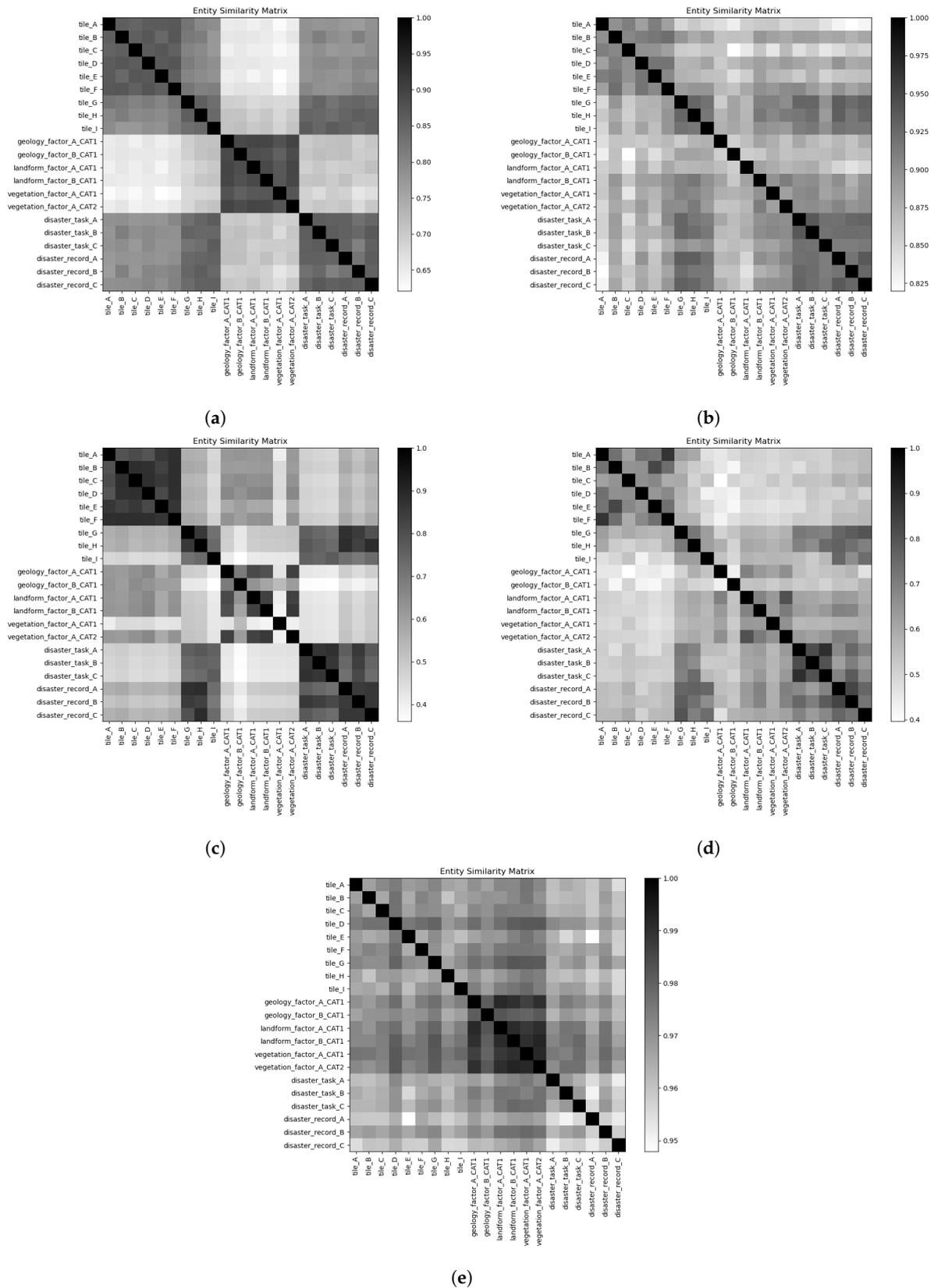


Figure 11. Comparison of vector similarity for different KGE models, including (a) TransE; (b) RESICAL; (c) DistMult; (d) ComplEx; and (e) ROTATE, after mapping entities to embedding space. Darker colors indicate higher similarity between corresponding entities on the horizontal and vertical axes.

5. Discussion

In our experiments, we compare the general landslide prediction system with our prediction system to demonstrate the feasibility and advantages of modeling landslide scenes based on a knowledge graph. On one hand, the experimental results show that the knowledge graph-based modeling of landslide scenarios is more useful for discovering spatial patterns in the landslide process than the traditional modeling method based on the “sequence” structure. This benefit stems from the nature of the knowledge graph in the process of data organization and representation, which has the ability to convey semantic information in model training, thus enhancing the model performance. Additionally, based on the results of entity similarity comparison, we found that the KGE model can indeed learn logical entity embedding representations, with semantically similar entities having similar distances in the vector space. This contributes to the correct results in the link prediction process. Similarly, the susceptibility map generated using the KGE model demonstrates this semantic representation capability. Moreover, the advantages of KGE models become more significant when the dataset is sparse. Data-driven models usually require substantial data for training and prediction. With sparse data, it is challenging for these models to effectively learn and generalize from missing data. The advantage of KGE models lies in the mapping of entities and relations to the embedding space, which assists the models in inferring missing values and gaining a deeper understanding of the underlying relationships in the data. It is worth noting that the tile level also affects the performance of the model. A smaller tile level can enhance the training and prediction speed of the model. However, due to the lower resolution, there are too many LCFs on each tile, preventing the model from fully learning the relationship between LCFs and grid cells, ultimately reducing the model’s performance. On the other hand, if the tile level is too high, it significantly decreases the training and prediction speed of the model. Additionally, it may lead to an overly sparse knowledge graph, hindering the model’s ability to effectively learn features. Therefore, selecting an appropriate tile level is crucial.

On the other hand, predicting landslides using KGE models is a novel and comprehensive end-to-end method. General data-driven methods typically involve manually selecting, designing, and extracting environmental features, and then using those features to train a model for a prediction task. These methods typically require multiple steps, including data preprocessing and LCF analysis and selection. These steps often necessitate the involvement of domain experts and multiple individual modules. For KGE models, in contrast, the data preprocessing step is performed only once when constructing the knowledge graph, enhancing data reusability. Moreover, KGE models generate predictions directly from the embedding space, eliminating the need for manual feature selection or multiple preprocessing steps. Thus, using the KGE model to predict landslides reduces the complexity of manual intervention and engineering design, making the model more easily scalable to other hazard tasks. Based on our experiments, we believe that this method is promising.

However, there are some limitations to our method that are worth noting. In the case of complete data, the advantage of the KGE method over generalized machine learning methods is not very significant, although it exhibits a slight advantage in prediction performance. This is mainly due to the sparsity of the structure of the constructed knowledge graph and the inherent limitations of the KGE model’s learning capabilities. Additionally, when performing data-driven landslide prediction based on the negative samples, which are not truly non-landslide areas, errors may occur during the sample production process, subsequently affecting the quality of the test set. Furthermore, in terms of the training details of the KGE model, the way in which the head entity and tail entity are replaced in negative triples can also affect the performance of the KGE model.

6. Conclusions

Data-driven methods typically simplify landslide scenarios during modeling, resulting in information loss during the prediction process. To address this challenge, this paper

presents a novel approach to landslide prediction. We represent complex disaster scenarios by designing the schema layer of the knowledge graph and organize the multi-source heterogeneous disaster data into triples by constructing the data layer of the knowledge graph and mapping it to the schema layer. Subsequently, landslide prediction is conducted using the KGE model. The novelty of the experimental results lies in demonstrating the capability of knowledge graph modeling for complex disaster scenarios, addressing the issue of information loss that occurs in data-driven approaches during modeling. The primary contributions of this paper can be summarized as follows:

- For the first time, a knowledge graph embedding method is applied to landslide prediction, resulting in a performance improvement, marking an innovative approach in this field.
- With the assistance of a graph-based modeling method, we improve the exploration of spatial information within landslide scenarios.
- We introduced a novel end-to-end assessment method for the precise evaluation of landslide susceptibility, which holds extensive applicability.
- Our method empowers effective landslide prediction even with limited data, offering support for applications in resource-constrained environments.

In future research, a primary focus will be on refining the precision of landslide prediction, encompassing areas with sufficient sample data and those lacking historical landslide records, as experimental results suggest promising potential in both scenarios. Addressing the prediction bias stemming from the sparse knowledge graph structure will entail an exploration of a more sophisticated schema layer. Concurrently, improvements to the model structure will be actively pursued, including the integration of graph neural network models to capture higher-order interactions among entities and relationships. This strategic enhancement aims to enhance the predictive capabilities of the model. Furthermore, endeavors will be directed towards expanding the scope of downstream tasks to include other disaster types, thereby augmenting the method's versatility and the utility of disaster data.

Author Contributions: Conceptualization, L.C., L.P. and L.Y.; methodology, L.C. and L.P.; validation, L.C.; resources, L.C.; data curation, L.C.; writing—original draft preparation, L.C.; writing—review and editing, L.C. and L.Y.; funding acquisition, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ningxia Key R&D Program (2020BFG02013). This work was sponsored by Tianjin intelligent manufacturing special fund project (NO.20201198).

Data Availability Statement: Data are contained within the article.

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

References

1. Pavlinovic, D. Climate and Weather-Related Disasters Surge Five-Fold over 50 Years, but, Early Warnings Save Lives-WMO Report. *UN News* 1 September 2021. Available online: <https://news.un.org/en/story/2021/09/1098662> (accessed on 24 December 2023).
2. Li, H.; He, Y.; Xu, Q.; Deng, J.; Li, W.; Wei, Y. Detection and segmentation of loess landslides via satellite images: A two-phase framework. *Landslides* **2022**, *19*, 673–686. [CrossRef]
3. Tan, Q.; Bai, M.; Zhou, P.; Hu, J.; Qin, X. Geological hazard risk assessment of line landslide based on remotely sensed data and GIS. *Measurement* **2021**, *169*, 108370. [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. Ohki, M.; Abe, T.; Tadono, T.; Shimada, M. Landslide detection in mountainous forest areas using polarimetry and interferometric coherence. *Earth Planets Space* **2020**, *72*, 67. [CrossRef]
6. Hussain, Y.; Schlögel, R.; Innocenti, A.; Hamza, O.; Iannucci, R.; Martino, S.; Havenith, H.B. Review on the geophysical and UAV-based methods applied to landslides. *Remote Sens.* **2022**, *14*, 4564. [CrossRef]
7. Feng, K.; Huang, D.; Wang, G.; Jin, F.; Chen, Z. Physics-based large-deformation analysis of coseismic landslides: A multiscale 3D SEM-MPM framework with application to the Hongshiyuan landslide. *Eng. Geol.* **2022**, *297*, 106487. [CrossRef]

8. Alcántara-Ayala, I.; Ribeiro Parteli, E.J.; Pradhan, B.; Cuomo, S.; Vieira, B.C. Physics and modelling of landslides. *Front. Phys.* **2023**, *11*, 83. [CrossRef]
9. Akgun, A.; Sezer, E.A.; Nefeslioglu, H.A.; Gokceoglu, C.; Pradhan, B. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput. Geosci.* **2012**, *38*, 23–34. [CrossRef]
10. Muthu, K.; Petrou, M. Landslide-hazard mapping using an expert system and a GIS. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 522–531. [CrossRef]
11. Sabokbar, H.F.; Roodposhti, M.S.; Tazik, E. Landslide susceptibility mapping using geographically-weighted principal component analysis. *Geomorphology* **2014**, *226*, 15–24. [CrossRef]
12. Thiery, Y.; Malet, J.P.; Sterlacchini, S.; Puissant, A.; Maquaire, O. Landslide susceptibility assessment by bivariate methods at large scales: Application to a complex mountainous environment. *Geomorphology* **2007**, *92*, 38–59. [CrossRef]
13. Tang, Y.; Feng, F.; Guo, Z.; Feng, W.; Li, Z.; Wang, J.; Sun, Q.; Ma, H.; Li, Y. Integrating principal component analysis with statistically-based models for analysis of causal factors and landslide susceptibility mapping: A comparative study from the loess plateau area in Shanxi (China). *J. Clean. Prod.* **2020**, *277*, 124159. [CrossRef]
14. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhau, 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. [CrossRef]
15. Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhau, I.; Prakash, I.; Bui, D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total. Environ.* **2018**, *627*, 744–755. [CrossRef] [PubMed]
16. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženilek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [CrossRef]
17. 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]
18. Hogan, A.; Blomqvist, E.; Cochez, M.; d’Amato, C.; Melo, G.D.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S.; et al. Knowledge graphs. *Acm Comput. Surv.* **2021**, *54*, 1–37. [CrossRef]
19. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Philip, S.Y. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [CrossRef]
20. Ge, X.; Yang, Y.; Peng, L.; Chen, L.; Li, W.; Zhang, W.; Chen, J. Spatio-temporal knowledge graph based forest fire prediction with multi source heterogeneous data. *Remote Sens.* **2022**, *14*, 3496. [CrossRef]
21. Liu, X.; Zhang, Y.; Zou, H.; Wang, F.; Cheng, X.; Wu, W.; Liu, X.; Li, Y. Multi-source knowledge graph reasoning for ocean oil spill detection from satellite SAR images. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103153. [CrossRef]
22. Wang, Q.; Mao, Z.; Wang, B.; Guo, L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2724–2743. [CrossRef]
23. Huang, X.; Zhang, J.; Li, D.; Li, P. Knowledge graph embedding based question answering. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, Melbourne, VIC, Australia, 11–15 February 2019; pp. 105–113.
24. Zhang, C.; Liu, M.; Liu, Z.; Yang, C.; Zhang, L.; Han, J. Spatiotemporal activity modeling under data scarcity: A graph-regularized cross-modal embedding approach. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
25. Eyharabide, V.; Bekkouch, I.E.I.; Constantin, N.D. Knowledge graph embedding-based domain adaptation for musical instrument recognition. *Computers* **2021**, *10*, 94. [CrossRef]
26. Golon, D.K. *The Land Processes Distributed Active Archive Center (LP DAAC)*; US Geological Survey: Reston, WV, USA, 2016.
27. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The shuttle radar topography mission. *Rev. Geophys.* **2007**, *45*. [CrossRef]
28. OpenStreetMap. 2023. Available online: <https://wiki.openstreetmap.org/wiki/Key:water> (accessed on 8 August 2023).
29. Resource and Environment Science and Data Center. Data on the Spatial Distribution of Soil Types in China. 2023. Available online: <https://www.resdc.cn/data.aspx?DATAID=145> (accessed on 8 August 2023).
30. OpenStreetMap. 2023. Available online: <https://wiki.openstreetmap.org/wiki/Highways> (accessed on 8 August 2023).
31. Merghadi, A.; Yunus, A.P.; Dou, J.; Whiteley, J.; ThaiPham, B.; Bui, D.T.; Avtar, R.; Abderrahmane, B. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Sci. Rev.* **2020**, *207*, 103225. [CrossRef]
32. Steger, S.; Mair, V.; Kofler, C.; Pittore, M.; Zebisch, M.; Schneiderbauer, S. Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling—Benefits of exploring landslide data collection effects. *Sci. Total. Environ.* **2021**, *776*, 145935. [CrossRef] [PubMed]
33. Battersby, S.E.; Finn, M.P.; Usery, E.L.; Yamamoto, K.H. Implications of web Mercator and its use in online mapping. *Cartogr. Int. J. Geogr. Inf. Geovisualizat.* **2014**, *49*, 85–101. [CrossRef]
34. Decker, S.; Melnik, S.; Van Harmelen, F.; Fensel, D.; Klein, M.; Broekstra, J.; Erdmann, M.; Horrocks, I. The semantic web: The roles of XML and RDF. *IEEE Internet Comput.* **2000**, *4*, 63–73. [CrossRef]
35. McBride, B. The resource description framework (RDF) and its vocabulary description language RDFS. In *Handbook on Ontologies*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 51–65.

36. Battle, R.; Kolas, D. Geosparql: Enabling a geospatial semantic web. *Semant. Web J.* **2011**, *3*, 355–370. [[CrossRef](#)]
37. Car, N.J.; Homburg, T. GeoSPARQL 1.1: Motivations, details and applications of the decadal update to the most important geospatial LOD standard. *Isprs Int. J. Geo-Inf.* **2022**, *11*, 117. [[CrossRef](#)]
38. Brank, J.; Grobelnik, M.; Mladenic, D. A survey of ontology evaluation techniques. In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005), Citeseer, Slovenia, 17 October 2005; pp. 166–170.
39. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Adv. Neural Inf. Process. Syst.* **2013**, *2*, 2787–2795.
40. Nickel, M.; Tresp, V.; Kriegel, H.P. A three-way model for collective learning on multi-relational data. In Proceedings of the Icm1, Bellevue, WA, USA, 28 June–2 July 2011; Volume 11, pp. 3104482–3104584.
41. Yang, B.; Yih, W.t.; He, X.; Gao, J.; Deng, L. Embedding entities and relations for learning and inference in knowledge bases. *arXiv* **2014**, arXiv:1412.6575.
42. Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; Bouchard, G. Complex embeddings for simple link prediction. In Proceedings of the International Conference on Machine Learning, PMLR, New York, NY, USA, 20–22 June 2016; pp. 2071–2080.
43. Sun, Z.; Deng, Z.H.; Nie, J.Y.; Tang, J. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv* **2019**, arXiv:1902.10197.
44. Zheng, D.; Song, X.; Ma, C.; Tan, Z.; Ye, Z.; Dong, J.; Xiong, H.; Zhang, Z.; Karypis, G. Dgl-ke: Training knowledge graph embeddings at scale. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual, 25 July 2020; pp. 739–748.
45. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
46. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
47. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883. [[CrossRef](#)]
48. Zhou, Z.H.; Feng, J. Deep forest. *Natl. Sci. Rev.* **2019**, *6*, 74–86. [[CrossRef](#)] [[PubMed](#)]
49. Chen, L.; Ge, X.; Yang, L.; Li, W.; Peng, L. An Improved Multi-Source Data-Driven Landslide Prediction Method Based on Spatio-Temporal Knowledge Graph. *Remote Sens.* **2023**, *15*, 2126. [[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.